

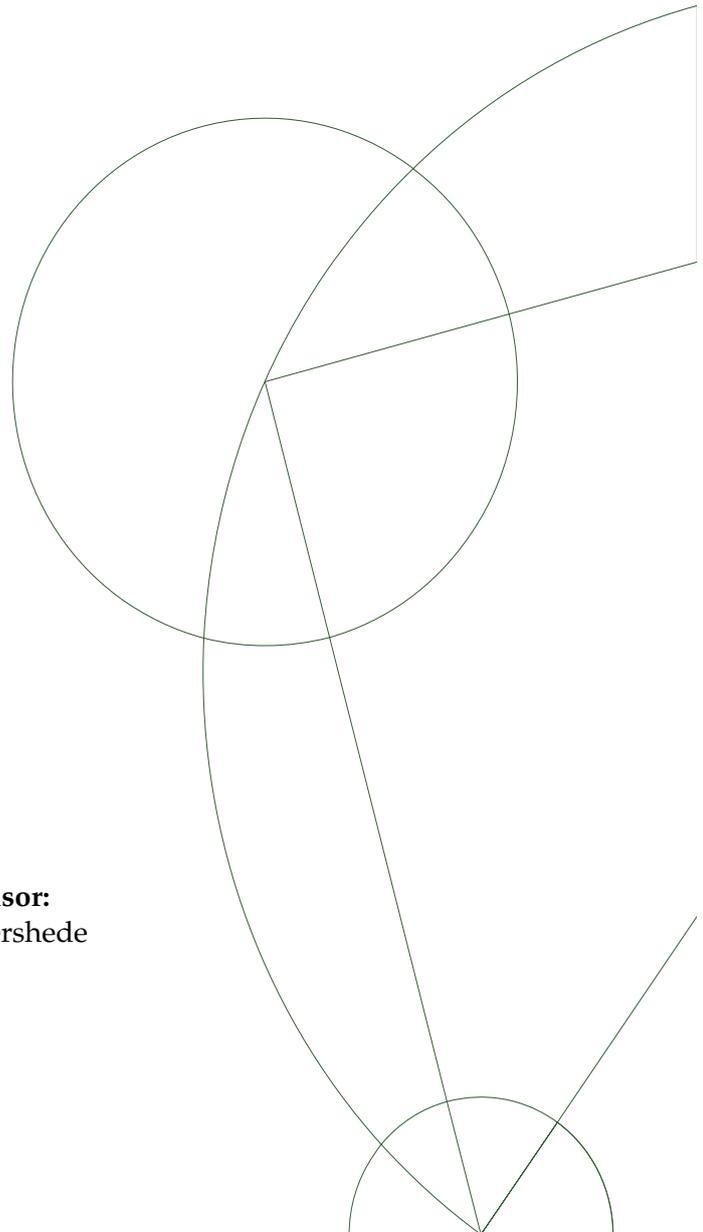


Master's thesis, April 27, 2012

Investigating RNA pseudoknots using a double optical trap

by
Rebecca Bolt Ettliger
Niels Bohr Institute, University of Copenhagen

Supervisor:
Lene Oddershede



Contents

Contents	i
Abstract	v
Dansk opsummering	vii
Acknowledgements	ix
1 Introduction	1
1.1 Pulling molecules with optical traps - historical context	1
1.2 Biology of RNA pseudoknots	2
1.3 Thermodynamics and kinetics	4
1.4 Focus of investigation and hypotheses	6
1.5 Structure of the report	6
2 Optical trapping - theory and calibration	7
2.1 Optical trapping theory	7
2.2 Optical trap setup	9
2.3 Calibration	10
2.3.1 Stokes calibration of force	11
2.3.2 Brownian motion calibration	12
2.3.3 Power spectrum calibration	13
2.3.4 Fitting to the power spectrum	15
2.3.5 Sources of calibration error	16
3 Theoretical background - molecular transitions	19
3.1 Pseudoknot opening - the molecular transition investigated	19
3.2 Polymer stretching theory	19
3.3 Kinetics	22
3.3.1 The Dudko-Hummer-Szabo kinetics approach	25
3.4 Nonequilibrium thermodynamics	26
3.4.1 Free Energy, the Jarzynski Equality and the Crooks Fluctuation Theorem	26
3.4.2 Using the Jarzynski Equality and the Crooks Fluctuation Theorem	27
3.4.3 Work distributions and handles	28
3.4.4 Measuring work distributions - two methods of finding ΔG_{PK}	29
3.4.5 Calculating the work to stretch handles and pseudoknots	31
3.5 Summary of molecular transition theory	31
4 Materials and Methods	33
4.1 The pseudoknots - structure and frameshift efficiencies	33
4.2 The molecular constructs	36

4.3	Biochemical methods	37
4.3.1	Preparing samples for pulling experiments	38
4.4	Pulling experiments	39
4.4.1	Sample chamber and trapping routine	39
4.4.2	Nanotracker particularities	41
4.5	Data processing and statistics	42
4.5.1	Using the EWLC to predict force-extension curves and rip/zip lengths	43
4.6	Summary of methods	44
5	Results Part I: Rips and zips	45
5.1	Typical force-extension curves	45
5.2	Criteria for accepting rips and zips for analysis	51
5.3	Data overview	52
5.4	Rips and zips: forces and lengths	54
5.4.1	Exploring possible subdistributions of rips and zips	57
5.4.2	Identification of rip subdistributions	62
5.5	Testing bias from “super tethers”	62
5.6	Summary of points made in this Chapter	66
6	Discussion Part I: possible transitions and intermediates	69
6.1	Rip forces	69
6.2	Measured versus predicted rip and zip length	70
6.2.1	EWLC parameter sensitivity	71
6.3	Subdistributions and intermediate states	72
7	Results Part II: Gibbs free energy and kinetic parameters	77
7.1	Work distributions and ΔG	77
7.1.1	Finding ΔG_{PK} using the simple area approach	78
7.1.2	Finding ΔG_{PK} using the Collin approach	83
7.2	Kinetics: distance to transition state	87
7.2.1	Kinetic parameters: Subdistributions and reverse transitions	88
7.2.2	Comparison of kinetic parameters to literature values	90
7.3	Summary of points made in this chapter	93
8	Discussion Part II: energy and kinetics	95
8.1	Energy Landscapes and intermediate states	95
9	Conclusion and perspectives	97
9.1	Background for the conclusion	97
9.1.1	Basic differences	97
9.1.2	Pseudoknot versus hairpin	98
9.1.3	Unexpected structures	98
9.1.4	Predicting tertiary structure	99
9.1.5	Conclusion summary	99
9.2	Perspectives	100
9.2.1	Further pseudoknot pulling investigations	100
9.2.2	Further thermodynamic investigations	100
9.2.3	Experimental improvements and ideas for expanded pulling experiments	101
	Bibliography	103
A	Power spectrum and other calibration treatment details	109

A.1	Fourier transforms and power spectrum construction	109
A.2	Solving the Langevin Equation	110
A.3	Selection of the best power spectrum fitting routine	111
A.3.1	Quantification of Nanotracker noise	111
B	Frameshift and stall efficiency investigation	115
C	Biochemical methods	117
C.1	Preparing RNA-handle constructs: downstream handle preparation	117
C.2	DNA plasmids and transcription of DNA to make RNA	117
C.2.1	Preparing RNA-handle constructs: template amplification	118
C.2.2	Preparing RNA-handle constructs: RNA synthesis	118
C.3	Preparing RNA-handle constructs: annealing handles to RNA	119
C.4	Annealing RNA constructs to beads for pulling experiments	119
D	Testing bias from super tethers	121
E	Additional data for separating Rip and Zip subdistributions	125
E.1	Force distributions of rips and zips	125
E.2	Results of statistical tests	125
F	Additional data on work and kinetics	129
F.1	Work distributions - single distribution for all work data treated with simple approach	129

Abstract

This thesis uses optical tweezers to investigate the mechanical properties of two closely related RNA molecules. The aim is to better understand a class of messenger RNA which contain so-called pseudoknot structures through examination of their mechanical properties. In particular, the correlation of their mechanical strength to their frameshifting interaction with translating ribosomes is investigated.

Pseudoknots are structures on messenger RNA (mRNA) that have been found in organisms as diverse as bacteria and humans. In viruses such as SARS and Infectious Bronchitis Virus they appear to play an important role in protein synthesis because they can cause frameshifting of ribosomal translation on the piece of mRNA in which they occur. Frameshifting allows the synthesis of two different types of protein product from the same mRNA.

The thesis is part of a larger project examining both function and mechanical properties of a range of pseudoknots; together these investigations may help show how RNA sequences fold into pseudoknots and how the pseudoknots affect ribosomal function and thus translation of messenger RNA into protein products.

The results indicate that pseudoknot-containing RNA sequences may dynamically take on many different structural conformations, some of which are strong and brittle pseudoknots, some weaker pseudoknots, and some not really pseudoknot-resembling at all. Instead the RNA sequence may form hairpins or other complex but unstable structures. The likelihood that a pseudoknot-containing mRNA sequence forms a strong pseudoknot does appear to be correlated to its frameshifting efficiency.

Dansk opsummering

Denne specialeafhandling benytter en forsøgsopstilling med to optiske pincetter til at undersøge de mekaniske egenskaber af to nært beslægtede RNA-molekyler. Målet er at opnå en forbedret forståelse af en hel gruppe af messenger-RNA, der indeholder en sekvens, som kan folde sig som en såkaldt pseudoknude. Undersøgelsen drejer sig i særdeleshed om at studere, om der er overensstemmelse mellem den mekaniske styrke af de undersøgte pseudoknuder og deres evne til at forårsage ændringer af læserammen af translaterende ribosomer.

Pseudoknudesekvenser på messenger-RNA (mRNA) er observeret i alle mulige former for organismer lige fra bakterier til mennesker. I virusser såsom SARS og bronchitis har pseudoknuder øjensynlig en vigtig rolle at spille i proteinsyntese, idet de kan forårsage skift af ribosomets læseramme på mRNA'et og dermed muliggør syntese af to forskellige proteinprodukter fra samme mRNA.

Nærværende speciale er del af et større projekt, som undersøger en række pseudoknuders mekaniske og biologiske funktioner og karakteristika. Tilsammen kan disse undersøgelser være med til at vise, hvordan og hvornår pseudoknuder folder ud fra en given RNA-sekvens, og hvordan pseudoknuder påvirker ribosomfunktion og proteinproduktion i vira og celler.

Resultaterne peger på, at RNA-sekvenser, som indeholder mulighed for at folde som pseudoknuder, ikke altid gør det. I stedet tyder det på, at RNA-sekvenser, som de, der er undersøgt her, dynamisk kan ændre struktur fra eet foldemønster til et andet. Nogle gange opstår en stærk pseudoknude, andre gange måske en svagere pseudoknude eller slet ingen pseudoknude afhængig af sekvens. Ofte ser det ud til at sekvensen folder som en simpel hårnålestruktur eller anden relativt ustabil struktur. Det virker sandsynligt, at jo mere tilbøjelig strukturen er til at antage en stærk pseudoknudestruktur, jo større er chancen for at den skaber ændringer i den ribosomale læseramme.

Acknowledgements

Thanks goes first of all to my advisor, Lene Oddershede, for your regular, thoughtful and conscientious guidance throughout this project. I want to thank you for suggesting a very concrete thesis topic that was part of a larger project and had a relatively high chance of success. It has meant a lot to me to be able to work with and discuss all my issues and findings with others. Thank you for your patience and good advice right through the process.

Thank you to Jesper Tholstrup for all our collaboration in the last year and a half, for your meticulous attention to detail and will to think things through, for providing unending explanations and help, for great discussions on many difficult concepts, and for guiding me through what was initially very daunting biochemical protocols.

Thank you to Michael Sørensen for your guidance and patient explanations of biochemical concepts, and for access to your lab and all the equipment needed for the biochemical part of the experimental work.

A special thank you to Nader Reihani for indispensable help with the NanoTracker setup when it got stuck last summer and I could not get the stage to budge.

Thank you to Christian Stokholm for helping me to get to know the NanoTracker, for helping to solve problems along the way and for sharing frustration and advice on the equipment and data treatment.

Thank you to Mette and all the other friendly people who have helped me in the biochem lab at the BioCenter.

Thank you to the OT group for lots of interesting Journal Clubs and for helpful and encouraging informal discussions, board games, and companionship.

Thank you to the incredibly friendly people in the C-room and in the BioComplexity group for great fooze ball sessions, pizza expeditions, common room lunches, tea drinking sessions, DLH racing and much more. The atmosphere of the group is amazingly warm and welcoming; thank you all for making it that way.

Thank you to the people who have helped read through this thesis and make my writing better - Lene first of all of course, but also Pernille, Lise, and Ayoe.

Finally a huge thank you to my mother for all the dinners and love, to Ayoe for all the walks, talks, texts, tears and laughs, to Pernille for warm encouragement, patient ears, and good advice, to Sarah for sisterly love and for hospitality when I needed a London getaway, and to all the friends I won't even begin to list who have shared this year with me and helped carry me through it.

1 Introduction

Using a tightly focused laser beam it is possible to trap particles of micrometer dimension and move them with precision of nanometers [1]. Parts of living cells or whole cells can be captured and manipulated, as can artificial beads with functional proteins attached [2, 3]. Among numerous biological applications, optical trapping has made it possible to unzip single DNA molecules and RNA hairpins, measure the step size of molecular motors, and monitor diffusion rates of molecules inside living cells and single proteins within cell membranes [4, 5, 6, 3].

In this work, messenger RNA (mRNA) pseudoknots are unfolded and refolded using a setup with two steerable optical traps. The aim is to investigate the mechanical properties of two structurally different pseudoknots and in the context of similar investigations made by others to look for correlations between structure, mechanical strength, and frameshift efficiency of the pseudoknots, as will be explained in detail below.

First of all, this chapter will introduce optical traps, messenger RNA and RNA pseudoknots, protein translation, programmed frameshifting and other general background for the project. The questions that led to this thesis and the methods used to investigate them will follow, and finally the structure of the remainder of the thesis report will be presented.

1.1 Pulling molecules with optical traps - historical context

Laser trapping of small particles was developed by Arthur Ashkin and colleagues at AT&T Bell Laboratories in the 1970's and 80's; the history of the development is described by Ashkin in his 1997 inaugural article in PNAS upon his election to the US National Academy of Sciences [7]. The group first reported trapping of silica beads of micrometer size by a single laser in 1986 [8] and soon afterwards they and others started using this technique, which they called optical tweezers or optical traps, for investigating biological systems, starting with viruses and bacteria. The first bacteria to be trapped were killed by what Ashkin called optocution - overheating due to the green laser being used - but in 1987 the group was able to trap live *E. coli* using an infrared laser which heated the cells and their environment much less [9].

Other groups began to use optical traps to investigate the forces and movements of molecular motors in the following years. Examples of new areas of study were the protein complexes that drive the rotation of flagellae propelling the movement of bacteria, the myosin molecules that move on actin making muscle cells work, and the kinesin molecules moving cell compartments along microtubuli (part of the cytoskeleton). Some of these early applications are surveyed in an easily accessible Nature overview of optical trapping and its applications from 1992 by Stephen Block [2]. A recent and highly readable review of optical trapping applications and methods by Thomas Perkins was published in Laser & Photonics Reviews in 2009 [3]. It focuses especially on single molecule and nucleic acid applications which are highly relevant in the context of this thesis.

In the early and mid-nineties, groups started investigating the strength and elasticity of long chain-like molecules such as DNA, RNA and protein polymers using optical tweezers and other new instruments, for instance atomic force microscopes, AFMs. The results of the first experiments using optical tweezers to stretch DNA were published in 1996 by Smith et al. [10], and theory of single- and double-stranded DNA stretching was developed in the same period while additional DNA stretching experiments were carried out. An article by Wang et al. from 1997 [11] contains careful description of the experimental procedure and the theoretical fit to stretching data. Unzipping and melting of DNA due to stretching was investigated in other studies, and in 2001 the unzipping of an RNA hairpin with optical tweezers was first documented [4]. A review of many of these DNA/RNA force investigations up to 2002 is found in [5]. The goal was often both to learn more about DNA and RNA in mechanical terms and to learn how enzymes such as polymerases interact with and modify them. Already in 1998, Wang et al. [6] was able to describe the movement of a single RNA polymerase molecule on DNA. Measuring molecular motors moving on RNA proved more difficult, and only in the last five years great advances have been made in using the technique to investigate how ribosomes translate mRNA [12, 13, 14].

RNA pseudoknots are one type of RNA structure that has also been investigated using optical trapping in recent years. Their folding energy landscapes are highly complex with many different possible tertiary interactions influencing the stability of intermediate states; see, e.g., [15]. Therefore the 3D structures that pseudoknots may fold into are more numerous and difficult to predict than the structure of a hairpin [16]. Pulling a pseudoknot open to a single strand of RNA is consequently less predictable, as it can occur in several steps [17, 18, and others]. The main focus of this thesis will be to examine the unfolding and refolding transitions for a two types of RNA pseudoknots. What an RNA pseudoknot is and what role it plays in biology will be described in the next section. The aims and results of other experiments using optical trapping to investigate RNA pseudoknots and ribosomal action will be summarized also. How optical tweezers work will be described in 2.

1.2 Biology of RNA pseudoknots

Pseudoknots are tertiary structures of RNA that when they occur in mRNA and are preceded by a so-called slippery sequence of RNA nucleotides can induce minus one nucleotide frameshifting of translating ribosomes. They are important for protein synthesis since mRNA contain molecular copies of the information in the genes on DNA that are used by ribosomes to assemble amino acid chains. Thus mRNA molecules ensure that the correct sequences of amino acids are made into the protein products encoded by the genes. Frameshifting alters how ribosomes translate the three-nucleotide codons of mRNA to the amino acid chain of the protein and can therefore cause vital changes in the protein product made from a particular gene. See cartoon of a ribosome encountering an mRNA pseudoknot in Figure 1.1.

Occasionally ribosomal frameshifting happens by chance; frameshifting at higher rates is called programmed frameshifting. Programmed frameshifting by pseudoknots has been observed in viruses including retrovirus such as HIV, in bacteria, e.g. *E. coli*, in yeast and in mammalian mRNA [19]. In minus one frameshifting the ribosomal reading frame is shifted one nucleotide backwards. Plus one frameshifting can also occur at pseudoknots [20] and plus two frameshifting has been observed at other mRNA structures [19].

How exactly ribosomes translate mRNA and how minus one frameshifting occurs is not well understood in detail, but some steps in the process have been established [19]. Programmed minus one frameshifting is usually observed at the occurrence of a pseudoknot when the pseudoknot is preceded by a slippery sequence and a spacer about

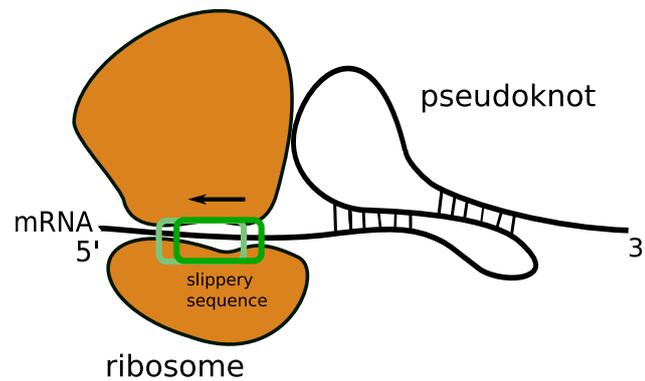


FIGURE 1.1: Cartoon of mRNA pseudoknot interfering with ribosomal translation (conceptual, not to scale). The ribosome is represented in orange as a small and a large subunit. The pseudoknot contains two stems and two loops. Frameshifting of the ribosomal reading frame by -1 nucleotide may be observed when a pseudoknot is preceded by a slippery sequence of nucleotides. This corresponds to a shift from the dark green reading frame to the light green reading frame in the figure. A slippery sequence is a series of nucleotides X XXY YYZ, where X, Y and Z stand for some or all of the ribonucleotide bases A, U, G or C. See text.

5-10 nucleotides long [21]. The basic steps involved in their interaction is explained by Giedroc and Cornish in their 2009 Viral Research review [19] and may be summarized as follows: The movement of the ribosome along the mRNA is interrupted by the pseudoknot, which must open for translation to continue. This interruption occurs when the 5' end of the pseudoknot (stem1 and/or loop2) interferes with the front of the ribosome. Thus the codon-reading A and P sites of the ribosome are still some distance (the 5-10 nucleotide space) from the pseudoknot. If a slippery sequence occurs just where the A and P sites are located when the ribosome is stalled, it may slide back by one nucleotide before the pseudoknot is opened and translation continues, now in a new reading frame.

A slippery sequence is a series of nucleotides X XXY YYZ, where X can stand for any of the ribonucleotide bases A, U, G or C, Y must stand for A or U, and Z must stand for A, U, or C [22]. The letters X, Y and Z may stand for different nucleotides or they may be the same. The sequence is believed to be slippery because transfer RNA's bind best to the first two nucleotides in each codon. Transfer RNA's located in the A and P sites of the ribosome bind the ribosome to the mRNA during translation. Therefore, if the last nucleotide of the previous codon is the same as the two first nucleotides of the next codon, the transfer RNA's will bind the ribosome almost as well to the mRNA after a one nucleotide backward shift the ribosome as they did in the original reading frame.

Though RNA pseudoknots are present in organisms from bacteria to mammals and frameshifting mRNA pseudoknots are known to be vital to the survival of vira such as HIV and Infectious Bronchitis Virus, pseudoknot structure and in particular differences in frameshift efficiency due to pseudoknots of varying structure are not yet well understood [23, 19]. Frameshift efficiency is the percentage of ribosomes that are frameshifted by an encounter with a frameshift-inducing structure such as a slippery sequence and a pseudoknot. The fact that not all ribosomes are frameshifted but only a certain percentage, means that in principle two different vital proteins can be encoded by the same piece of mRNA, reducing the amount of space that the genome needs to take up.

A pseudoknot is made up of stems of RNA basepairs with loops of unpaired bases in

between. Stems and loops in naturally occurring pseudoknots vary widely in length and base composition [24]. Though the sequence of bases can be readily determined and possible pseudoknot formation can also be predicted, the 3D structure of the pseudoknot and thus its overall stability are difficult to predict. Small variations in stem length and base pair composition of artificial pseudoknots have been observed to result in surprisingly large differences in both mechanical strength and frameshift efficiency [17]. The differences cannot be explained simply by calculating the free energy change of unpairing the basepairs of the stems [15]; however, frameshift efficiency has been found to correlate with the mechanical strength of the pseudoknots in some instances [17, 21]

Recently Tholstrup et al. [20] showed a third possible outcome from an encounter between a ribosome and a pseudoknot: apart from assembling a frameshifted or non-frameshifted protein product, the ribosome may also get stuck within the pseudoknot. This ribosome blocking appeared to correlate with predicted pseudoknot strength. Thus although stronger pseudoknots may more often cause frameshifting, they may also cause ribosomes to stall if they become too strong, and the pseudoknots that appear to give rise to most frameshifted protein products probably reflect a balance between these two effects.

A number of previous studies [17, 18, 25, 21] all use optical trapping to investigate how the mechanical properties of pseudoknots may reflect the interaction between ribosome and pseudoknot. In each, a series of artificial pseudoknots with systematically varied structures are produced. The designed pseudoknot structures have some similarities to naturally occurring ones, but are seldom wild type. The effect of the structural variation on observables such as unfolding force, evidence of intermediate states, the Gibbs free energy of unfolding and the molecule's brittleness are investigated. These parameters will be presented in more detail below.

Other optical trapping studies have examined ribosome-mRNA interaction. In 2007, Uemura et al. [12] measured the force required to pull a ribosome away from mRNA at the start of translation. They discovered that the bond was strongest during the first part of protein synthesis and weakened in a subsequent step which might correspond to the moment where the ribosome moves along the mRNA to read the next codon. The following year Wen et al. [13] measured the translation of mRNA hairpin structures by single ribosomes using optical trapping. They were able to detect distinct steps of hairpin opening corresponding to one codon each with pauses between steps.

In 2011 the same group demonstrated variation in the rate with which the ribosome opened an mRNA hairpin depending on the force with which the hairpin was pulled by the optical traps and depending on the strength of the bonds in the hairpin (GC base pairs are stronger than AU base pairs, so by substituting GC for AU, the strength can be varied without changing the overall geometry much). The group saw evidence that ribosomes open the hairpin through two mechanisms: 1) by creating a passive enzymatic bias for the open conformation; and 2) by using mechanical force to pry the hairpin open in a step that requires energy [14]. This matches the two mechanical models highlighted in [19] of how frameshifting pseudoknots may open when encountered by a ribosome: that the pseudoknot acts as a barrier to the ribosome, tension builds up as the ribosome "attempts" to move during translocation, and the tension is released when the ribosome frameshifts by minus one nucleotide. The build-up of tension would create force on the pseudoknot.

1.3 Thermodynamics and kinetics

As in previous studies of pseudoknots by optical trapping, in this work pseudoknots are attached to artificial beads by linkers and the beads are held and moved by two optical traps to coerce the pseudoknots to open and close. The basic layout of the single molecule experiment carried out is shown in Figure 1.2. The forces on the beads and their positions in

the traps are measured as one trap is moved, extending the molecule so that the pseudoknot opens. The resulting signal is converted into a graph of force versus extension in which sudden openings and closings of the pseudoknot can be clearly seen as “rips” and “zips”.

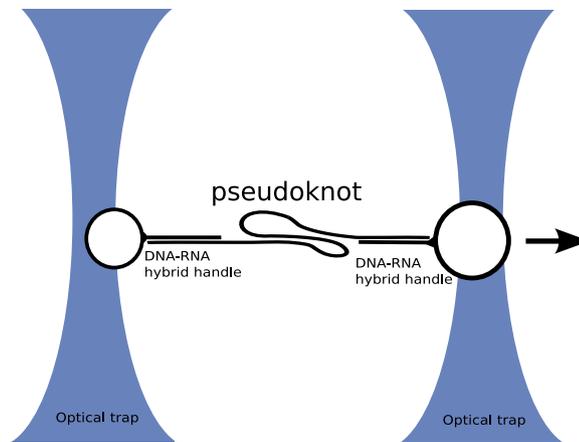


FIGURE 1.2: Experimental layout - not to scale. An RNA pseudoknot is caught between two polystyrene beads held in optical traps. The pseudoknot is attached to DNA strands so that it is flanked by DNA-RNA hybrid handles. One handle is bound by a digoxigenin-antidigoxigenin complex to a small polystyrene bead, the other is bound by biotin-streptavidin bonds to a slightly larger bead. One trap is moved laterally away from the other, pulling the pseudoknot open by shear force.

The observations that can be extracted directly from these measurements include the unfolding force and length of unfolding, the number of steps that the unfolding occurs in and thus possible observation of intermediate states or alternatively folded states. The transition from the folded to the unfolded state may take place in a single clear step, in several steps or gradually. The lengths of the unfolding rips can be measured and compared to values calculated from theories of elasticity and molecular conformation, using the expected lengths of the stems and loops that unfold. Observations of the unfolding/folding pathway and the distributions of unfolding and refolding forces and lengths were in [18] and [21] used as clues that folded states were observed which were not the expected “native” pseudoknot and that in some cases the mRNA hopped between one partially folded state and another partially folded state before fully unfolding or folding.

From the distribution of unfolding forces, kinetic parameters of the pseudoknot may be extracted that help characterize these states, namely the transition rate and the distance to the transition state, called x^\ddagger , which quantifies the brittleness of the molecule. These parameters tell us about the energy landscape of the pseudoknot: how high and how steep are the energy barriers between different conformations when, e.g., the extension of the molecule or its thermal energy are varied.

Finally, the Gibbs free energy difference between the folded and unfolded state may be determined. In pseudoknot unfoldings, the transition may not be reversible and some energy will be dissipated as heat [17], but the Gibbs free energy change can still be determined from the distribution of unfolding and refolding energies using either the Crooks’ fluctuation theorem [26] or Jarzynski’s equality [27]. The energies are found from the areas under the clear transitions that occur on the graph of force applied versus molecular extension. The energy difference again can tell us about the size of the barrier

that the ribosome must overcome to open the pseudoknot and continue translation. These concepts will be explored in much more detail in Chapter 3.4 and Chapter 3.3.

1.4 Focus of investigation and hypotheses

As outlined in this chapter, optical tweezers have in this thesis been used to characterize two RNA pseudoknots. The aim was to find the distributions of unfolding and refolding forces, the structural brittleness and the Gibbs free energy of pseudoknot formation from a single strand. Through these data the goal was to learn how the frameshifting and stalling effect that these pseudoknots have on ribosomes may be correlated to the structure, mechanical strength, and kinetic characteristics of the pseudoknots.

The main hypotheses are:

- That the pseudoknot will transition from closed to open state during elongation of the molecular construct and back again from open to closed state during relaxation
- That these transitions can be used to quantify the mechanical force needed to open the pseudoknot, the Gibbs free energy of the transition and the rate of opening of the pseudoknot at zero force.
- That the transition from open to closed pseudoknot occurs either in one step corresponding to complete dissociation or formation of the pseudoknot or in two steps corresponding to the unfolding/refolding of first one stem, then the other.
- That the mechanical strength of the pseudoknots will be correlated to their frameshift efficiency as suggested by other studies [17, 21].

1.5 Structure of the report

The coming chapters will describe the optical trapping and molecular transition theory used in this work as well as the experimental methods and results. The results will be discussed and finally conclusions will be drawn and suggestions for improvements and further work presented. In Chapter 2, the principles behind optical trapping and calibration are outlined. Chapter 3 provides other necessary theoretical background including a description of the molecular transitions expected in the single molecule experiments and explanation of the theory of polymer stretching, energy landscapes, and kinetic and thermodynamic principles used later on. Chapter 4 covers the experimental procedures in detail and also briefly covers the data processing routines. Chapter 5 presents the molecular transition data that was measured. Chapter 6 discusses some initial observations made from the data and Chapter 7 presents further results from data treatment, namely the calculations of transition energies and kinetic parameters. Chapter 8 discusses the overall picture shown by the data and Chapter 9, the conclusion, gathers all the information in a final synthesis.

2 Optical trapping - theory and calibration

This chapter introduces optical trapping theory and optical trap calibration. It will also describe the optical trapping setup used in this work and some of the particularities encountered in working with the setup, in particular those encountered during calibration.

2.1 Optical trapping theory

An optical trap captures objects of micrometer scale such as cells or tiny manufactured beads through a balance between the scattering and gradient forces exerted by a highly focused laser beam. Essentially, variation in the electric field of the laser light causes a flux of energy across the object which gives rise to force. Reflection from the object's surface creates a scattering force which pushes the object in the direction of the light's propagation. Refraction of the light creates a gradient force which pulls the object towards the light's focus.

A full analysis of the balancing of the gradient and scattering forces requires complicated calculations using Maxwell's equations which take into account the exact shape and size of the object [1]. However, we can gain some understanding of the trapping from the optical theory for objects much smaller or larger than the wavelength of the light (the Rayleigh and Mie regimes respectively). We will look briefly at both explanations with inspiration from the approach and figures in [1].

2.1.0.1 The Rayleigh regime

The Rayleigh regime applies to objects much smaller than the wavelength of light, such as a sphere with a radius of 20 nm or less captured by an infrared laser with $\lambda = 1064$ nm. The laser light induces a dipole in the object, $\vec{d} = \alpha\vec{E}$, with potential energy $U = -\vec{d} \cdot \vec{E}$. A gradient in the intensity of the light will create a gradient in the electric field and thus in the potential energy. This energy gradient creates a force which acts on the object:

$$\vec{F}_{grad} = -\nabla U = -\nabla(\vec{d}\vec{E}) = \alpha\nabla E^2$$

Since the square of the electric field is proportional to the intensity, \vec{F}_{grad} grows with the gradient of the intensity of the field and points to the area of highest intensity. Thus if the gradient is sharp the object is pulled towards the focus of the laser.

2.1.0.2 The Mie regime

The Mie regime applies to objects much larger than the wavelength of light - tens if not hundreds of micrometers in diameter compared to a laser in the visible spectrum ($\lambda = 400 - 700$ nm). Here, classical ray optics apply and a spherical object can be treated as a lens that reflects and refracts light according to Snell's law. The total momentum

change of the light due to its interaction with the sphere is paralleled by an opposite momentum change of the sphere towards the focus of the laser. To simplify the optical analysis, the forces due to the momentum change arising from refraction and reflection are shown separately in Figures 2.1 and 2.2 respectively.

Note that the force towards the focus is far stronger between the source of the light and the focus than beyond the focus away from the light source. This is because the reflective force always pushes the bead away from the light source while the refractive force always pulls it towards the focus, as may be seen in Figures 2.1 and 2.2. If the bead strays too far from the focus away from the light source, the scattering force from reflection dominates and the bead is pushed away. This was seen practice in this thesis where polystyrene beads with diameters of a few micrometers readily flew off through the sample chamber if displaced just a little from the focus in the direction away from the light source.

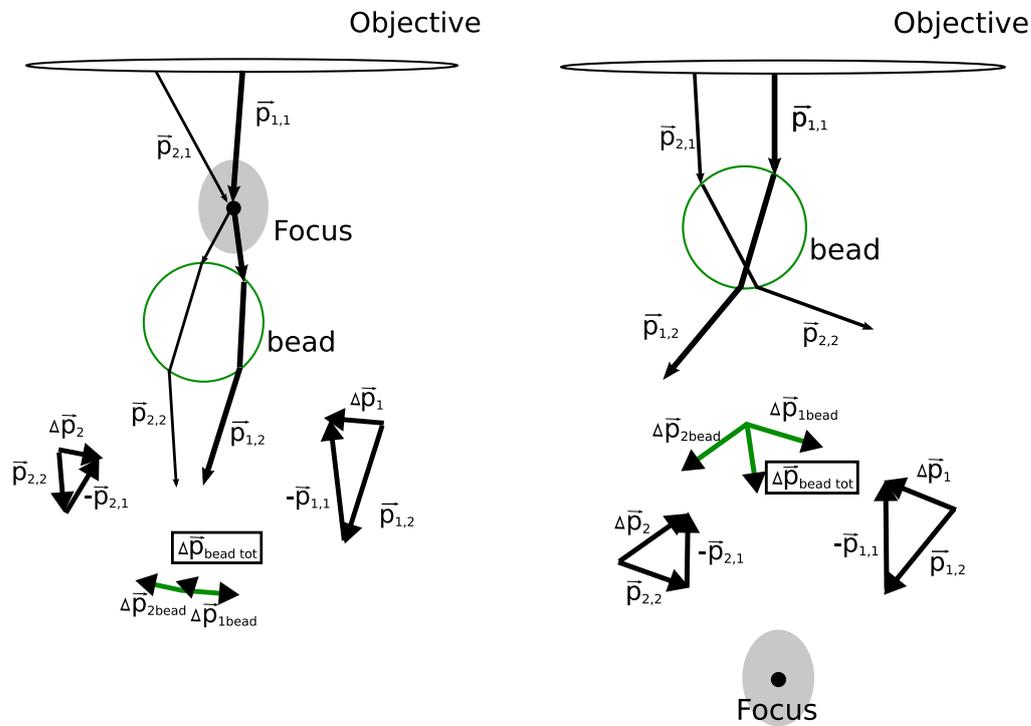


FIGURE 2.1: Ray optics explanation of laser trapping: Refractive (scattering) forces pull the bead towards the laser focus. Black arrows emerging from the objective and encountering the bead indicate the path of the laser light. Wider arrows indicate greater light intensity. At the bottom of the images, the momentum changes of the light and the bead are shown by vector addition. Black vectors represent light momentum; their length reflects the relative light intensity and thus the magnitude of the momentum change. Green vectors with black triangular heads represent the opposing change of momentum of the bead. Left: The bead is below the objective focus and scattering forces pull it up towards the focus, even though the example rays shown here balance out to create only a rather weak pull towards the focus. Right: The bead is above the focus and scattering forces push it down towards the focus.

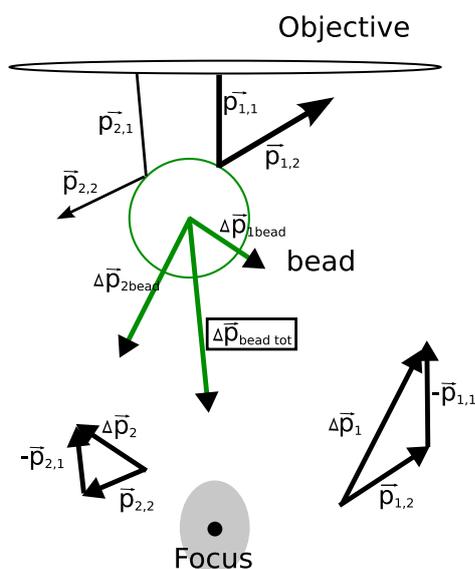


FIGURE 2.2: Ray optics explanation of laser trapping: Reflective forces push the bead away from the source of the light. Black arrows emerging from the objective and encountering the bead indicate the path of the laser light. Wider arrows indicate greater light intensity. At the bottom of the images, the momentum changes of the light and the bead are shown by vector addition. Black vectors represent light momentum; their length reflects the relative light intensity and thus the magnitude of the momentum change. Green vectors with black triangular heads represent the opposing change of momentum of the bead.

2.2 Optical trap setup

Usually the setup for optical trapping is an inverted microscope with a built-in laser. The laser light passes from the bottom of the microscope up through the objective and the sample chamber and is collected above the condenser by a quadrant photo diode (QPD), which is used to detect the position of the object in the trap. The specimen can be observed through the objective with a camera (or through the microscope eye-piece if the laser is off). Often the laser may be steered with some independence from the movement of the objective and sometimes the beam is split to provide multiple traps.

In this thesis, a double trap prototype commercial instrument called a NanoTracker from JPK Instruments was used for pulling experiments. The traps are made by a 1064 nm laser split by a beamsplitter. Each laser trap can be controlled by a galvanic mirror. The layout of the dual trap NanoTracker setup is shown in Figure 2.3. Some particularities of this system will be described below in Chapter 4.4.2. The experimental routine employed to capture beads and measure molecular transitions will be described in Chapter 4.4

The detection unit consists of four QPDs, two for each trap, whose centers are aligned to optimally capture changes in light intensity and interference patterns of the trapping laser beams. Each of the four quadrants in the QPDs generates a voltage proportional to the amount of light impinging on it. When an object is caught in the laser trap it casts an interference pattern onto the QPD which is symmetrical if the object is spherical and the laser and QPD are properly aligned. From the interference pattern, the movement of the bead in the x , y and z direction can be approximated by simple sums and differences between the voltage outputs of the four diodes. However, "crosstalk" (contamination of the

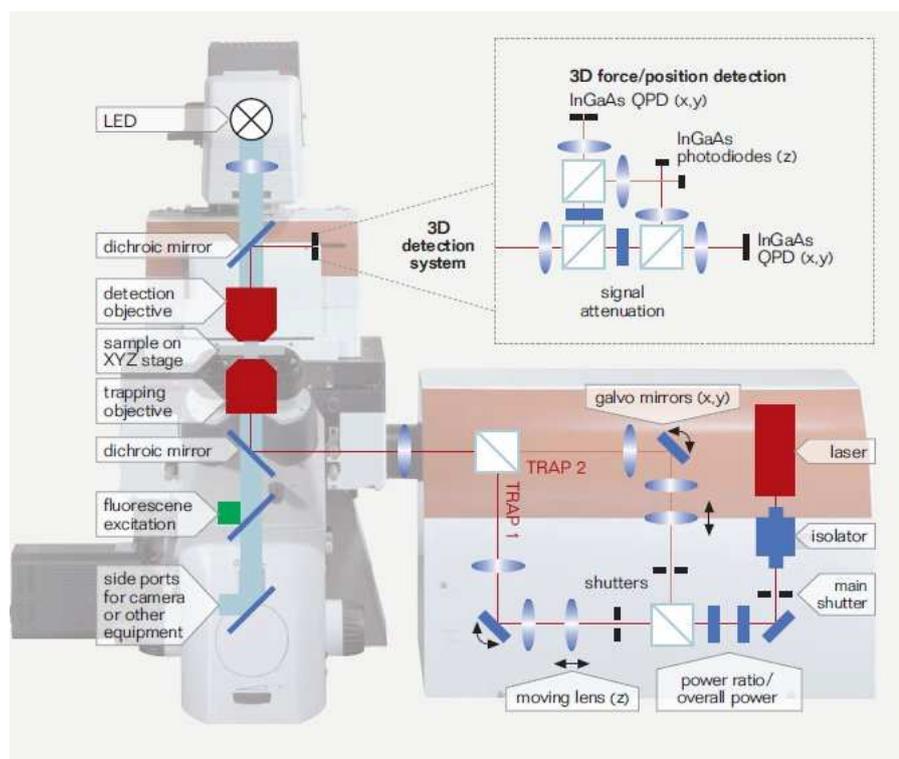


FIGURE 2.3: NanoTracker setup showing the system for detailed detection of trap movement by photodiodes and the steering system for the traps (right). Beads are placed in a chamber in the inverted microscope (left) and viewed with the camera or through the eyepieces. Figure from [28].

signal for the movement in one direction or another) can occur and is seen, e.g., as a signal for movement in the x -direction though movement should only occur in the y -direction. Crosstalk may occur if the laser intensity profile is not completely symmetrical or if the quadrants are not equally sensitive to light [29]. In the case of a dual laser trapping system, it can also occur if laser light from one trap is detected by the QPDs for the other. The latter appears from experience with the NanoTracker to be highly likely if the two traps are operated close to each other.

There are two QPDs per trap for optimal measurement of both axial (z -direction) and lateral (x - y -direction) displacement of the beads. This is necessary because the optimal angle for detecting the laser beam as it exits the condenser is different for measuring axial and lateral displacement [30].

2.3 Calibration

To translate the voltage output of the QPD into a measure of both the distance moved by a spherical particle in the trap and of the force exerted on it, it is necessary to calibrate the trap. The alternative would be to measure displacement directly and to calculate the force exerted on the particle at each position theoretically, but this is in practice more or less impossible. In the following, the calibration theory used in this work will be described.

The most important assumption underlying the calibration and subsequent extraction of data for force on and movement of the particle is that the optical trap creates a harmonic

potential equivalent to that of a Newtonian spring. Under this assumption, the force on the particle due to the trap is:

$$F_{trap} = -\kappa x_{trap} \quad (2.1)$$

where x_{trap} is the displacement of the particle from the trap center and κ is the stiffness of the trap. This relationship is valid up to a certain distance from the trap center, the linear regime, which is limited by the QPD and whose size depends additionally on the particle diameter, the strength of the trap, and the direction in which movement is measured. For the distances involved in the experiments in this work, the linear regime was rarely exceeded. The linear relationship between force and distance close to the trap center was originally found experimentally (in [31] and others cited in [32]) but is also predicted to some extent by the theory of ray optics [33].

The conversion factors, α and β , that convert the QPD signal, X_{QPD} , which is measured in volts, into force and distance will be defined as follows:

$$F_{trap} = -\alpha X_{QPD} \quad (2.2)$$

$$x_{trap} = \beta X_{QPD} \quad (2.3)$$

so that $\kappa = \frac{\alpha}{\beta}$. Note that the trap stiffness must be found for the particular direction in which force and movement is to be measured. It is usually much stronger in the lateral plane perpendicular to the laser than in the axial direction of laser propagation. Due to polarization of the laser it is also somewhat asymmetric in the lateral plane. This is certainly the case for the NanoTracker traps.

2.3.1 Stokes calibration of force

To find the conversion factor, α , between voltage output, X_{QPD} , and force, a Stokes calibration may be performed. Stokes' law states that the force on a small spherical particle being moved in a viscous media far from any surfaces is a constant given by:

$$F_{drag} = -\gamma v$$

where v is the velocity that the media is moved with and γ is the drag coefficient given by $\gamma = 6\pi\eta r$, r is the diameter of the sphere, and η is the dynamical viscosity of the media. "Far" from the surfaces in our case meant a distance of at least about 10 μm .

During the Stokes calibrations carried out in this thesis, the measurement chamber was moved at constant velocity relative to the trap and the displacement of the bead from the trap center was measured. The theoretical drag force was calculated on the assumption that the viscosity of the media (η) was that of water at room temperature and that the radius of the sphere was correctly given by the manufacturer. Assuming also that the voltage output of the QPD mostly arises while the bead is in constant motion (constant v), the conversion factor was calculated from:

$$\langle F_{drag} \rangle = -\langle F_{trap} \rangle = -\alpha \langle X_{QPD} \rangle = -\gamma v \quad \implies \quad \alpha = \frac{\gamma v}{\langle X_{QPD} \rangle}. \quad (2.4)$$

Data for a Stokes calibration made with the NanoTracker is shown in Figure 2.4. The sample chamber was moved back and forth a number of times with a piezoelectric stage that could be controlled with high accuracy. The smoothed data was fitted to a double Gaussian distribution (Figure 2.4(c)) and the average displacement, $\langle X_{QPD} \rangle$, from the central value was found. The calibration was performed for each set of beads at four different speeds and since $\langle X_{QPD} \rangle = \frac{\gamma v}{\alpha}$, α was found from the slope of the plot of $\langle X_{QPD} \rangle$ vs γv as shown in Figure 2.4(d).

Since the Stokes calibration only gives a value for the conversion from voltage to force, another calibration is needed to convert voltage to distance. Two methods, namely Brownian motion calibration and power spectrum calibration, will be explained in the following sections.

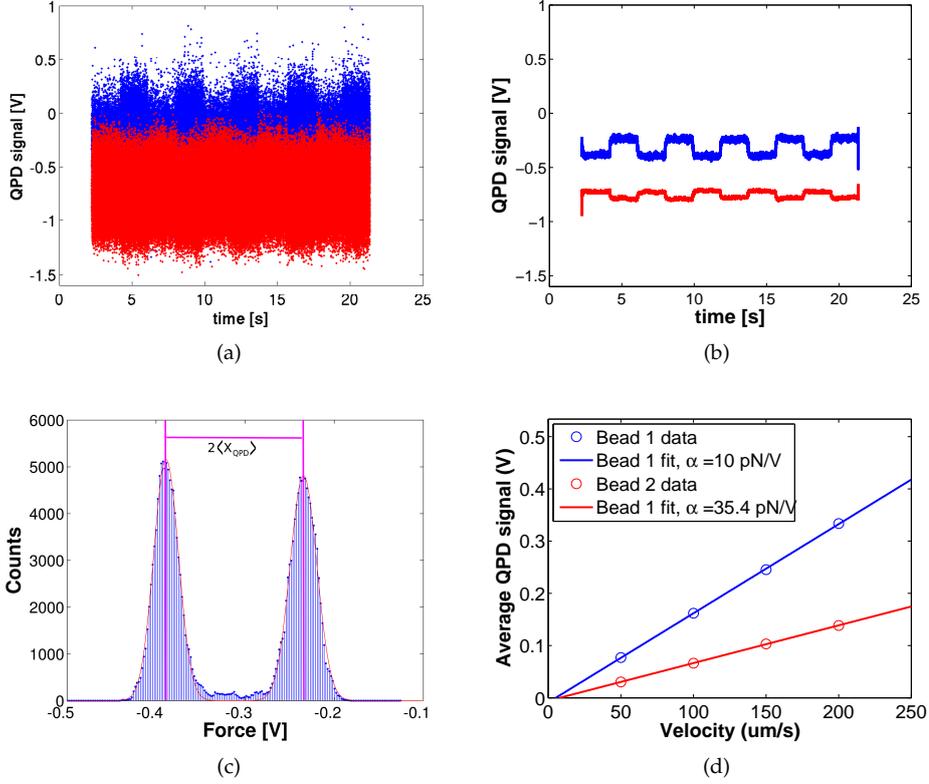


FIGURE 2.4: Stokes calibration: (a) raw data collected at 10 kHz during a particular Stokes Calibration - the piezo here moves $95 \mu\text{m}$ at $50 \mu\text{m/s}$ in each direction. (b) Data smoothed by a moving average window of 201 datapoints. (c) Data for bead 1 in (b) is fitted to a double Gaussian distribution (the same is done for the data for bead 2, not shown). (d) Fits to Equation (2.4) for each bead. The distance between the peaks of the Gaussian fits to the position data is halved and used as the value of $\langle X_{QPD} \rangle$ for the corresponding piezo speed.

2.3.2 Brownian motion calibration

The motion of the particle in the trap is constrained Brownian motion. Maintaining the assumption of a harmonic potential, the particle will display a Gaussian distribution of positions around the center of the trap. If the noise and drift of the trap is negligible, the variance of the positions will be determined by [1]:

$$\sigma_x^2 = \langle x^2 \rangle = \frac{k_B T}{\kappa_x}$$

where σ_x^2 is the variance of the position measurements, k_B is Boltzmann's constant and T is the temperature in Kelvin.

This equality derives from the equipartition theorem,

$$\frac{1}{2} m \langle v^2 \rangle = \frac{1}{2} k_B T$$

because the average potential energy of the particle in the trap given by $U = \frac{1}{2} \kappa_x \langle x^2 \rangle$ is equal to the average kinetic energy, $\frac{1}{2} m \langle v^2 \rangle$.

Due to this relationship, measuring the width of the distribution of positions in volts as the particle moves randomly in the trap makes it theoretically possible to find the relationship between the spring constant κ and the conversion factor β between QPD signal and distance. Since $\kappa = \frac{\alpha}{\beta}$ it is possible to find β from the distribution of position measurements if α is known:

$$\sigma_{QPD}^2 = \frac{\langle x^2 \rangle}{\beta^2} = \frac{k_B T}{\kappa_x \beta^2} \iff \beta = \frac{k_B T}{\alpha \sigma_{QPD}^2}$$

Unfortunately, any noise in the system will add to the motion of the particle beyond its intrinsic thermal motion, so that the equipartition theorem no longer describes the full distribution of the particle's positions. This is the case for the NanoTracker, in which the noise of the system adds significantly to the random motion of the particle and the width of the distribution of particle positions is much too wide to be used for calibration. See Figure 2.5 where the real position distribution for a bead in each trap is compared to that which would be predicted just based on thermal motions using values of β found with power spectrum calibration (next section, Chapter 2.3.3). For this experiment, the real distribution of positions is three times as wide as expected for trap 1 and seven times as wide as expected for trap 2. This means that the magnitude of the total noise experienced by a molecule attached to both beads is on the order of the expected unfolding length of the pseudoknots. The calculations made to quantify this level of noise are presented in Appendix A.3.1.

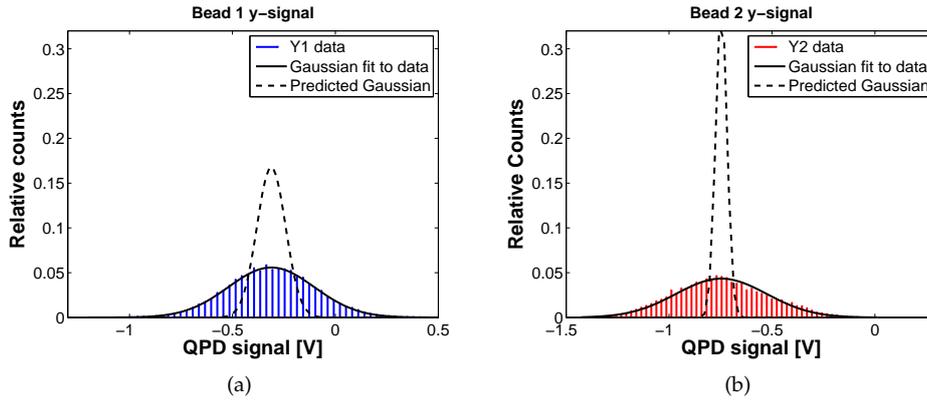


FIGURE 2.5: Histograms of bead positions for (a) Trap 1 ($r_{bead} \approx 1\mu\text{m}$) and (b) Trap 2 ($r_{bead} \approx 1.5\mu\text{m}$) with Gaussian fits to position distribution (full lines) and predicted Gaussian distribution based on values for β found using power spectra (dashed lines) (see Chapter 2.3.3).

2.3.3 Power spectrum calibration

An alternative approach often used to calibrate for both force and distance is through the plotting of a power spectrum of the motion of the particle in the trap. A power spectrum is the square of the Fourier transform of a series of position measurements. It has a highly characteristic shape for a particle trapped by a harmonic potential. Fitting to the power spectrum ideally yields calibration factors for both force and distance as will be shown below. However, in the case of the NanoTracker, noise in the most important frequency window made it impossible to use the power spectrum to calibrate fully, and it was used only to calibrate for distance. An ideal (simulated) power spectrum is shown in Figure 2.6 together with a power spectrum measured by the NanoTracker. Note that the axes are

logarithmic and the noise peak at the center of the spectrum therefore delivers more than 10 times more signal than Brownian motion power spectrum itself.

The reason the noise is so problematic is that the power spectrum force calibration is made using a quantity called the corner frequency, $f_c = \frac{k}{2\pi\gamma}$ which is found by fitting the theoretical power spectrum to the experimental one. The value of f_c corresponds to the frequency at the bend in the curve shown in Figure 2.6(a). Due to the noise in the NanoTracker power spectrum at exactly these frequencies, the fit here is not accurate and the value of f_c cannot be determined precisely.

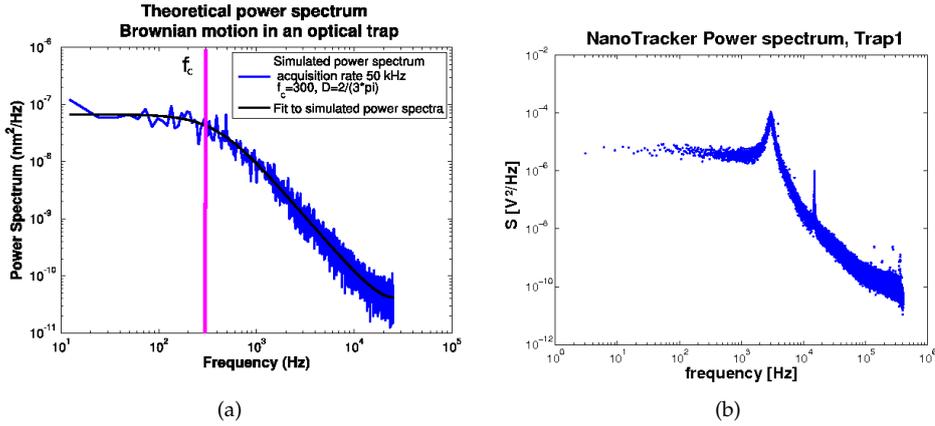


FIGURE 2.6: Power spectra: (a) Theoretical power spectrum for Brownian motion in an optical trap; (b) power spectrum measured by the Nanotracker. Notice the high noise in NanoTracker power spectrum at the frequencies just around f_c . The theoretical power spectrum was generated by an algorithm originally made by Line Skotte and Henrik Flyvbjerg.

Numerous descriptions of the construction of the power spectrum and subsequent interpretation exist in the literature (e.g. [34], [35], [36], [29]). Here I will mainly follow the approach of Gittes and Schmidt [34], who explain very clearly how the discrete power spectrum (based on numerical data) is calculated from the Fourier transform of a series of measurements of bead positions in the trap over a period of time. As noted by these authors, the Fourier transform of a data series contains exactly the same amount of information as the original data, it just presents the information differently. The power spectrum, on the other hand, is derived from the Fourier transform with loss of half of the information but a significant gain in usefulness. Essentially, Gittes and Schmidt explain, the power spectrum can be seen as a measure of the intensity of the signal at a given frequency. They go on to define the one-sided power spectrum - which exists only for positive frequencies - for a discrete data series as is needed in practice. The one-sided power spectral density for a signal that is continuous in time is simpler to define, as is done in [37]:

$$S_x(f) = |\tilde{x}(f)|^2 + |\tilde{x}(-f)|^2 = 2 |\tilde{x}(f)|^2$$

Note that $S_x(f)$ is defined only for $f \geq 0$ (which seems to make physical sense when we are dealing with a time series of distance measurements) and the last equality holds only for functions $x(t)$ that are real. This is the power spectrum definition that is used for calibrating optical traps in [34] as well as [35] and [36]. The two-sided power spectrum can be used as well as long as the treatment is internally consistent.

The theoretical expectation value of the one-sided power spectrum for an optical trap is derived in Appendix A following [34] and [38]. It is:

$$S_x(f) = \frac{k_B T}{\gamma \pi^2 (f_c^2 + f^2)} \quad (2.5)$$

where $f_c = \frac{\kappa}{2\pi\gamma}$ as mentioned is the corner frequency whose value corresponds to the frequency at the bend in the curve shown in Figure 2.6(a). The shape of the power spectrum curve is sometimes referred to as a Lorentzian because it has the general form $f(x) = a/(b^2 + x^2)$.

The power spectrum for the distance X_{QPD} measured in volts is related to the power spectrum for the distance in meters $x(t)$ by $S_x(f) = \beta^2 S_v(f)$. Thus least squares fitting to a measured power spectrum yields both the calibration factor β from volts to distance and, theoretically (though in not in practice for the NanoTracker data), the calibration factor α from volts to force. The latter is found through the fitted value for f_c , which reveals the trap stiffness, $\kappa = \alpha/\beta$.

2.3.4 Fitting to the power spectrum

As noted in above in Chapter 2.3.3, the corner frequency, f_c , which is usually used for calibration, could not be reliably found for the Nanotracker due to noise at about 1000 Hz, just about where f_c was located. However, fitting the experimental power spectrum measured by the Nanotracker to its theoretical expectation value still yielded a fairly good estimate for the conversion factor β^2 for converting voltage to distance (Equation 2.3). Three fitting methods were tested for this purpose and one was selected as the one that most consistently yielded a reasonable fit. This was a routine that performed an unweighted fit of the low and high frequency portions of the experimental power spectrum to the theoretical one. It was implemented in the programme IgorPro by Jesper Tholstrup, as despite the efforts of this author neither Gnuplot nor Matlab could be coaxed into performing this type of fit with any stability. The two other fitting methods investigated by this author were:

- Weighted least squares fitting only to the region of high frequency, $f \gg f_c$, where $S_x(f) \approx \frac{k_B T}{\gamma \pi^2 f^2}$ performed in Gnuplot.
- Weighted least squares fitting of the low and high frequency regions of the data to the full equation for the power spectrum, excluding the noisiest regions from the fit, also performed in Gnuplot.

The three methods yielded slightly different values for β , differing by up to about 30 %. Sample fits made to the same experimental power spectrum using the three different methods are shown in Appendix A, Figure A.1 (a), (c) and (e). The accuracy of the calibration factors yielded were tested as follows: Two beads were caught in the traps and brought close enough to stick together by what appeared to be an inelastic polystyrene connection (method suggested by Jesper Tholstrup). One trap was held immobile while the other was moved away at constant speed. Since in this scenario the distance between the beads is expected to be constant over time though the force is increasing linearly, a plot of force versus tether extension should yield a vertical line if the traps have been correctly calibrated. This is because change in tether extension (i.e., change in bead-bead distance) is calculated as the distance between the traps minus the displacement of the beads from the trap centers. If the bead-bead distance is constant while the traps are moved, the change of displacement of the beads from the trap centres, which is calculated using the calibration factors, should exactly cancel the known movement of the traps.

In fact it turned out that none of the calibration factors yielded a completely straight line in the test plots, as may be seen in in Appendix A, Figures A.1 (b), (d) and (f). In general, the

unweighted least squares fit to the data in the least noisy regions performed best, so this was the routine which was subsequently used in data analysis. The ranges of data points fitted were 50-150 Hz and 20-55 kHz.

It was a surprise to this author that the most accurate calibration method turned out to be an unweighted least squares fit, since a weighted least squares fit seems most statistically correct (as noted in the detailed treatment of power spectrum fitting in [29], the variance of $S_x(f)$ is simply equal to $S_x(f)$ since the values of $S_x(f)$ are exponentially distributed about their expected value). The reason the unweighted fit is best may be as follows: The weighted fitting method only works well if the data in the regions of relatively low noise truly correspond to the ideal power spectrum for the optical trap. If we examine the fit to the theoretical powerspectrum, $S_v(f) = \frac{1}{\beta} \frac{k_B T}{\gamma \pi^2 f^2}$, in the region of high f in Figure A.1(a), we see that the slope of the data in this log-log plot is actually steeper than the line fit to it. Thus contrary to what might be expected, namely that due to aliasing $S_x(f)$ would be too large in the region of high frequency (see, e.g., [29], [34]), there is a loss of intensity at high frequency. This intensity loss most likely derives from loss in the photo diodes due to diffusion of electrons between the silicon wafers as explained in [39], even though the NanoTracker photodiodes are made of a type (InGaAs) that should minimize this loss. The loss of data in the high frequency region makes this area unreliable for fitting. It therefore makes intuitive sense that the unweighted fit is better, as it effectively gives more weight to the more reliable low frequency data than to the high frequency data.

2.3.5 Sources of calibration error

Apart from the uncertainty in fitting to the power spectrum (which is not an ideal Lorentzian as it should be) or to the Gaussian position distributions for the Stokes calibration (which are not ideal Gaussians), error in calibration arises from the values used for:

- viscosity η
- temperature T
- bead radius r

which all influence the force calibration more than the distance calibration, since error in α is directly proportional to error in ν and r while error in β is proportional to the square root of the error in ν , r , and T .

Bead radius differs significantly from experiment to experiment. Apart from difficulty fitting to the power spectrum, this is probably the most significant source of error. The manufacturer provides the mean value and for the smaller beads also a measure of the standard deviation. The larger beads varied very much. Though visual evaluation was used to limit variation in size of the larger beads used for experiments, variation on the order of about 5 % of the larger beads' radii could probably not be avoided. See examples of bead variation in Figure 2.7.

As noted, the calibration will also contain some systematic error from approximating the dynamical viscosity, which depends on temperature as well as on exact media composition. The variation due to the uncertainty in temperature itself is not large, but the uncertainty in the viscosity due to temperature variation is noticeable - η is about 10 % lower for water at 25 °C than at 20 °C. The temperature at the focus of the optical trap will be higher than laboratory temperature; how much depends on the intensity of the laser at the trap, which is not easy to measure. In [40] the temperature difference between laboratory and optical trap is measured to about 8 °C/W around silica beads held in water by a 1064 nm laser. Using a nominal laser power of about 1 W per trap during experiments, we do not expect that the power delivered at the traps in the sample chamber after loss in the optical path will be much more than 100 mW. Thus the heating is not expected to exceed 1 °C above

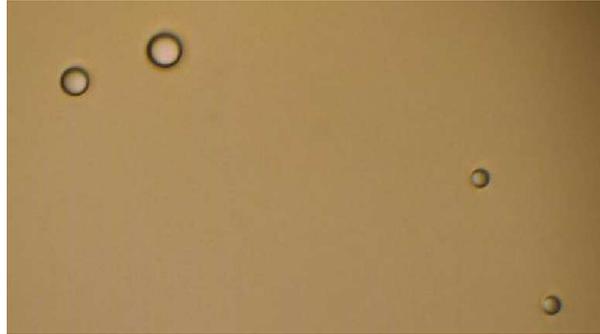


FIGURE 2.7: Size variation of beads. The two large beads at top left of image should nominally both have a radius of $1.5 \mu\text{m}$. The two smaller beads have $1 \mu\text{m}$ radii. The beads are stuck on the glass surface at the bottom of the sample chamber; photograph taken from below.

laboratory temperature. The viscosity of the buffer used in experiments was measured by Jesper Tholstrup and found to be very similar to that of water. The value used is therefore the viscosity determined for water at 25°C , easily found in literature as $\eta = 0.89$ (e.g. [41]).

3 Theoretical background - molecular transitions

This chapter will introduce the molecular transition investigated in greater detail along with a brief theory of polymer stretching and some kinetics and thermodynamics relevant for single molecule experiments.

3.1 Pseudoknot opening - the molecular transition investigated

As described in the introduction, this thesis investigates the opening and closing of an RNA pseudoknot as it is pulled and relaxed by optical traps. The basic experimental layout was shown already in Figure 1.2: a molecular construct containing the RNA pseudoknot is attached to polystyrene beads which are held and manipulated by two laser traps. A full description of the experimental conditions will follow in Methods and Materials, Chapter 4, but to give the theory presented in this chapter some context, the expected transition is shown in Figure 3.1

In short, the transition envisioned is that a molecular construct including a folded RNA pseudoknot structure is pulled so that the pseudoknot opens up into a single strand of RNA. The construct may then be pulled even further before being relaxed so that the single strand may refold into the pseudoknot.

Using the optical trap, the force on the beads holding the molecule and the change in extension of the molecule are continuously measured. From these measurements, curves of force versus extension may be drawn for each pulling cycle. A theoretical curve of force versus extension is shown in Figure 3.2. Note the abrupt changes (rips and zips) on the force-extension curve, which are expected when the pseudoknot structure unfolds into the single strand at A or refolds from the single strand at C. As will be explained below, the single stranded RNA is expected to be much more flexible than the folded pseudoknot structure, and the force-extension curve therefore rises more steeply from point D to A, where the pseudoknot is assumed to be folded, than from point C to B, where a long section of the molecular construct is expected to occur as single stranded RNA.

3.2 Polymer stretching theory

Several models have been developed to describe the stretching of molecular polymers such as DNA and RNA. One of the most commonly used is the Worm Like Chain (WLC) model [42], which describes the stretching of DNA by envisioning it as a continually flexible cord (like an electric cord or a rope, the molecule can according to the model be rearranged more or less flexibly depending on how stiff it is, but it cannot be lengthened elastically). This model considers only entropic changes in the molecule as it is stretched but ignores enthalpic changes.

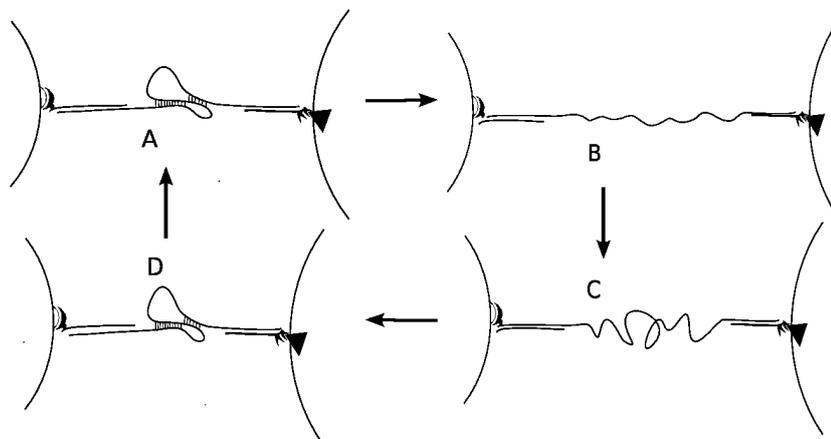


FIGURE 3.1: Cycle of pseudoknot unfolding/refolding expected for the pseudoknot-containing RNA strand as the beads are moved apart and together again (open circle sections on either side of the pseudoknot strands are meant to show the bead attachments). At A, the beads are fairly close together and the pseudoknot is folded. At B, the pseudoknot has opened up into the single strand and the beads are able to move much further apart. Subsequently the molecular construct is relaxed back down to the conformation at C before the pseudoknot structure is formed again and the structure passes into conformation D. From conformation D the beads may again be moved apart, elongating the molecule back to conformation A and the cycle can be repeated.

The WLC is characterized by the persistence length, L_p , of the polymer, which is a measure of how “bendy” it is with respect to thermal fluctuations (a telephone cord is much more flexible than the electrical cord for a fridge when both are kept at the same temperature, for instance). The model predicts the extension, x , of the polymer as a fraction of its total contour length, L_c , when a given force, F , is applied to it. The model equation is:

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

Essentially, the WLC accounts for the decrease in entropy as an initially completely randomly arranged polymer is straightened out by force allowing less and less random polymer bending and curving. It agrees well with experimental stretching of double stranded DNA up to a force of about 10 pN according to [17]. An expanded version of the WLC model called the Extensible Worm Like Model or EWLC adds a parameter, the elastic modulus, K_0 , to account for the intrinsic ability of the polymer to stretch. This change is enthalpic rather than entropic; it has to do with the conformation of the chemical bonds of the polymer and releases or absorbs heat (when you stretch a rubber band it actually feels a little cooler!). The EWLC was originally applied to measurements of the stretching of double stranded DNA by Wang et al. [11]. It is an implicit equation, which reads as follows:

$$F = \left(\frac{k_B T}{L_p} \right) \left[\frac{1}{4(1 - x/L_c + F/K_0)^2} - \frac{1}{4} + \frac{x}{L_c} - \frac{F}{K_0} \right]. \quad (3.2)$$

L_p has been found experimentally to vary with ionic buffer strength from about 40–50 nm for double stranded DNA while K_0 lay around 1.0–1.3 nN with no clear trend in value

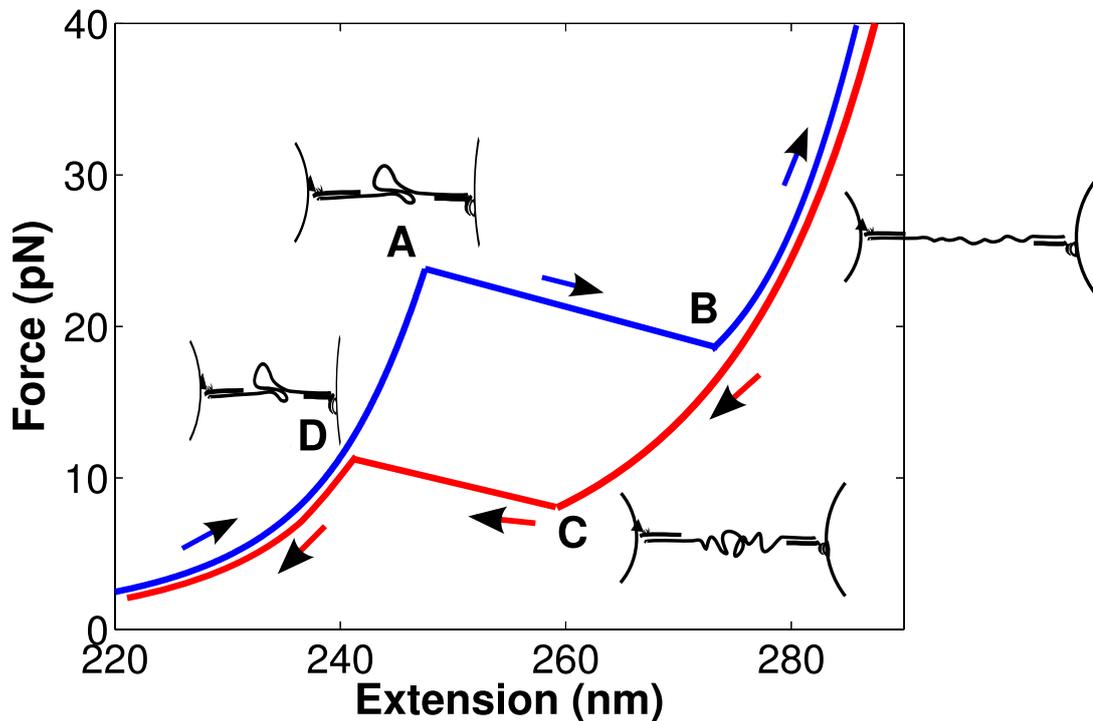


FIGURE 3.2: Expected curve of force versus extension for the simple cycle of molecular transitions described in Figure 3.1. The letters correspond to the molecular conformations shown in that figure. The pseudoknot is closed during the extension from D to A and open during the relaxation from B to C. The transition from A to B is called a “rip” and the transition from C to D is called a “zip”.

depending on buffer [11]. Both the EWLC and the WLC model assumes $L_c \gg L_p$. It must be kept in mind that the EWLC does not necessarily hold in the case of the molecular constructs investigated in this thesis, since they have $L_c \sim L_p$ rather than $L_c \gg L_p$.

The EWLC predicts the stretching of double stranded DNA well up to about 30 pN [43, 44]. Above 30 pN the curve starts to deflect away from the EWLC (its curvature is reversed) and at about 60 pN the stretching curve of force versus extension plateaus due to force-induced melting of the double strand [45]. To predict the exact shape of the force-extension curve above 30 pN it is therefore necessary to take into account the twisting and melting of the DNA under tension as described by [44].

The WLC has been used to estimate the stretching of the molecular constructs used in optical tweezer investigations of RNA hairpins [46, 4], while in investigations of RNA pseudoknots both the WLC [25] and EWLC [17] have been used. These investigations either treat the stretching of the DNA-RNA hybrid handle as negligible and ignore it [4, 17] or they simply treat the the single strand of RNA and double stranded RNA-DNA hybrid as separate entities with each their own L_p [25, 46, 47] and add their contributions to get the total curve.

Figure 3.3 shows the theoretical EWLC force-extension curve calculated for the expected stretch of the folded and unfolded PK 11/6 pseudoknot structure. In the figure, the predicted curve is plotted together with a measured force-extension curve which displays a transition between a longer and a shorter conformation. The predicted curves are plotted

with two different values of L_p for the double stranded (RNA-DNA hybrid) part of the construct, since $L_p = 10$ nm is the value used in other RNA studies [46, 47] while $L_p = 43$ nm is a value found for double stranded DNA in [11]. It is difficult to tell which value is better based on the curves shown here or on other curves examined - perhaps a value in between the two extremes would fit the data best. Note that the experimental force-extension curve has an arbitrary zero point for extension as the initial separation of the beads was not known. The predicted curve has been offset from zero by minus 85 nm to match the experimental curve.

In the EWLC prediction, the value used for K_0 is 1 nN and the persistence length L_p for single stranded RNA is set to 1 nm as in [17]. The precise method used to make the predicted EWLC curve will be described in Chapter 4.5.1.

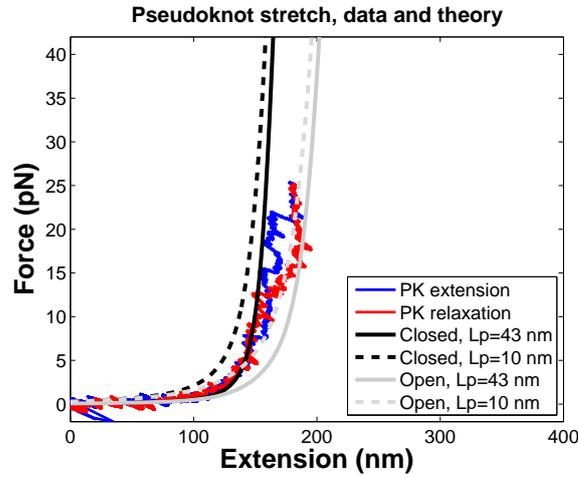


FIGURE 3.3: EWLC prediction of force versus extension for the PK 11/6 pseudoknot constructs investigated in this thesis. Shorter traces (black): closed pseudoknot. Longer traces (gray): open pseudoknot. The fully drawn traces use $L_p = 43$ nm for the DNA-RNA hybrid handles which is the persistence length calculated for double stranded DNA in Wang et al. [11]. The dashed traces use $L_p = 10$ nm as in Collin et al. [46]. Note that the predicted curve has been offset from zero by minus 85 nm to match the experimental curve. The initial separation of the beads in the experimental setting is not known and the extension scale therefore has an arbitrary zero.

3.3 Kinetics

From the distribution of forces at which the pseudoknot opens and closes, information may be extracted about the rate of opening at different forces, $k(F)$. The rate of opening at zero force, k_0 , is especially of interest. Finding these rates is not easy when the transition is as complicated as the one involving a pseudoknot, since several intermediate states and thus several rates of reaction are most likely involved. Nonetheless, if some simplifications are accepted, several models exist for finding $k(F)$ from the distribution of unfolding forces. The simplest is based on an expression called Bell's formula as cited in, e.g., [48]:

$$k(t) = k_0 e^{\frac{F(t)x^\ddagger}{k_B T}}. \quad (3.3)$$

Here x^\ddagger is a quantity called the distance to the transition state, which is a “distance” in the energy landscape of the pseudoknot. Fx^\ddagger is the energy required for the molecule to reach the transition state. Bell’s formula is thus very similar to the Arrhenius equation,

$$k(t) = k_0 e^{\frac{\Delta E}{k_B T}}. \quad (3.4)$$

The idea of an energy landscape which contains a distance to a transition state is easiest to understand by looking at a conceptual illustration. See Figure 3.4, which is inspired by an illustration found in [49]. In Figure 3.4 (a), an energy landscape surface $U_0(x)$ at zero force varies along an arbitrary reaction coordinate and the distance to the transition state is a distance along this virtual coordinate between the closed state and the transition state. The apparent energy of activation, ΔG^\ddagger , quantifies the energy difference between the closed and transition states at zero force *along the reaction coordinate*. It is not equivalent to the actual energy of activation at zero force, $\Delta G_{\text{transition}}$, unless the molecular energy levels attained during the zero force transition actually correspond to the energy levels along the reaction coordinate shown.

In Figure 3.4 (b), the effect of adding force to the molecular system is shown. Here the reaction coordinate is the pulling coordinate in a pulling experiment and thus may be imagined as corresponding to an actual physical molecular extension. The more the construct is pulled, the more force is applied to the system and the more energy is added to it. This can be visualized as a tilting of the energy landscape that alters the relative energy of different conformational states and lowers the transition barriers between some states.

An example of a naturally occurring reaction where the transition from one molecular state to another would as in a pulling experiment be directly correlated to molecular extension is the unzipping or unwinding of an RNA hairpin by a ribosome. In such reactions, the basepairs have been observed to be opened one at a time from the end of the helix [13]. Here, ΔG^\ddagger might truly correspond roughly to $\Delta G_{\text{transition}}$ and x^\ddagger to a change in the end-to-end extension of the molecule caused by the ribosome.

Bell’s formula for molecular unfolding kinetics is based on the following assumptions:

1. That the reaction is a first order two-state transition so that the survival probability of the molecule is an exponential function of time.
2. That the quantity x^\ddagger is constant with increasing force.

Using Bell’s formula we can derive an expression for k_0 and x^\ddagger in terms of the probability that the molecule has not yet unfolded. Following [50], we assume that the reaction of the pseudoknot opening and closing is a simple two-step reaction with probability of being in the closed state P_c and probability of being in the open state $P_o = 1 - P_c$ where

$$P_c(t) \approx N_c(t)/N_{\text{tot}}$$

and $N_c(t)/N_{\text{tot}}$ is the fraction of molecules in the closed state at time t . Then the change over time in the probability of being in the closed state is governed by the differential equation:

$$\frac{dP_c(t)}{dt} = -k_f(t)P_c(t) + k_r(t)(1 - P_c(t)), \quad (3.5)$$

where $k_f(t)$ and $k_r(t)$ are the rates of the forward (opening) and reverse (closing) reactions.

Assuming that very few closing transitions take place compared to opening transitions, we simplify the expression to

$$\frac{dP_c(t)}{dt} = -k_f(t)P_c(t), \quad (3.6)$$

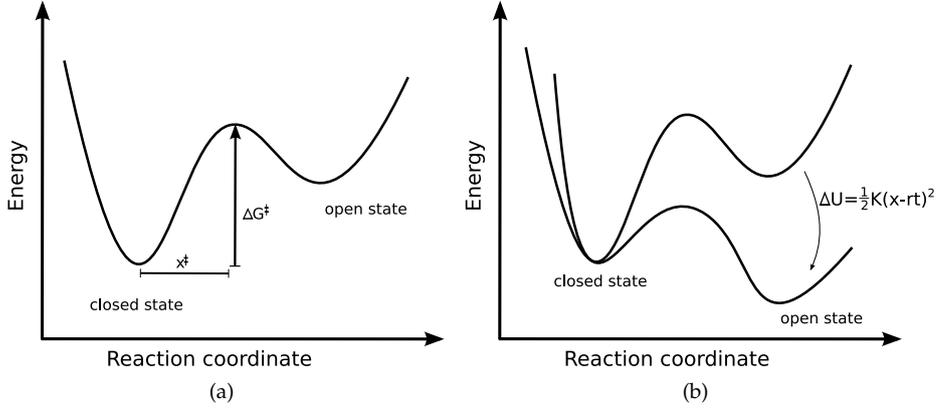


FIGURE 3.4: Energy landscapes. (a) Generic energy landscape described by the surface $U_0(x)$ for a two-state system with an open and a closed conformation separated by an energy barrier. The energy barrier is characterized by x^\ddagger , the distance from one state to the transition state along the (arbitrary) reaction coordinate, and by ΔG^\ddagger , the energy of activation along the reaction coordinate. (b) The transition curve $U_0(x)$ is tilted by $\kappa(x-rt)^2/2$ due to the application of force from the optical trap. The energy needed for the transition is reduced and the rate of transition, $k(F)$, is increased. During a pulling experiment, the reaction coordinate being varied is the molecular extension (or actually the distance between the traps). In this direction the tilt is exactly $\kappa(x-rt)^2/2$. On the actual energy landscape path along which the transition takes place, the effect of the force may be different.

which is first order decay of the number of molecules in the closed state over time.

Following the approach in [51], we then insert Bell's expression for $k_f(F)$:

$$\frac{dP_c(t)}{dt} = -k_0 e^{\frac{F(t)x^\ddagger}{k_B T}} P_c(t).$$

We substitute in $F = rt$ where F is the force applied to the system by pulling the molecule and r is the pulling velocity. Then we integrate by parts and rearrange and finally get:

$$r \ln P_c = -\frac{k_0 k_B T}{x^\ddagger} \left(e^{\frac{F(t)x^\ddagger}{k_B T}} - 1 \right). \quad (3.7)$$

Thus from the empirical probability distribution of the pseudoknot opening at a given force, we can plot $r \ln P$ against F (the forces at which molecular transitions are observed) and fit to Equation (3.7) to find k_0 and x^\ddagger . This way of finding these kinetic parameters will later in this thesis be called the *rlnP method*.

The rlnP method allows us to check how well the model fits the data by using the parameters to go back to the probability distribution of unfolding forces (or the distribution of any other type of transition forces), $p(F)$, as is demonstrated by [48]. We find $p(F)$ remembering that $P_o(F)$ tells us the probability that the pseudoknot has opened at force F or below:

$$P_o(F) = \int_0^F p(F) dF$$

and $P_o(F) = 1 - P_c(F)$, so

$$p(F) = \frac{d}{dF}(1 - P_c(F)) = -\frac{d}{dF}P_c(F).$$

Thus $p(F)$, the expected distribution of unfolding forces, is found by differentiating the expression for $P_c(F)$ found in 3.7 [48]:

$$p(F) = \frac{k_0 k_B T}{r} \exp \left[k_B T F x^\ddagger - \frac{k_0}{r x^\ddagger} \left(e^{\frac{F(t)x^\ddagger}{k_B T}} - 1 \right) \right]. \quad (3.8)$$

3.3.1 The Dudko-Hummer-Szabo kinetics approach

Although the rlnP kinetic approach presented above is useful, neither of the assumptions made in Bell's formula (Equation (3.3)) to derive it are likely to be true in the experiments undertaken here or in similar single molecule experiments. In the last few years, several theoreticians have attempted to take this into account in their work. Specifically, in [49] and [52], the authors Olga Dudko, Gerard Hummer and Attila Szabo attempt to address the violation of the second assumption: As the force increases, the whole "energy landscape" of the pseudoknot (including x^\ddagger) changes. They therefore explicitly describe a theoretical free energy surface $U(x)$ along the pulling coordinate x :

$$U(x) = U_0(x) + \kappa(x - rt)^2/2$$

where $U_0(x)$ is the free energy surface at zero force and $\kappa(x - rt)^2/2$ is the potential due to the pull from the optical trap. As before, κ is trap stiffness and r is pulling speed. They use two different simple geometric formulas for $U_0(x)$, one that is "cusp" shaped, i.e. parabolic with a sharp drop-off at the transition state, and one that is cubic with a soft hill at the transition state. They then use Kramer's theory for particle escape over an energy barrier to find the rate of transition, $k(F)$.

The inverse of the transition rate, $\tau(F) = 1/k(F)$, which can be thought of as the average lifetime of the closed construct at a given force, is according to this formalism found as [52]:

$$\tau(F) = \tau_0 \left(1 - \frac{\nu F x^\ddagger}{\Delta G^\ddagger} \right)^{1-1/\nu} e^{-k_B T \Delta G^\ddagger [1 - (1 - \nu F x^\ddagger / \Delta G^\ddagger)]} \quad (3.9)$$

In this equation two new parameters have been introduced, namely ΔG^\ddagger , the "apparent energy of activation" that was discussed above in connection to Figure 3.4, and ν , a parameter describing the assumed shape of the energy landscape. As the authors describe, using fixed values of ν , Equation (3.9) can be fitted from experimental values for $\tau(F)$ found from histograms of unfolding forces from single molecule pulling experiments. This will be attempted in Chapter 5.

The shape parameter ν elegantly defines whether the assumed underlying energy landscape is cusp shaped ($\nu = 1/2$) or cubic ($\nu = 2/3$) [49]. Figure 3.5 illustrates the energy landscape shapes assumed by Dudko, Hummer, and Szabo (this figure is inspired by [48] as well as [49]). From the figure it is clear that the kinetic theories treated here still assume a two-state transition. It is also clear that the cusp-shaped and cubic-shaped energy landscapes depend on the first order reaction approximation made in Equation (3.6): there is no returning from the open state once the energy barrier has been passed.

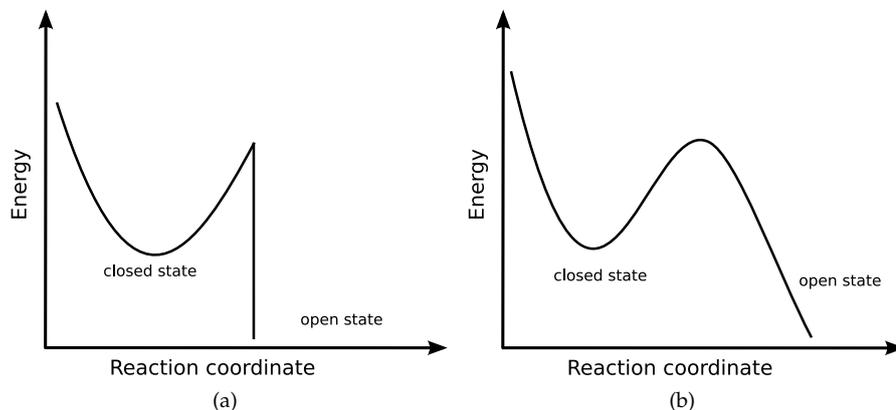


FIGURE 3.5: Energy landscape visualizations. (a) Cusp shaped two-state transition curve $U_0(x)$, corresponding to $\nu = 1/2$ in equation 3.9. (b) Cubic two-state transition curve $U_0(x)$, corresponding to $\nu = 2/3$ in equation 3.9.

3.4 Nonequilibrium thermodynamics

When molecular transitions are induced by a rapidly changing force as in the single molecule force spectroscopy used here, the system is not in equilibrium and classical thermodynamics cannot be used to find the transition energy. In the last fifteen years, however, new thermodynamic theories have been developed that are able to derive information from nanoscale nonequilibrium experiments. Below the theories used in this work are described.

3.4.1 Free Energy, the Jarzynski Equality and the Crooks Fluctuation Theorem

When the unfolding of a molecular construct takes place as an irreversible reaction, heat is dissipated and the work performed on the system by the surroundings during unfolding is not all transferred to internal energy in the construct. Therefore the work used in an irreversible reaction cannot be directly used to estimate the Gibbs free energy difference, ΔG , between the unfolded and folded construct. However, using the Crooks Fluctuation Theorem or the Jarzynski Equality, which describe how work energy and Gibbs free energy are related in a reaction taking place away from equilibrium, ΔG can be found.

The relation first discovered that may be used to find ΔG for a non-equilibrium transition is the Jarzynski Equality (JE) [27]:

$$\langle e^{-k_B T W} \rangle = e^{k_B T \Delta G}, \quad (3.10)$$

where W in the average on the left hand side is the work measured for *either* the forward or the reverse transition. It assumes that the transition being measured (whether forward or reverse) begins in equilibrium. Though formulated two years before the Crooks Fluctuation Theorem, it is actually a consequence derived from it. In fact when Crooks first derived his fluctuation theorem (though not in its final form), he was out to prove the Jarzynski equality [53].

The Crooks Fluctuation Theorem (CFT) states [26]:

$$\frac{P(W_F)}{P(-W_R)} = e^{(W - \Delta G)/k_B T}, \quad (3.11)$$

where W is an amount of work done on the construct, $P(W_F)$ is the probability distribution of the work exerted by the system on the construct during the forward transition, $P(-W_R)$ is the probability distribution of the work absorbed by the system during the reverse transition, and ΔG is the *reversible* change in the Gibbs free energy of the construct. Note that in (3.11), ΔF in the original formulation [26] has been replaced by ΔG , as is also the way the theorem is presented in the work of Collin et al. [46].

As shown by both Crooks [26] and Jarzynski [54], all one needs to do to derive the Jarzynski Equality, Equation (3.10), from the Crooks Fluctuation Theorem, Equation (3.11), is to multiply both sides by $P(-W_R)e^{-W/k_B T}$ and integrate for all transition work values remembering that ΔG is independent of the individual transitions and their associated work (variables as above):

$$P(W_F)e^{-W/k_B T} = P(-W_R)e^{\Delta G/k_B T} \quad (3.12)$$

$$\int_{-\infty}^{\infty} P(W_F)e^{-W/k_B T} dW_F = \int_{-\infty}^{\infty} P(-W_R)e^{\Delta G/k_B T} dW_R \quad (3.13)$$

$$\langle e^{-W/k_B T} \rangle = e^{\Delta G/k_B T} \int_{-\infty}^{\infty} P(-W_R) dW_R \quad (3.14)$$

$$= e^{\Delta G/k_B T}. \quad (3.15)$$

The derivation of the Crooks Fluctuation Theorem will not be repeated here, nor will the original derivation of the Jarzynski Equality. For the derivation of the CFT, the reader is referred to [53] which is detailed and more intuitive than the proof in [26] (the CFT appears in the first line of equation 10 in [53]. This author first noticed this derivation through [55]).

3.4.2 Using the Jarzynski Equality and the Crooks Fluctuation Theorem

The Jarzynski equality has the advantage that it requires data only for transitions in one direction. The disadvantage, which is quite substantial, is that since it uses an average of exponential functions, it is disproportionately influenced by the smallest work values measured. Very large numbers of measurements are required to give good statistics and make up for this skewing. The further the process is from equilibrium, the more data are required. In practice, Jarzynski noted, it would probably be very hard to obtain enough data if the heat dissipation is more than about $k_B T$ [27]. Others have calculated that if the dissipation is more than about $9k_B T$, about 8000 experimental repetitions are needed (!) [47]. However, a better estimate appears possible to obtain if the Jarzynski equality is used on both unfolding and refolding work distributions and the two are averaged, as is done in [46].

The Crooks Fluctuation Theorem, on the other hand, is easy to employ even with relatively limited data sets. The theorem implies immediately from Equation (3.11) that at the point where $P(W_F) = P(-W_R)$, we have $W = \Delta G$. Thus even though the reaction is taking place under non-equilibrium conditions and is irreversible, ΔG for the reversible transition can be found from the intersection of the distributions of work used in the unfolding and refolding transitions. This is illustrated in Figure 3.6. The further the reaction is from equilibrium, the more heat is dissipated and the less $P(W_F)$ and $P(-W_R)$ overlap. This makes it more difficult to find their intersection, but using a Bennett's acceptance ratio [56], which describes how the distribution of states must be between two energy levels that obey detailed balance, it can still be done [46, 57].

Note that the CFT applies under the following assumptions:

- The state in which the system begins the unfolding transition is the same as the one in which it ends the refolding transition, and vice versa, that the state in which the unfolding transition ends is the one in which the refolding transition begins.

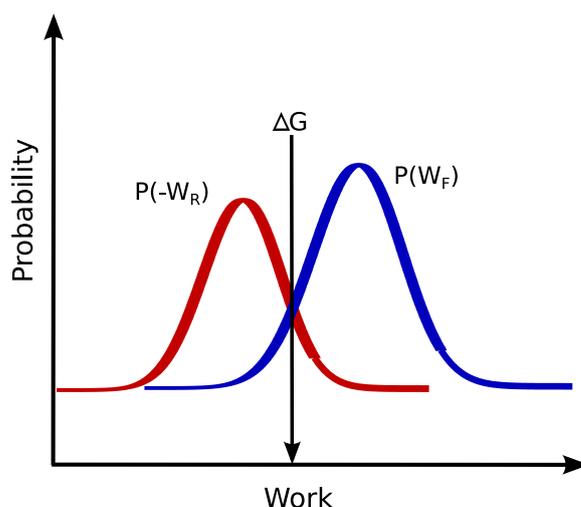


FIGURE 3.6: The Crooks Fluctuation Theorem states that the Gibbs free energy, ΔG , of an irreversible transition is equal to the work value where the forward and reverse transition work distributions overlap. If the transition were reversible, $P(W_F)$ and $P(-W_R)$ would overlap and both center on ΔG . Note that the work distributions are here approximated by Gaussian distributions. They are not truly Gaussian, but resemble them [58].

- The transition, though overall irreversible, is microscopically reversible, i.e., at any given moment, if the velocity were reversed, the system would be just as likely to move in the reverse direction as it was to move in the forward direction with original velocity.

The latter condition means that at any given moment we cannot tell whether we're looking at the forward or the reverse reaction. This is true even though overall, when looking at many irreversible forward and backward reactions, we can tell that there is hysteresis in the system, i.e., that the forward and backward reactions on average require/return different amounts of energy.

3.4.3 Work distributions and handles

To employ the Crooks Fluctuation Theorem (CFT) and the Jarzynski equality in practice, we must determine the work distributions. The work is found as the force applied to the construct times the distance moved by the system. Formally the distance moved should be measured as the distance between the traps, i.e., rt where r is the pulling rate and t is the time since the beginning of the pull. Thus the work is the area under the curve of force vs rt [58]. However, more often in literature the area under the curve of force versus molecular extension is used [46, 47, 58]. When the data is smoothed, the error introduced by this approximation becomes negligible [58].

The work for the transition measured from each force-extension curve includes several contributions that do not relate to the change in Gibbs free energy, ΔG_{PK} , during the chemical transition from closed to open pseudoknot. Following the lettering in Figures 3.2, the contributions during the cycle are:

- Curve through D to A: The handles and the closed pseudoknot are stretched

- A to B: Rip. The pseudoknot opens, the handles relax, the closed pseudoknot is relaxed to zero extension and the open pseudoknot is stretched from the extension of the closed pseudoknot at A to its own equilibrium extension at B
- Curve rising above B: The handles and the open pseudoknot are stretched
- Curve descending through B to C: The handles and the open pseudoknot are relaxed
- C to D: Zip. The pseudoknot closes, the handles stretch, the open pseudoknot is relaxed to zero extension and the closed pseudoknot is stretched from zero extension to its equilibrium extension at D
- Curve descending below D: the handles and the closed pseudoknot are relaxed

Collectively the contributions from stretching the handles and the pseudoknot will be called W_{stretch} . These must be subtracted from ΔG for the full transition found using the CFT or the JE to yield ΔG_{PK} . This is done by calculating the theoretical value of W_{stretch} using the EWLC as will be described in Chapter 4.5.1. The assumption is that the EWLC describes force versus extension of the stretching of the construct well. Implicitly this entails that the stretching and relaxing of the handles and the pseudoknot are reversible on the time scales of our experiments [4, 46, 47].

3.4.4 Measuring work distributions - two methods of finding ΔG_{PK}

For the CFT, the work must formally be measured for transitions which all start and end in the same states. For the Jarzynski Equality they must formally all start in equilibrium. However, since we are interested only in ΔG_{PK} and will subtract W_{stretch} from ΔG for the full transition to find it, we can ignore the part of W_{stretch} that is the same for all transitions. Therefore for finding the work distributions for both the CFT and the Jarzynski equality we can find the integral of the force-extension curves for all pulls between a starting force corresponding to the lowest opening force observed and an ending force corresponding to the highest force observed. This is what is done by Liphardt et al. [4] in their 2001 paper experimentally validating the Jarzynski equality and by Collin et al. [46] in their 2005 paper experimentally verifying the Crooks fluctuation theorem (the methods that these papers use are explained in more detail than in the original articles by Tinoco et al. in a 2006 Quarterly Reviews of Biophysics article [47]). Both groups measure the opening and closing of an RNA hairpin and Collin et al. also investigate another more complex RNA structure; these experiments are thus possible to relate to the ones undertaken here.

The approach of [4] and [46] is theoretically stringent but highly vulnerable to error arising from the estimation of W_{stretch} from the EWLC when the difference between the lowest and highest opening and closing forces is large and W_{stretch} therefore also becomes large. This may be problematic since as noted above the EWLC assumes $L_c \gg L_p$, which is not the case here. Therefore, in this work two different variant approaches are applied which reduce this error but introduce extra assumptions.

The first approach is closely related to the approach applied by Liphardt et al. and Collin et al. [4, 46]. The modification made in this thesis is that rather than integrating between the same forces for all pulls, I integrate only between the refold force and unfold force recorded in each pull (points D and B in Figure 3.7). Thus the states of the construct at point D for all pulls and at point B for all pulls are treated as though they is the same no matter at what force they occurs (in other words, the state at point B is not considered equivalent to the state at point D, but differences in state between points D occurring at different forces in

different pulls are ignored, and likewise for the different points B found in different pulls). Despite this difference between the approach here and the one used in the literature, this approach will be called the "Collin approach" in the remainder of this thesis.

The second approach, which will be called the "simple approach", estimates the work required during the opening transition as the area only beneath the rip of the force extension curve. See Figure 3.7. The work required during the closing transition is similarly found from the work beneath the zip of the force extension curve. This approach extends the approximation made above by treating the transition as though the state at the beginning of the opening rip (point A in Figure 3.7) and at the end of the closing zip (D) is the same even though it occurs at different forces and extensions. Similarly, the state at B and C in Figure 3.7 is treated as though it is the same. Thus any stretching or relaxing of the construct between D and A and between B and C is ignored. The advantage is that the contribution W_{stretch} to the work energy is much smaller in this approach. As W_{stretch} is difficult to estimate correctly due to the uncertainty of the EWLC prediction, this simple approach may therefore be more reliable.

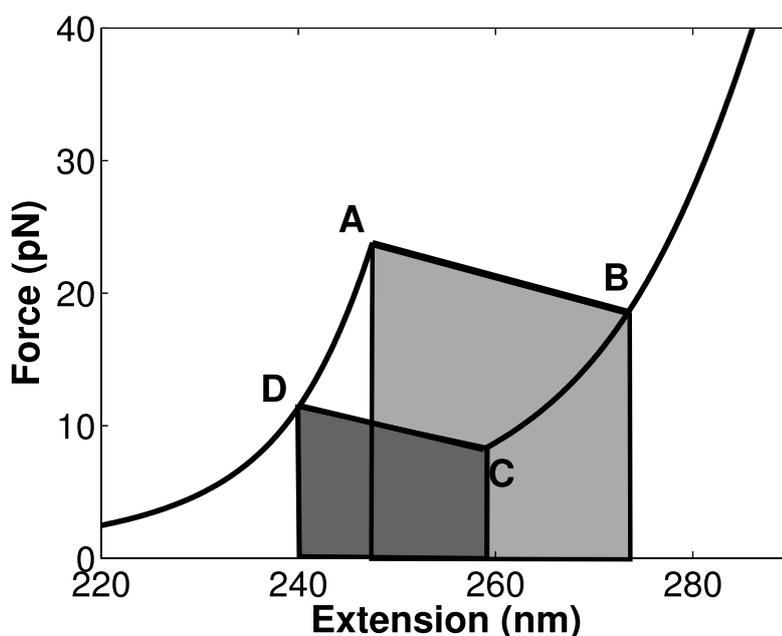


FIGURE 3.7: In the "Simple" approach to finding the work distributions, W_F for each pull is found as the area beneath the force extension curve's rip from A to B while W_R is found as the area beneath the zip from C to D. In contrast, with the Collin approach, the area beneath the curve from D to A to B is found for the forward reaction and the area beneath the curve from B to C to D is found for the reverse reaction.

The second approach is the one used in the supplementary material in Hansen et al. [17, 43]. The approach used in Green et al. [25] is not clear. This is the only other source found where an attempt is made to calculate the Gibbs free energy of the transition between the open and the closed pseudoknot.

How the contributions to W_{stretch} is calculated using each method will be described briefly below.

3.4.5 Calculating the work to stretch handles and pseudoknots

For the Collin approach, W_{stretch} is the same for the forward transition measured from D to B in Figure 3.7 and the reverse transition from B to D. It is equal to the work required to stretch (or in some cases gained by relaxing) the handles from D to B, the work gained by relaxing the closed pseudoknot from D to zero and the work spent stretching the open pseudoknot from zero to B.

Since in the Collin approach the work is calculated between different forces F_D and F_B for each pull, W_{stretch} is also calculated for each of these sets of F_D and F_B . These values were averaged for subtraction from ΔG to give ΔG_{PK} .

In the simple approach illustrated in Figure 3.7, W_{stretch} is different for the forward (rip) transition and the reverse (zip) transition. For the work exerted during the rip, $W_{\text{stretchRip}}$ is the work gained by relaxing the handles from A to B, the work gained by relaxing the closed pseudoknot from A to zero extension and the work used to stretch the open pseudoknot from zero to B. During the zip, $W_{\text{stretchZip}}$ is the work spent stretching the handles from C to D, the work spent stretching the closed pseudoknot from zero to D, and the work gained by relaxing the open pseudoknot from C to zero.

An average W_{stretch} for the simple approach is calculated from $W_{\text{stretchRip}}$ and $W_{\text{stretchZip}}$ weighted by the numbers of rips and zips included in the analysis. The numbers of rips and zips is not necessarily the same, since this approach requires only a clear, one-step opening or closing of the pseudoknot, not that both are observed in the same pull, which may be the case as will be seen in Results, Chapter 5. The average W_{stretch} is subtracted from ΔG to give ΔG_{PK} .

3.5 Summary of molecular transition theory

This chapter has presented the main characteristics of the molecular transitions generally expected in single molecule pulling experiments. It has shown the theory behind the kinetic equations used to find parameters such as the distance to transition state, x^\ddagger , and the rate of reaction at zero force, k_0 . It has also shown newer kinetic theories developed by Dudko, Hummer, and Szabo [52] that may be able to provide more information on the underlying energy landscape of the molecular transitions. Finally we have seen how the Crooks Fluctuation Theorem and the Jarzynski Equality may be used to extract information on the Gibbs free energy (ΔG) of a molecular transition even if the transition occurs irreversibly with heat dissipation. We have also examined in detail how to find the distributions of work energy needed to make the calculation of ΔG in practice.

It is now time to look at the detailed experimental routines used in this thesis before presentation of the results of the experiments.

4 Materials and Methods

This section will present the experimental setup and protocols used during the thesis work including the the molecular constructs, biochemical assays, optical tweezer setup, calibration routines, error sources, and the data processing routines.

4.1 The pseudoknots - structure and frameshift efficiencies

Two pseudoknots whose structures are related to a wild-type frameshifting pseudoknot found in Infectious Bronchitis Virus (IBV) are studied in this project. They are named PK 11/6 for a pseudoknot with eleven basepairs in Stem1 and six in Stem2 and PK 6/11 for a pseudoknot with six basepairs in Stem1 and eleven in Stem2. Figure 4.1 shows the expected structures of PK 11/6 and PK 6/11 together with one of the pseudoknots treated by Hansen et al. [17], PK401, which like PK 11/6 has stem lengths of eleven and six base pairs for Stem1 and Stem2 respectively. The wild-type IBV pseudoknot also has eleven basepairs in Stem1. In previous studies, artificial pseudoknots with Stem1 length shorter than eleven basepairs have proved much less efficient at provoking frameshifting than the wild-type, whereas a longer Stem1 still worked equally well [59].

It is important to note here that our study, Hansen et al. [17] and Green et al. [25] all rely on theoretical programs to predict how variations in base composition will change the basic structure of the pseudoknot, i.e. which stems and loops will form. In Hansen et al., pknotsRG [61] is used, and this was also the program used during the design of this experiment to predict the structures of PK 11/6 and PK 6/11 as seen in Figure 4.1. Newer program options include DotKnot [63] and HotKnots [64]. All three of these programs also predict the Gibbs free energy of formation of the pseudoknot from the free single strand based on the base pairings expected, but not on tertiary interactions. The crucial point to note is that whereas the program pknotsRG only displays a single predicted structure, the one it finds optimal, DotKnot and HotKnots present a number of different conformations with slightly different energies of formation for a single input sequence of RNA bases. For the PK 11/6 and PK 6/11 sequences, some of these outputs are pseudoknots with entirely different stem or loop lengths than those expected from the pknotsRG output. Others are not pseudoknots at all and others again strongly resemble the pknotsRG output, with variation only in the hairpins formed within the loops or on the single stranded parts of the handles. A number of alternative possible structures are shown in Chapter 6, Figures 6.1 and 6.2, where they will be discussed further.

The upshot of the uncertainty in the structural prediction is that the loops and stems being opened in a pseudoknot experiment might be shorter than expected or that the structure might not be a pseudoknot at all. Since several possible structures are predicted which have similar energies of formation it is possible that several different molecular conformations are randomly encountered during experiments. Chen et al. consider this possibility explicitly in two different papers as they discuss evidence for misfoldings occurring during their experiments on human telomerase RNA pseudoknots [18, 21]. A

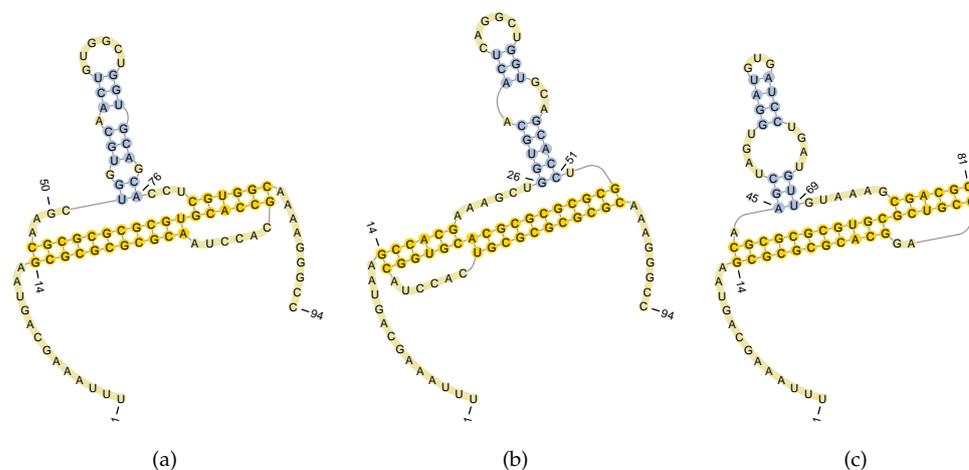


FIGURE 4.1: Sequence and predicted structure of a) PK 11/6, b) PK 6/11 and c) PK401. PK 11/6 and PK 6/11 are investigated here. PK 11/6 causes some frameshifting while PK 6/11 does not [20, 60]. PK401, investigated by Hansen et al. [17] is similar structurally to PK 11/6 but frameshifts more effectively. Predicted structures found with pknotsRG [61] and rendered by PseudoViewer 3.0 [62]. Pseudoknot stems in yellow. Hairpins (blue) are not expected to significantly influence pseudoknot strength. PK401 is shown without downstream single strand handle.

similar scenario is discussed by Li et al. [65], who examined the folding and unfolding of a HIV transactivation response region RNA hairpin with a bulge.

Assuming that the structures shown in Figure 4.1 are correct, PK 6/11 is more or less a reversal of PK 11/6: Stem1 has become Stem2 and vice versa, while Loop1 has become Loop2 and except for a one-nucleotide insertion Loop2 has become Loop1. The insertion of an extra nucleotide is made to keep the amino acid sequence produced by the mRNA more uniform so that the frameshift products can be more accurately compared [66]. This addition is countered by the deletion of a nucleotide in the portion of the mRNA that follows the pseudoknot so that the remainder of the cistron being translated will be the same for both pseudoknots. The sequences and corresponding amino acids are shown in Figure 4.2.

11/6

```

TTTAAAGCA GTA AGC GCG CGC GCA ATC CAC GCC ACG TGC GCG CGC GCA AGC TGG TGC AAC TGT GGC TGG TGC AGC ACCTCG TGG CAA AAG GGC C
V S A R A I H A T C A R A S W C N C G W C S T S W Q K G

```

6/11a

```

TTTAAAGCA GTA AGC CAC GAA AGC TGG TGC AACTCA GGC TGG TGC AGC ACCTGC GCG CGC GCA CGT GGC ATC CAC TGC GCG CGC GCA AAG GGC C
V S H E S W C N S G W C S T C A R A R G I H C A R A K G

```

FIGURE 4.2: Pseudoknot nucleotide sequences and corresponding amino acids (rendered by Jesper Tholstrup)

TABLE 4.1: Comparison of pseudoknots. PK 11/6 and PK 6/11 are the pseudoknots investigated in this work. PK 10/6 was investigated by Jesper Tholstrup in parallel to this work [67]. PK 401 was investigated by Thomas Hansen [69]. The characteristics of the wild-type IBV pseudoknot are included for extra perspective. The numbers of basepairs and nucleotides in stems and loops are predictions based on the sequence of bases in the RNA single strand. $\Delta G_{\text{transition}}$ is a prediction of the free energy needed to form the basepairs of the expected structure from the single strand.

	PK 11/6	PK 6/11	PK 10/6	PK 401	Wild-type ^a
Stem1 basepairs	11	6	10	11	11
Stem2 basepairs	6	11	6	6	6
Loop1 nucleotides	6	33	6	2	2
Loop2 nucleotides	32	6	32	32	32
Predicted $\Delta G_{\text{transition}}$ (kCal/mol) ^b	-43.1	-45.9	-41.6	-39.4	-37
Frameshift efficiency (1D gel) ^c	10 %	~ 0	8 %	14 %	30 %
Stall efficiency ^d	2.3	1.0	1.5	~ 1.1	Unknown
Stall percentage ^e	~ 60	~ 0 %	~ 30	~ 10 %	Unknown

^a Information on wild-type pseudoknot structure found in [59].

^b Found using pknotsRG [61] except for the wild-type, whose ΔG was found using DotKnot [63]. Loop2 sequence assumed the same as in PK 11/6 and PK 401.

^c Frameshift efficiency is the percentage of frameshifted mRNA protein product out of all protein product. Does not count protein products from stalled ribosomes as frameshift products [60]. For PK 11/6, PK 6/11, and PK 10/6 measured by Jesper Tholstrup and Michael Sørensen [60]. Percentage for PK 401 from [17] and for wild-type from [68].

^d Stall efficiency is the ratio of total frameshifted protein products (including products from stalled ribosomes) to the amount of frameshifted protein products from non-stalled ribosomes. For PK 11/6, PK 6/11, and PK 10/6 numbers provided by Jesper Tholstrup and Michael Sørensen [60]. For PK 401, estimated from Figure 3B in [20].

^e For PK 11/6, PK 6/11, and PK 10/6 numbers provided by Jesper Tholstrup and Michael Sørensen [60]. For PK 401, estimated from Figure 3B in [20].

Apparently, the loop lengths are not as important as the stem lengths in determining how well the IBV-variant pseudoknots frameshift. The only requirement is that Loop2 must be long enough to span Stem1 [59]. The length of loop2 in PK 11/6 and PK401 is the same length, 32 nucleotides, as in the wild-type pseudoknot, while in the wild-type, loop 1 is only two nucleotides long [59].

The frameshift efficiencies of the pseudoknots described here were measured by Jesper Tholstrup and Michael Sørensen. With a conventional assay for measuring frameshift efficiency, which uses one-dimensional SDS-PAGE gel analysis, they found that the percentage of protein product that was frameshifted was about 10 % for PK 11/6 whereas for PK 6/11 the amount of frameshifting was negligible, meaning that it was indistinguishable from the spontaneous frameshift rate at only 0.1 % [60]. With novel usage of a two-dimensional SDS-PAGE gel to investigate frameshift efficiency, they additionally found that a little more than half of the ribosomes that were frameshifted by PK 11/6 actually got stuck ("stalled") within the pseudoknot [20, 60]. For PK 6/11 there was no detectable stalling of ribosomes [60]. A gel analysis performed by Michael Sørensen to investigate the stalling of PK 6/11 is shown (and explained) in Appendix B.

For comparison with the pseudoknots investigated in this thesis, an extremely closely related pseudoknot called PK 10/6, which was investigated by Jesper Tholstrup in parallel

to the work done in this thesis, had 8 % frameshift efficiency and displayed about 30 % stalling of frameshifted ribosomes [60]. PK401 gave rise to about 14 % frameshifted protein product (investigated with one-dimensional gel analysis) [17] and only a small proportion permanently stalled ribosomes, about 10 % [20]. PK401 is the one of the three pseudoknots shown in Figure 4.1 that most closely resembles a naturally occurring -1 frameshifting pseudoknot: its stem1 is much longer than stem2 and it has a very short loop1. The wild-type IBV pseudoknot originally investigated 20 years ago by Brierley et al. [68] has different stem nucleotides from PK401 but the two are otherwise structurally similar. This wild-type pseudoknot displayed 30 % frameshifting according to one-dimensional gel analysis [68]. The expected structural characteristics and measured frameshift data for all five pseudoknots are summarized in table 4.1.

4.2 The molecular constructs

For the pseudoknot pulling experiments, short strands of mRNA are attached to DNA handles so that RNA and DNA form hybrid double stranded handles with a short section of single-strand RNA in between. The RNA single strand region includes the pseudoknot sequence and very short sections on either side including on the upstream (5') end the slippery sequence and spacer region before the pseudoknot. The RNA flanking the single strand region is hybridized to DNA single strands. One of the DNA strands is attached to a digoxigenin group, the other to biotin. For the experiments the biotin-labelled end of the RNA-DNA construct is allowed to bind to a streptavidin-coated polystyrene bead while the other end binds to a smaller anti-digoxigenin coated bead. This is illustrated in Figure 4.3.

Both biotin-streptavidin and digoxigenin-antidigoxigenin complexes are commonly used in single molecule experiments to tether DNA to surfaces. Biotin-streptavidin interactions create a very strong non-covalent bonds, while digoxigenin-antidigoxigenin are connected by a weaker non-covalent antibody bond. When molecular constructs break mid-pull, it is assumed that it is the digoxigenin binding that is broken [43]. This is because it is by far the weakest link in the chain. Like the biotin-streptavidin bond, the covalent bonds of the single stranded RNA and DNA nucleotide backbones are far stronger than the digoxigenin-antidigoxigenin bond. The non-covalent base pairings joining the single stranded RNA to the DNA handles are not individually strong (as seen in RNA and DNA hairpin unzipping experiments, e.g. [4], they are easily unzipped if pulled in the direction normal to the double strand). When exposed to shear stress rather than normal stress, however, they only start melting at a force of about 65 pN, where their force-extension curve starts exhibiting a characteristic sawtooth pattern [45]. This sawtooth pattern was observed twice at about 65-80 pN during the experiments performed for this thesis (data not shown). This underscores the unlikelihood that the construct would break in any other location than at the digoxigenin-antidigoxigenin handle attachment.

While the upstream handle attached to the 5' end of the RNA may be connected by more than one biotin-streptavidin bindings to bead, the downstream handle attached to the 3' end of the RNA is always attached to the styrofoam bead by only one digoxigenin bond. The reason is how the handles are produced: They are made using PCR, polymerase chain reaction, which produces huge numbers of copies of a specified DNA sequence using only short complementary primers from which the reaction starts. The primers must correspond to the upstream (5') section of the sequence, so any marker attached before PCR must be attached to the 5' end so it does not interfere with the lengthening process. Thus when handles are produced, each downstream (5') handle already has a digoxigenin group attached to its primer, whereas biotin is added enzymatically to the upstream (3') handle after PCR [69].

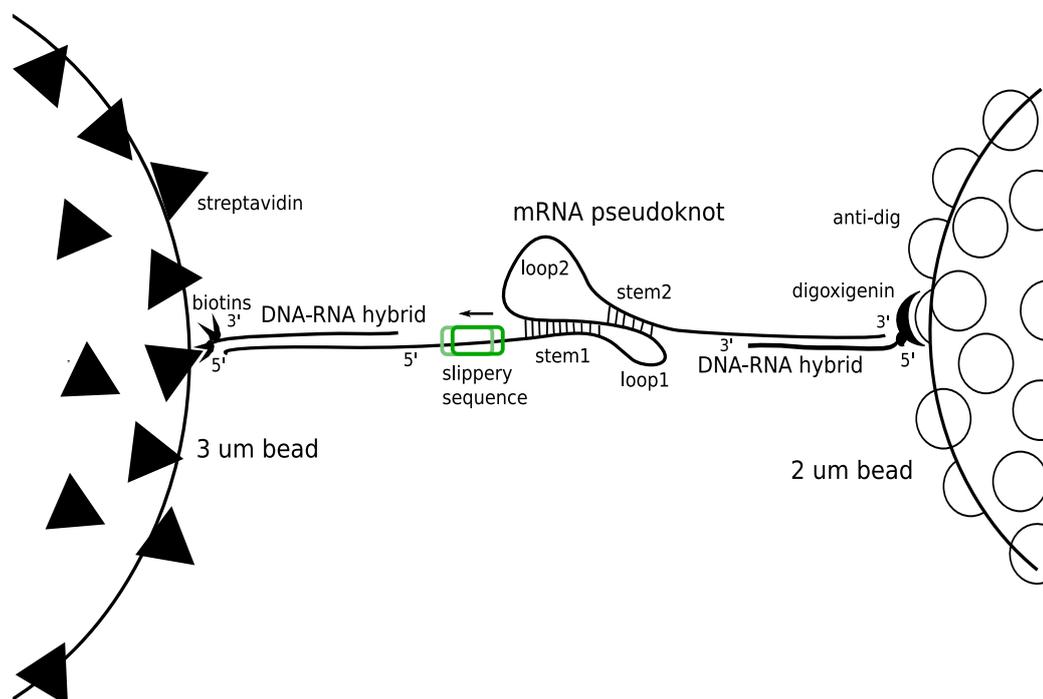


FIGURE 4.3: Conceptual image of molecular construct and attachments to beads. Not to scale. Polystyrene beads $3\ \mu\text{m}$ in diameter coated with streptavidin are bound to biotin on the DNA-RNA upstream handle, while beads $2\ \mu\text{m}$ in diameter coated with anti-digoxigenin antibodies are bind to a digoxigenin group on the downstream handle.

4.3 Biochemical methods

For this thesis, the author participated in synthesizing the downstream DNA handles and the mRNA pseudoknot strand, annealing handles to RNA, and preparing RNA-handle constructs with beads for pulling. The biochemical protocols are described in detail in Appendix C. All protocols were suggested by and carried out with the guidance of Jesper Tholstrup, who also synthesized the plasmids, upstream handles, and DNA oligos used for downstream handle synthesis, which were used in the processes described below.

The basis for the molecular construct is the pseudoknot-containing RNA strand. It is made from a DNA template by *in vitro* transcription by RNA polymerase. Even when kept frozen at -20 , RNA is not as durable as DNA, so the constructs need to be made every few months.

Downstream handles for the construct were prepared from DNA oligos which had digoxigenin attached to the 5' end of the strand that would become the handle strand. The oligos had been generated by PCR by Jesper Tholstrup. They were extracted with phenol and digested with λ -exonuclease which breaks down DNA strands from the 5' end (digoxigenin protected the handle strand, allowing the exonuclease only to remove the opposite single strand, thus leaving a single stranded handle). After digestion, the handles were washed to remove the exonuclease, carefully extracted with phenol again, washed with salt, measured and frozen for later attachment to RNA constructs.

The RNA portion of the molecular construct was transcribed from linear DNA templates, which were made by PCR amplification of plasmid DNA. Plasmid DNA is explained in

Appendix C; the plasmid DNA used here had been synthesized and purified by Jesper Tholstrup and encoded the RNA-pseudoknot sequence plus handles, slippery sequence and spacer. The DNA templates were transcribed to make RNA using T7 RNA polymerase, a polymerase originally found in T7 phages. The process requires great care, since RNA is easily destroyed by omnipresent RNA-ase enzymes. To counteract this, RNA-ase inhibitor, RNA-sin, is added to the reaction mixture. Nonetheless, this is the reason that the RNA constructs do not last indefinitely.

After the RNA strands were produced, they were annealed to the DNA handles by heating. The process was not always successful as the relative amounts of RNA solution and the two types of handle solutions had to be appropriately adjusted. Careful checking of the partial products and final outcome of the process using an agarose gel was necessary. See Figure 4.4 for a successful example. Once the handles had been attached, the constructs were ready to be annealed to polystyrene beads for the pulling experiments.

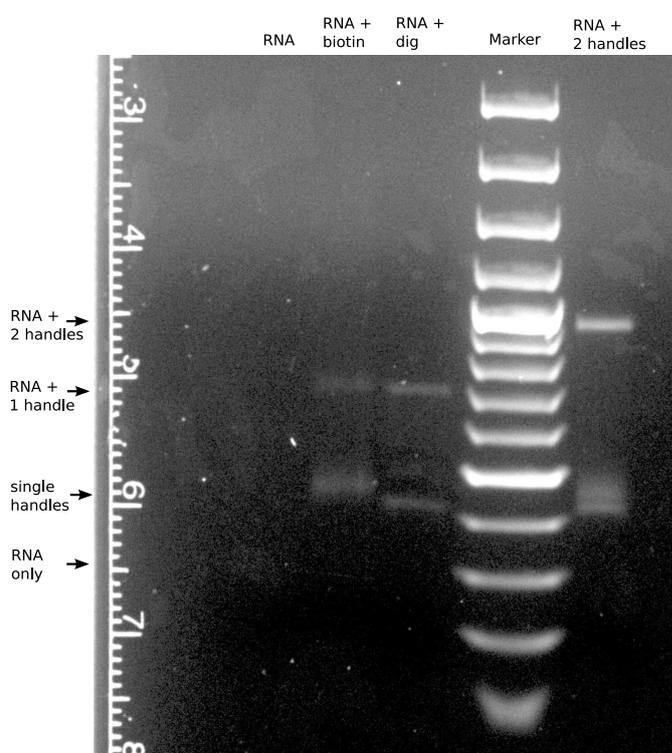


FIGURE 4.4: Checking for synthesis of full construct with two handles on an agarose gel. This example was from the synthesis of PK 11/6. The smallest molecules run the furthest down the gel. The marker, Ethidium Bromide, binds best to double stranded DNA and RNA, so the single stranded RNA with no handles bound to it does not show up clearly. Samples from left to right: RNA only, RNA + upstream handle with biotin, RNA + downstream handle with digoxigenin, marker, and RNA with both handles

4.3.1 Preparing samples for pulling experiments

For pulling experiments, RNA-handle constructs were mixed at high concentration with one type of beads (streptavidin coated) and allowed to bind for some time. Then the

mixture was diluted before the second type of beads (different diameter, antidigoxigenin coated) was added, so that there would be high likelihood that RNA constructs had bound to the streptavidin coated beads but not to the antidigoxigenin coated beads before being added to the sample chamber. A detailed guiding protocol is given in Appendix C.

The relative amounts of beads and construct were adjusted over the course of experimentation to give optimal likelihood of catching beads and observing single molecule binding events. The amount of RNA construct and the time for construct binding to streptavidin beads was adjusted to make it more likely that one and only one molecular tether would form between beads. The amount of beads was adjusted so that one of each kind could conveniently be caught while the remaining beads in solution were not so numerous that they would interfere with the experiment by diffusing into the traps. In particular, there had to be enough anti-digoxigenin coated beads to allow the experimenter to catch one with an optical trap before they attached permanently to the glass surface of the sample chamber.

The buffer used during pulling experiments was called buffer R. It contained 10 mM Tris-HCl pH 7.5, 250 mM NaCl and 10 mM MgCl₂. Tris-HCl keeps the buffer at the correct pH and is commonly used for biochemical solutions with nucleic acids. The amounts of other salts are also important for how the nucleic acids fold. In particular, Mg²⁺ ions, which like K⁺ and Na⁺ are present under physiological conditions, shield the negative charges on the nucleic acid backbone, creating more favorable conditions for double stranded nucleic acid conformations. Therefore RNA pseudoknots and hairpins are expected to be significantly stronger in the presence of Mg²⁺. This has been experimentally confirmed, e.g. by Liphardt et al. [4].

4.4 Pulling experiments

As described in Chapter 2.2, a double optical trap setup was used for the pulling experiments. During the experiments, one trap was moved while the other was held steady. As also described in Chapter 2, photodiodes measured the displacement of the beads from the centres of the optical traps. Stokes calibration and power spectrum analysis of the unattached beads were subsequently used to translate volt output from the photodiodes into force and distance.

The molecular extension was calculated from the distance between the traps minus the displacement of the bead centers from the trap centers. It is important to note that the absolute distance between the traps was not known. Zero was set at the point at which the beads started moving. The trap-trap distance measurement was then updated as a function of time and the velocity of the moving trap, while the displacements of the beads from the trap centers were calculated from the photodiode outputs. Since the absolute separation of the beads could not be measured accurately, the absolute molecular extension is not known.

The movement of the traps, the objectives, the mechanical stage and the piezo that allows fine control of stage position were all controlled from the software that accompanies the NanoTracker. The traps can be moved in the x, y, and z direction with x-y the sample plane and z the vertical direction through the microscope, i.e., the direction of the laser through the objective and sample to the condenser.

4.4.1 Sample chamber and trapping routine

Sample chambers were made from glass slide object covers and double adhesive tape. They were sealed by silicone wax so that the sample solution did not evaporate because of the laser heat during trapping. Silicone wax is also used to attach the larger glass slide to

the sample holder. Images of a chamber are shown in Figure 4.5. The objectives used for viewing the sample and focusing the laser were water immersion objectives. To avoid evaporation of immersion liquid due to the heat of the trapping lasers, special immersion oil was used with a refractive index very close to that of water ($n_r \approx 1$). A droplet was placed on either side of the sample chamber near its center before the chamber was placed inside the NanoTracker.

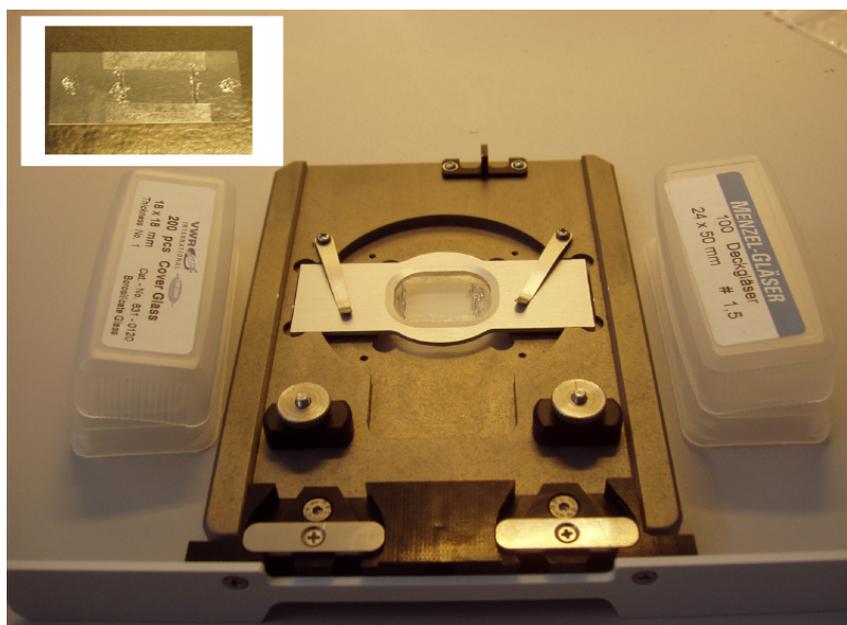


FIGURE 4.5: Sample chamber stuck to sample holder flanked by boxes of objective slides used for chamber construction. The inset shows the sample chamber itself. It is made from two glass objective covers separated from each other by small pieces of double adhesive tape and sealed along the other edges by silicone wax. Silicone wax is also dotted near the ends of the larger glass slide to allow it to stick to the sample holder.

Before experimentation, the objectives had to be adjusted in height using the NanoTracker software to allow trapping at correct sample level and to give a focused camera image. For further experiments, note that the focusing routine requires great care to avoid crashing the upper objective into the sample. The lower objective may need to be lifted slightly using the lever for changing objectives so that it attains the correct height.

Once the objectives were focused, beads were located and trapped, and calibration was carried out. The beads were then approached to each other so that a molecular tether might form between them. The traps with the beads could be placed near each other using the computer mouse to move the Nanotracker trap symbols on the camera image. Fine approaching was subsequently done in 20 nm increments until the voltage signal increased nonlinearly during small (~ 40 nm) reseparations of the beads, showing that a molecular tether was caught (a linear increase in the voltage indicated that the beads were inelastically stuck to each other by a polystyrene tether).

When a possible RNA tether had been observed, a programmed routine was used to repeatedly move the beads apart and together again a set distance at a set velocity. The voltage signal was recorded continuously during each pull. In this way long series of pulls could be recorded for subsequent analysis. Since the pulling speed was $0.1 \mu\text{m/s}$ and the

TABLE 4.2: Average effective trap stiffness of all experiments, pN/ μm . κ_1 is the stiffness of Trap1, κ_2 is the stiffness of Trap2, and κ_{eff} is the effective stiffness experienced by a molecule held between the traps.

	κ_1		κ_2		κ_{eff}	
	μ	std	μ	std	μ	std
PK 11/6	153	33	224	35	91	13
PK 6/11	209	46	225	29	108	14

pull distance was about 400-700 nm in each direction, each pull took about 8-14 seconds. The waiting time between pulls varied, usually between a few seconds and half a minute.

4.4.2 Nanotracker particularities: Trap interference and trap intensity variation

Figure 4.6 (a) shows the proximity of the beads as seen through the camera when a tether is caught. Since the beads were so close, some interference occurred between the traps. This meant that the larger bead in trap2 was easily pushed above the laser focus and lost. The z height of the trap foci therefore had to be adjusted during each experiment so that the focus of trap2 was located slightly above the focus of trap1.

The effect of trap1 on the bead in trap2 and vice versa meant that when the beads were moved apart during a pull, each adjusted slightly in z height, changing the voltage output a little bit also for the signal for the y-direction (the pulling direction in which we measured displacement of the beads and traps). It was important to try to minimize the disturbance of the y-signal while maintaining the stability of the beads in the traps. This was done by repeated adjustment of the trap z height interspersed with "baseline" measurements in which the traps were moved as they would be for pulling a molecular tether while no tether was caught. A baseline measurement ideally should show no change in the displacement of the beads from the trap centers as the traps were moved relative to each other. In practice the change in displacement of the beads from the trap centers due to interference as the traps were moved could be limited to a signal that corresponded to an error of up to about 2-3 pN (usually much less) in the force-extension curves.

Figure 4.6 (b) shows an image of the reflection of the laser beams at the upper surface of the sample chamber. In this image the two traps are nominally set at the same z-height and laser intensity. It is impossible to know whether the traps are in fact at the same height, but attempts to focus the beam reflections as much as possible indicated that they were offset somewhat, perhaps up to 0.4 μm . Despite any possible difference in trap height, what stands out in the image is the difference in the size and intensity of the laser reflection spot, which gives an indication of the difficulty of predicting and controlling the strength of each trap.

Not only was it hard to predict the distribution of laser intensity between the two traps. The trap strength also varied quite a bit from experiment to experiment despite nominally being set to the same power level. This meant that trap stiffness varied significantly from experiment to experiment. Table 4.2 shows the average trap stiffness of all the experiments with each type of pseudoknot. The size of the standard deviation gives some indication of the great variation in measured stiffness. The effective trap stiffness, κ_{eff} , which is the trap stiffness experienced by a molecule caught between the beads in the traps, is calculated as:

$$\kappa_{eff} = \frac{\kappa_1 \kappa_2}{\kappa_1 + \kappa_2}.$$

The loading rate, r , is calculated as $v\kappa_{eff}$ where v is the pulling speed.

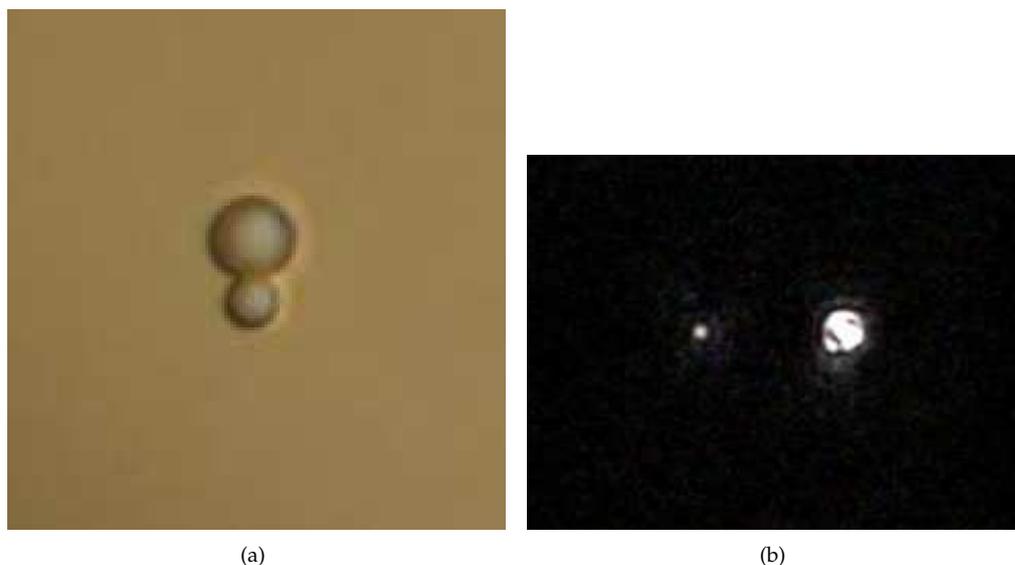


FIGURE 4.6: Pull layout and laser strength. (a) : Layout of a typical pull. Trap2 holding the large bead is located just next to trap1 with the small bead ("above" it in the y-direction). The z-height of the two traps has been adjusted so that the beads remain stable in the traps when this close. That means that the z-height of the center of the bead in trap2 most likely is above the center of the bead in trap1 (based on appearance in photo, taken from below, in which the bottom of the two beads appear to be at about the same height though their diameter is different). (b) Reflection of the laser light at the bottom surface of the sample chamber. Laser enters from below and photo is taken from below. The two traps are nominally at the same z-height and laser intensity.

4.5 Data processing and statistics

Most of the data processing was carried out using Matlab scripts written by the author. This included calculation of the conversion factor α based on Stokes calibration, generation of force-extension curves, recording of rip and zip forces on force-extension curves, Maximum Likelihood Estimation of fits of unimodal and multimodal Gaussian distributions to the data sets of of rips and zips and of rip and zip work energies, finding the intersections and other defining characteristics of multimodal Gaussian distributions, generation of scatter plots, generation of theoretical plots of the stretching of the closed and open pseudoknot structure using the EWLC model, statistical analysis of various possible subdistributions of rips and zips, and even the simple analysis of average effective trap stiffness.

However, as noted, Matlab could not be used for fitting the power spectrum data with an unweighted fit, so the conversion factor β for finding bead displacement from the trap center in terms of distance rather than volts was calculated using a script written in IgorPro by Jesper Tholstrup.

Additionally and importantly, a script written by Jesper Tholstrup in Python were used to generate theoretical predictions of the rip and zip lengths using the EWLC model, as will be described in more detail below. A bootstrapping script by Jesper Tholstrup written in Python was also used to estimate the uncertainty in the intersection point found between intersecting Gaussian distributions based on the uncertainty in their parameter values.

This allows estimation of the uncertainty in the values found for ΔG using the Crooks Fluctuation Theorem.

Finally, Gnuplot was used for several least squares fitting routines, particularly to fit the kinetic data to Equation (3.7) and Equation (3.9). Some of the preprocessing of data for these fits was still performed in Matlab, i.e., finding rip and zip forces and associated survival probabilities as well as estimating values of $\tau(F)$ from histograms of rip forces.

Detailed data analysis routines will not be provided here. It is worth noting, however, that fortunately most of the data routines were cross checked with the routines written to perform the same tasks by Jesper Tholstrup in Python and IgorPro. By ensuring that the same results were obtained for the same data sets using the two types of routines we have both tried to provide a control that the data routines were not too plagued by undiscovered bugs.

4.5.1 Using the EWLC to predict force-extension curves and rip/zip lengths

As described in Chapter 3.2, predictions of the force-extension curves for the closed and open expected pseudoknot structures were made using the Extensible Worm Like Chain. The predictions were made with fixed parameter values for L_c , L_p , and K_0 for the double and the single stranded portions of the molecular construct (the digoxigenin-anti-digoxigenin and streptavidin-biotin linkers were considered inelastic in these calculations).

This author used a relatively simple Matlab routine to generate values of the expected force-extension curves by solving for the extension that minimized the difference between the right and left hand sides of Equation (3.2) for a series of fixed force values. These predictions were made using the EWLC parameter values specified in Chapter 3.2. Subsequently, in order to calculate W_{stretch} , the energy needed to stretch the handles and the pseudoknots during a pull, the predicted EWLC curves for the appropriate double or single stranded portions of the construct were integrated to give the energy required to stretch that portion of the construct between the appropriate forces.

The predicted force-extension curves could also be used to give a rough indication of expected rip and zip lengths at high forces. However, because the calculation of rip length based simply on static force extension curves assumes that the same force occurs at the start and the end of the rip, the length prediction made is an overestimate, especially at low forces. In reality, any change in construct length is accompanied by a change in force. For a change in force, a new construct length needs to be calculated, and as the change in construct length is rather large at low forces, proper prediction of expected rip length requires an iterative script. Such a script was implemented by Jesper Tholstrup and used by this author to generate the curves of predicted rip length that will be shown in Chapter 5.

Crucially, the calculations made to predict both force-extension curves and rip/zip lengths is made based on the a priori assumption of the pseudoknot's structure presented in Figure 4.1. The stem and loop lengths as well as the lengths of the surrounding single and double strand portions are needed. The EWLC is then used to predict the change in extension that is expected on the explicit assumption that one or both stems of the pseudoknot open at a given force, causing a change in the double and single stranded lengths of the construct of a particular number of base pairs or nucleotides.

Note additionally that any experimental mismatch in the z-height of the traps during the pull would make the measured force-extension curve appear to have a slightly lower persistence length, L_p [67]. This was a relevant concern based on the Nanotracker particularities noted above (the beads were almost certainly misaligned in z height during experiments). Jesper Tholstrup found that the error caused by this in the force-extension curves generated was on the order of a few percent, less with increasing force [67].

4.6 Summary of methods

This chapter has presented the main experimental and (briefly) the main numerical methods used in this thesis. Detailed protocols for some of the experimental routines may be found in the appendixes while data processing routines may be obtained directly from the author. The next chapter, Chapter 5, will present the first batch of the experimental results: measured force-extension curves and observed rip and zip lengths and forces. Chapter 7 will present the remaining results, those deriving from work and kinetic analysis of the rip and zip data.

5 Results Part I: Rips and zips

As described in Chapter 2 and Chapter 4, the raw data for this thesis were voltage outputs from the Nanotracker's photodiodes. These were recorded during each attempt to pull a molecule by moving one optical trap relative to the other and were converted to graphs of force vs extension. The force-extension curves that showed the presence of a single molecular tether which displayed clear molecular transitions (rips or zips) were the basis of further investigation.

In this chapter, typical force-extension curves for the two pseudoknots will be presented followed by an overview of the number of molecular tethers observed, the number of pulls with rips, etc. The distributions of forces at which rips and zip occurred and the lengths of the rips and zips will subsequently be presented.

To refresh the terminology, a "*tether*" is a molecular construct linking two beads. A "*pull*" is a recording of the elongation and relaxation of a molecular tether. It may also be an attempt to pull a molecular tether that breaks or an attempt in which there turn out to be several molecular tethers between the beads. A "*rip*" or "*zip*" is an abrupt change in molecular extension on the force-extension curve that is assumed to indicate the opening or closing of the pseudoknot.

Throughout this chapter and the next the uncertainties indicated by " \pm " are the standard error of the mean (SEM) unless otherwise noted.

Histograms of data shown in this chapter and the remainder of the thesis are depicted with actual data counts on the y-axis rather than relative counts. This has been chosen to explicitly show the number of data points included. Probability distributions shown together with the histograms are therefore multiplied by a constant to show up in the figures, although this is not always noted in the figure captions.

Note additionally that for all experimental force extension curves shown, the scale on the x-axis, i.e., the extension, has an arbitrary zero as the initial separation of the beads is not known (see Chapter 2.3.4).

5.1 Typical force-extension curves

Figures 5.1 and 5.2 show examples of force-extension curves recorded for PK 11/6 and PK 6/11. The curves have been chosen to display samples of typical rip and zip patterns observed. Some rips are short and at low force, some far longer and at high force. Some curves show two rips during molecule elongation, indicating a two-step transition from open to closed pseudoknot, others show only a single rip, a rip followed by a zip and another rip during elongation, or other more complex pathways. The criteria used to distinguish rips and zips from noise will be presented below. When several rips and rerips occur during elongation of the molecule or several zips and rerips occur during relaxation it is called "*hopping*". In the range of forces where hopping occurs, the state of the molecule is apparently multistable.

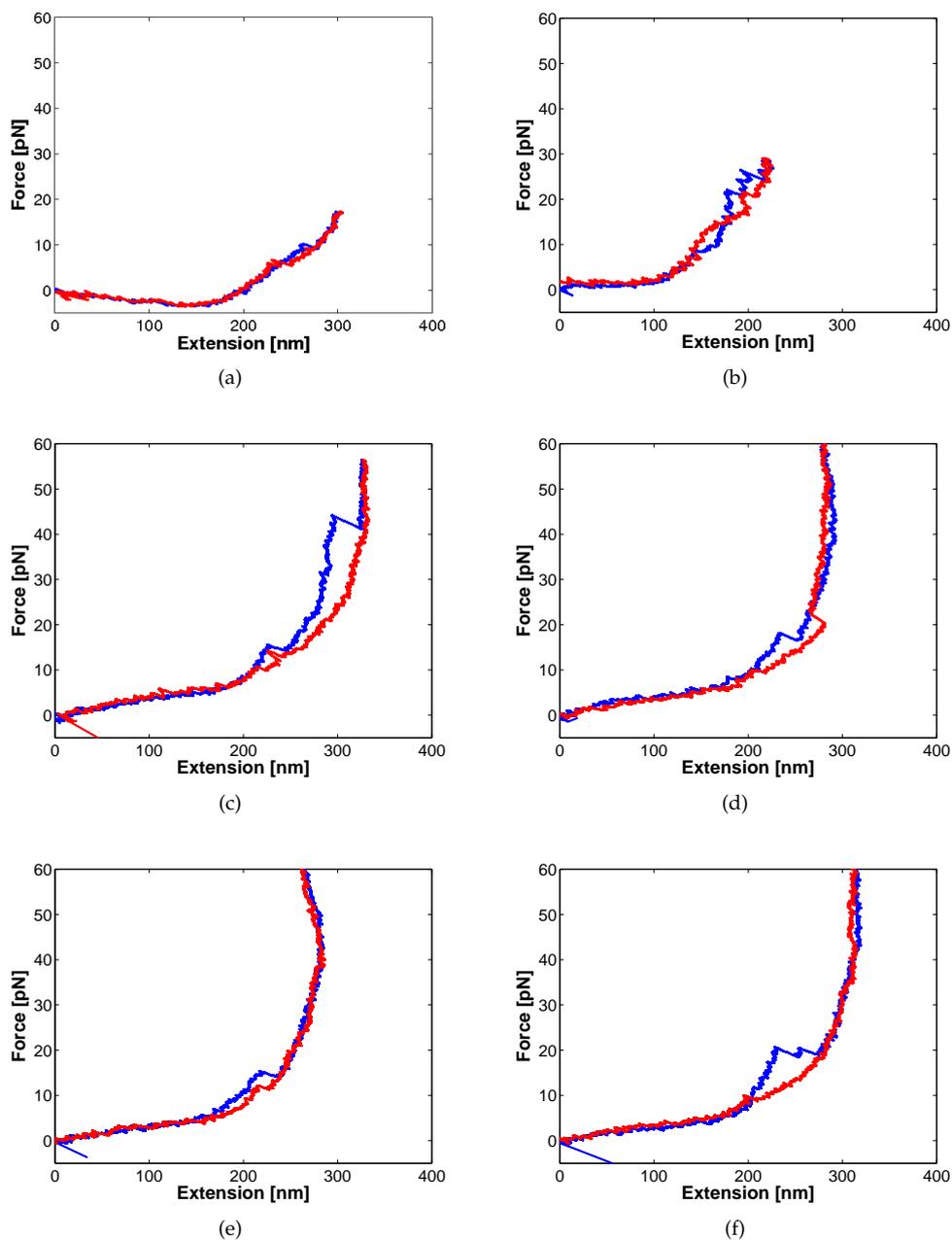


FIGURE 5.1: Examples of force-extension curves with rips and zips observed for PK 11/6. Blue: elongating the molecule (moving the beads apart). Red: relaxing the molecule (moving the beads together). (a) Clear rip and zip at low force. (b) Rips/zips in several steps with hopping: the pseudoknot opens, closes, and reopens. (c) Two-step opening of the pseudoknot. (d) Clear rip during molecule elongation, extra rip during molecule relaxation. (e) Zip during refolding constitutes the first step of a two step closing transition, the remainder occurs gradually. (f) Two-step opening of the pseudoknot, gradual closing. Pulls (c)-(f) originate from the same molecule and were recorded sequentially within a short time span.

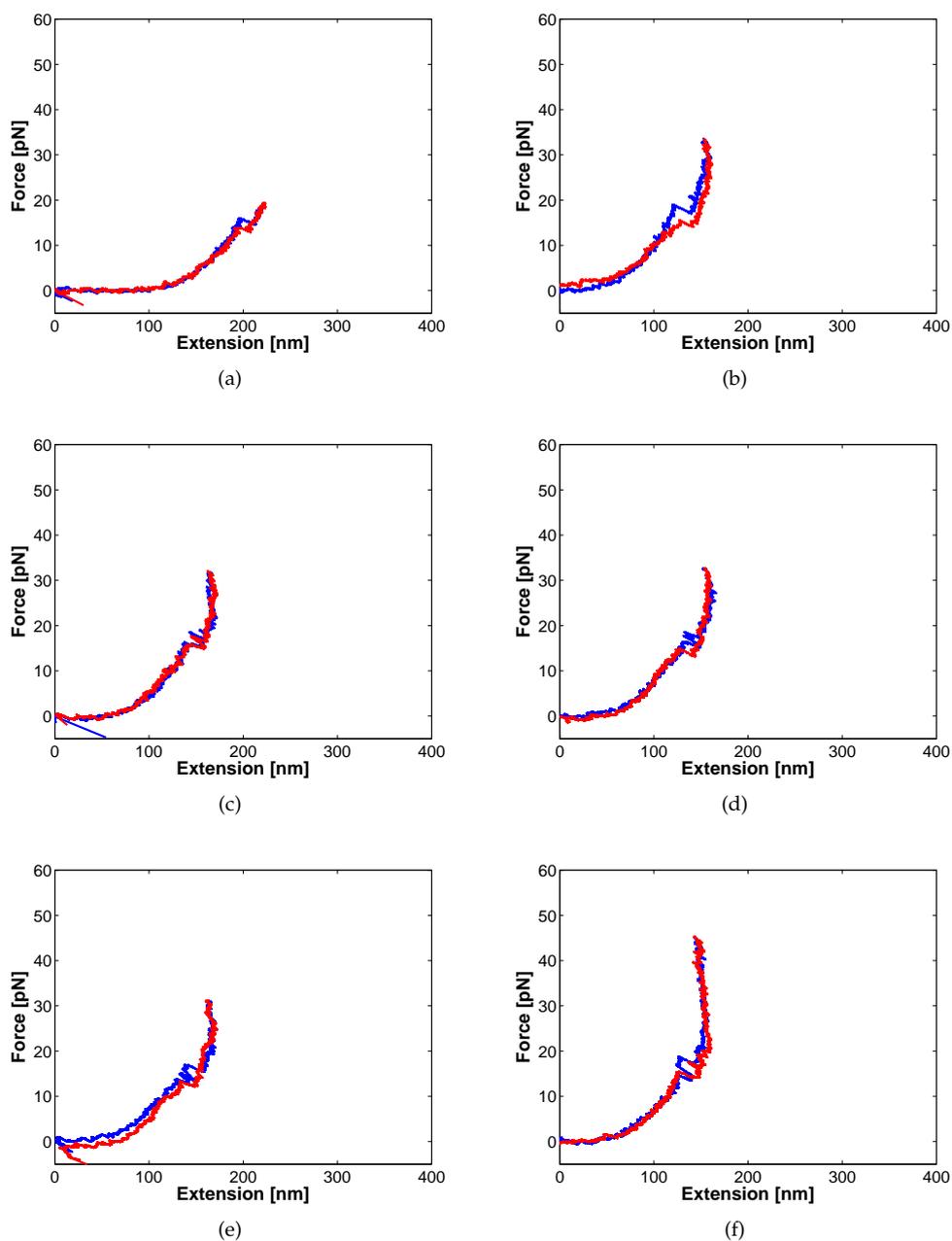


FIGURE 5.2: Examples of force-extension curves with rips and zips observed for PK 6/11, displayed on the same axes as the curves for PK 11/6 in Figure 5.1. Blue: elongating the molecule. Red: relaxation. (a) A clear rip and clear zip. (b) One-step rip during elongation, first step of a two-step zip and gradual completion of closing during relaxation. (c), (d), (e) and (f): Complex opening and closing transitions with hopping; see zooms in Figure 5.4. Pulls (b) through (f) originate from the same molecule and were recorded sequentially within a short time span; note similarity in appearance.

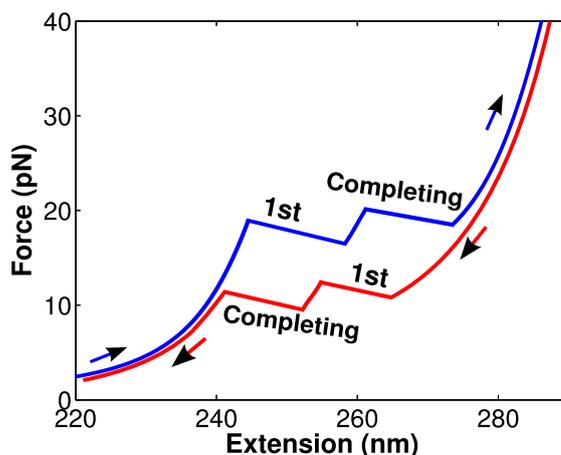


FIGURE 5.3: Labeling of the first and completing rips and zips of hypothetical two-step opening and closing transitions. The first (1st) step during elongation (blue) is the one at lowest force while the 1st step during relaxation (red) is the one at highest force. The second step is called the “completing” step as some hopping might occur before it is observed. Scale on x- and y-axis from a theoretical plot of the EWLC-model of the closed and open pseudoknot constructs as described in Chapter 3.2.

The rips and zips observed were classified as *one-step* or *two-step* transitions from the assumption that pseudoknot opening would occur either in one single step or in two steps corresponding to the opening of first one pseudoknot stem and loop, then the other (see original hypothesis in Chapter 1.4). As illustrated in Figure 5.3, the first step of a two-step rip on the elongation curve (sometimes called two-step opening) was the one occurring at the lower force. The first step of a two-step zip on the relaxation curve was the one occurring at the higher force. The following step of a two-step transition is called the “completing” step because it might occur after a bit of hopping rather than straight after the first step.

To better display the differences in force and extension of the rips observed, the force-extension curves in Figures 5.1 and 5.2 are all shown on the same scale. The details of some of the multistep and hopping transitions observed for PK 6/11 in Figure 5.2 (c)-(f) are shown in magnification in Figure 5.4. Hopping was also observed for PK 11/6 but less often. Note that, e.g., in Figure 5.4 (c) and (d) it is not completely obvious whether we are seeing part of a two-step transition or a full one-step transition.

Multistep transitions were characterized as the 1st step or the completing step of a two-step transition even when only one clear step was observed as long as it was obvious that the clear step was part of a larger transition. This could occur for instance because the other step was obscured as a gradual transition (Figure 5.1 (c) and (e), Figure 5.2 (b)) or because the structure hopped back to the original state before a longer transition was observed (Figure 5.4 (b)). *For this reason the numbers of 1st and completing two-step transitions in the later analysis are not the same.*

In some experiments, long series of pulls were recorded for the same molecular tether (up to 250 sequential recordings in one case), while in most experiments the tether broke almost immediately. For each pull, a separate force-extension curve was generated. Many force-extension curves were recorded which showed no clear rips or zips during molecule elongation and relaxation. Others showed large rips but no zips or more gradual

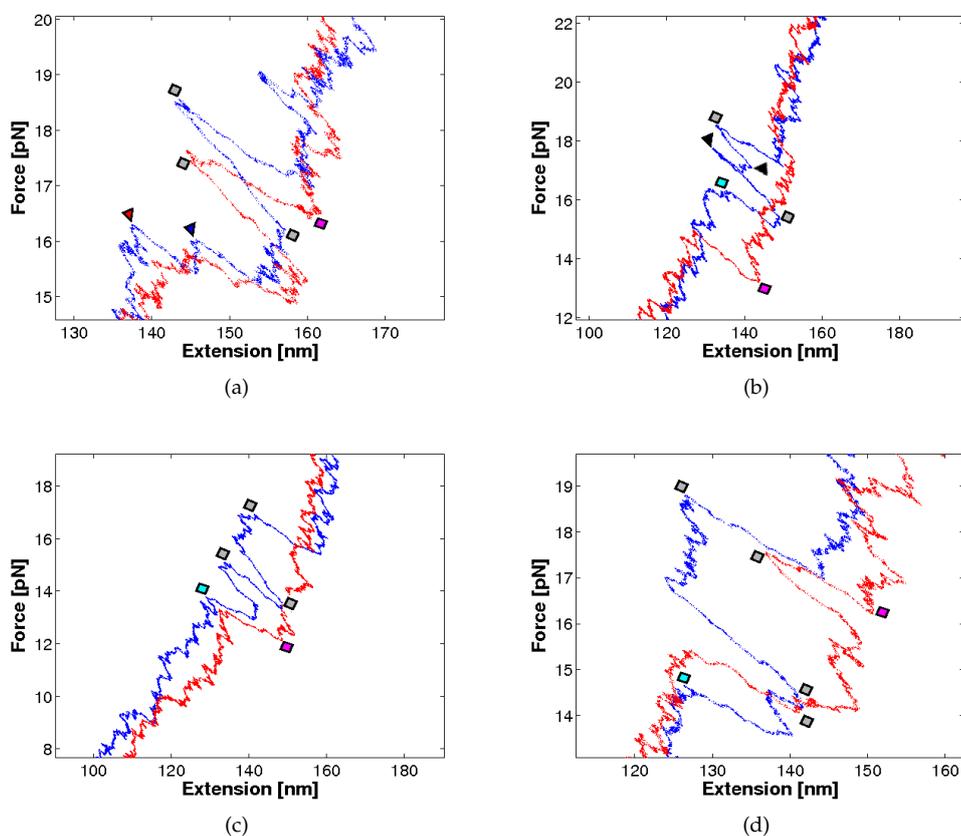


FIGURE 5.4: Force-extension curves for PK 6/11 with complex stepping and hopping: Details from Figure 5.2 (c)-(f). Note different levels of zoom/different scales. Blue: elongating the molecule. Red: relaxation. (a) Two-step opening, then hopping back to closed state and to intermediate state before the open structure continues elongating during the remaining pull; similar hopping during relaxation. (b) One-step opening, hop back to closed state, 1st step of two-step opening, then hop back to closed state before a full one-step opening during the elongation of the tether; closing in a single one-step zip. (c) and (d) Hopping where all transitions are classified as one-step rips or zips. As noted in Figure 5.2, all four pulls originate sequentially from the same molecule. The beginnings of each transition used in further analysis of rips and zips is marked by a little geometrical shape. Red triangles are first parts of two-step rips, blue triangles are completing two-step rips, black triangles are partial hop transitions. Blue boxes are first one-step rips, pink boxes are first one-step zips and gray boxes are one-step hop transitions. The same symbols will be used in Figures 5.8 and 5.9.

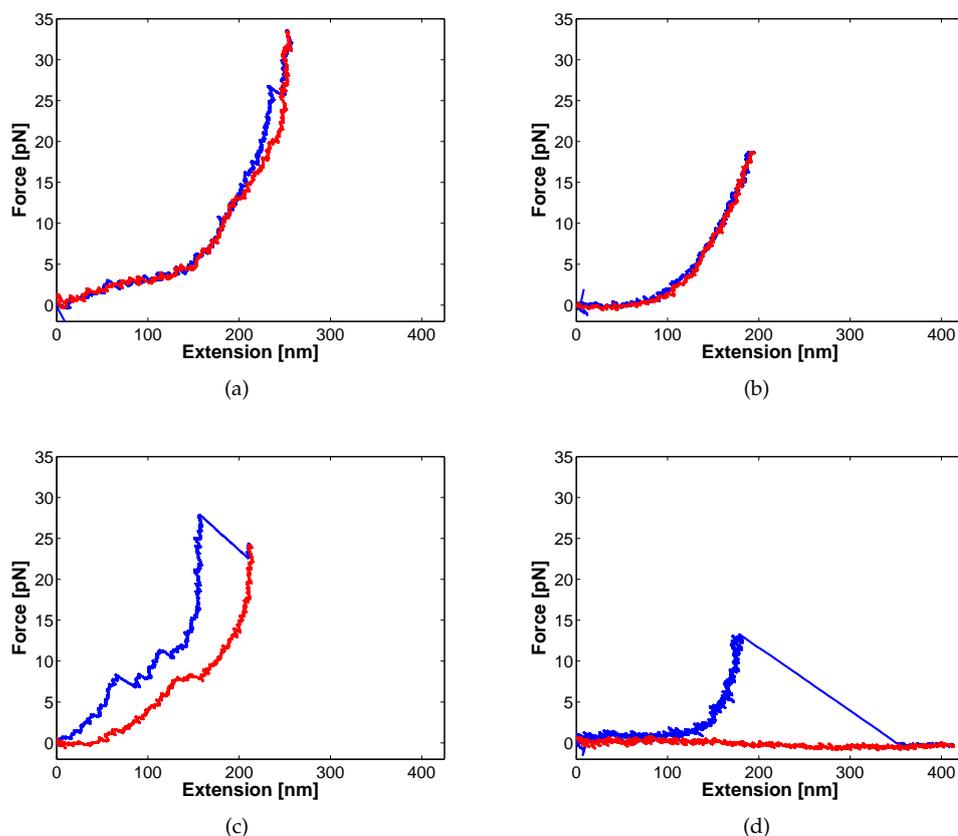


FIGURE 5.5: Examples of force-extension curves highlighting that rips and zips do not occur in all pulls. Blue: elongating the molecule. Red: relaxing the molecule. (a) A clear rip and gradual closing of PK 11/6. (b) A PK 6/11 tether elongates and relaxes without a rip. (c) Multiple PK 6/11 tethers: At least two tethers have been caught and one breaks mid-pull at ≈ 27 pN. Notice that the elongation and relaxation curves do not coincide at low force the way they do in (a) and (b). (d) A PK 11/6 tether breaks during the pull.

refoldings, which was interpreted as evidence that multiple molecular tethers had formed and one or more of them ripped during the pull. Figure 5.5 displays some examples of pulls with and without rips, with multiple tethers and with a tether that breaks.

The tethers for which many sequential pulls were recorded showed a pattern that is not easily quantified: A number of pulls with similar transition pathways would be observed sequentially (e.g., high force long rips) followed by a number of pulls with another type of transition pathway (e.g., low force short rips). See Figure 5.2 (b)-(f) for one such series. The force-extension curves in Figure 5.1 (c)-(f) were also recorded sequentially for a single molecule within a similar timespan, but the rips in these curves vary much more in appearance. With the full sequence of force-extension curves for this molecule, however, sequential patterns of force-extension curves with similar appearance are also observed.

5.2 Criteria for accepting rips and zips for analysis

As described above, rips (openings) and zips (closings) were defined as clear, abrupt transitions on the force extension curves such as the transitions shown in Figures 5.1 and 5.2. Distinguishing rips or zips from noise was not always easy, however. See Figure 5.6 for an illustration of force-extension curves in which it may be difficult to discern rips.

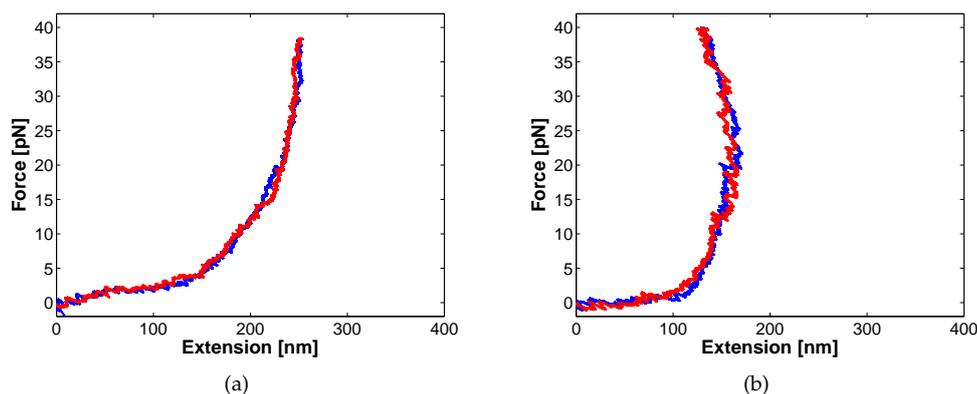


FIGURE 5.6: Two examples of pulls where discerning rips and zips from noise may be difficult. Blue: elongating the molecule. Red: relaxation.

The criteria used to filter rips and zips were:

1. The rip or zip observed had to be unusually large relative to the inherent noise of the force extension curve in that region.
2. A transition from one equilibrium curve to another had to be apparent beneath the noise. This meant that relatively large sudden changes in force and extension were not included if the curve immediately returned to its original path.
3. The transition had to be abrupt enough that the slope force extension curve was approximately equal to the trap strength, $-\kappa$ (estimated by eye). This follows from the fact that an immediate change in the force had to be accompanied by a proportional change in extension according to $F = -\kappa x$. Thus the slope of the force extension curve during a rip or zip had to be negative. This meant that:
 - gradual transitions were never included as rips or zips
 - if both rips and zips occurred on the same plot, they had to be approximately parallel.
4. The curve had to display both an opening and a closing, though one of the transitions (most often the closing) might occur gradually. In other words, the transition from the "open-state" curve to the "closed-state" curve had to occur before the force dropped all the way to the baseline level during molecule relaxation.
5. Two-step transitions had to display a clear pause at an intermediate force-extension curve before full opening or closing (see Figure 5.3 for a theoretical illustration and Figure 5.4 (a) for a real example). In other words, three distinct "states" of the molecule's force extension curve had to be observed.

The fourth criterion was made to exclude rips which simply represented the breaking of one tether in a multi-tether pull. Contrast Figure 5.5 (a) and Figure 5.5 (c). During control experiments performed by Jesper Tholstrup using molecular constructs with no pseudoknot-forming sequence on the single-strand RNA, large rips were sometimes observed but never with an apparent refolding back to the original curve [67]. This was interpreted as the breaking of one out of several molecular tethers linking the beads, since once broken it would be unlikely for the tether to form again before the beads were back at the baseline level [67].

The fifth criterion means that if only one stem of the pseudoknot unfolded at all during a pull, that one unfolding would be classified as a one-step rip. A scenario in which one stem of the full pseudoknot unfolds while the other stays intact is expected to be unlikely based on the assumption that once one stem has opened, the other is basically a simple RNA hairpin, which should be much weaker than the full pseudoknot and therefore relatively easily opened. However, it is possible that only one hairpin rather than a full pseudoknot was formed at the beginning of the pull, so that no more than a single hairpin could be opened at all.

The assumption behind the idea of a transition between equilibrium curves is that the molecule's force extension curve (criterion two) is approximated by the Extensible Worm-Like Chain model (EWLC) as discussed in Chapter 3.2. Thus when it transitions from closed to open and back it passes from one EWLC-curve to another with a different contour length, persistence length and stiffness as shown in Figures 3.3 and 5.3. From the examples in Figures 5.1, 5.2, 5.5, and 5.6 it is seen that in some cases rips and zips are easily identified and delineated and in some cases not.

5.3 Data overview

Table 5.1 gives an overview of the data obtained in this work. The number of pulls with rips and/or zips is the maximum number of uncorrelated openings and closings of the pseudoknot that may be used for further analysis. As mentioned above, the transitions in a given pull may actually not have been completely uncorrelated with the transitions in previous and subsequent pulls. However, transitions in different pulls are certainly less correlated than transitions measured during the same pull.

The number of experiments with rips and/or zips in Table 5.1 is the minimum number of separate molecules observed that displayed abrupt opening or closing transitions. More molecules may have been observed, as in some experiments a molecular tether broke but another later formed which could be the same molecule or a different one. The number of separate tethers indicate how often this occurred. It is the maximum number of separate molecules for which pulls without breakages were observed.

Table 5.1 thus gives an idea of how many molecules were observed and how likely it was that a rip or zip would be observed once a molecular tether had been caught, extended, and relaxed without breaking. The percentage of molecular tethers that did not break and which contained rips or zips may have depended on e.g. pull distance (varied during the experiment) and precise loading rate (though the piezo stage was always moved at 0.1 $\mu\text{m/s}$, the actual loading rate varied with the trap strength and had a standard deviation of a little over 10 %). Nevertheless this percentage could give a rough indication of the strength of that particular pseudoknot: A low percentage of tethers with rips and zips out of the total number of tethers that do not break would indicate that the closed state is relatively favorable. However for the two pseudoknots here the difference in the percentage of pulls that exhibit rips and/or zips out of the total number of pulls with unbroken tethers is not significant (tested with a two-tailed Student's t-test, $p=0.75$).

TABLE 5.1: Overview of experimental data

	PK 11/6	PK 6/11
Experiments w tethers ^a	14	23
- of which w rips/zips	8	6
- %-age with rips/zips	57 %	26 %
Separate tethers ^b	43	26
- of which w rips/zips	20	9
- %-age with rips/zips	47 %	35 %
Pulls w unbroken tethers ^c	429	306
- of which w rips/zips	171	64
- %-age with rips/zips	40 %	21 %
Total number of rips/zips	411	227
Average rips/zips per pull w rips/zips	2.4	3.5
Number of "hop" rips/zips	101	80
%-age hop rips/zips	25 %	38 %
Pulls w multiple tethers (minimum)	51	15
Pulls w broken tethers	205	152
Total pulls w tether(s)	685	473
- %-age with rips/zips	25 %	14 %
Mean loading rate (pN/s) \pm std	9.1 \pm 1.3	10.8 \pm 1.4

^a Experiments in which RNA tethers were caught, i.e. minimum number of separate molecules observed

^b Maximum number of separate molecules observed. During some experiments a tether broke after a number of pulls but another tether was observed later in the same experiment with the same beads. This is considered a separate tether, though it may or may not be the same molecular construct.

^c Total number of pulls recorded where the molecular tether does not break. May include pulls with multiple tethers where none are observed breaking.

The table also compares the number of pulls with rips and/or zips to the total number of rips and zips recorded and shows how many of these represent "hopping" rips or zips. This number could give some indication of how easily the structure reforms once broken during a pull. However, again the difference between the percentages of hopping rips/zips is not significant (tested with a two-tailed Student's t-test, $p=0.84$).

Figure 5.7 graphically breaks down the distribution of pulls in the categories of broken tethers, multiple tethers, tethers that neither broke nor exhibited rips, and tethers with rips for PK 11/6 and PK 6/11 (examples of pulls in each category were seen in Figure 5.5). Note that it was usually not possible to distinguish pulls with multiple tethers from single-tether pulls unless at least one tether broke during the pull.

The number of pulls with rips and/or zips observed per molecule ranged from one to 109 for PK 11/6 and from one to 50 for PK 6/11. As the total number of pulls with rips or zips were 171 and 64 respectively for the two pseudoknots, clearly further analysis of the rip and zip characteristics may have been biased by the circumstances of the particular experiments where many rips were observed. The possibility of such bias will be examined below in Section 5.5.

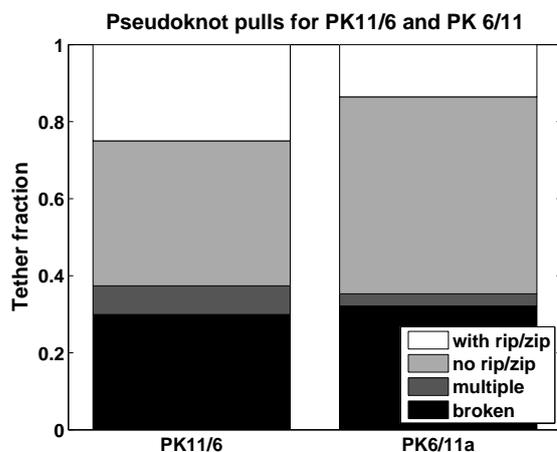


FIGURE 5.7: Distribution of pulls recorded among broken tethers, multiple tethers (of which at least one breaks, so that the pull is identifiable as such), tethers that do not break but also do not exhibit rips or zips, and tethers that exhibit rips and/or zips.

5.4 Rips and zips: forces and lengths

After filtering, the rips and zips identified were analyzed through examination of the forces at which they began and the change of construct length they represented. This was straightforward in the many cases where a single clear opening step was observed as the molecule was elongated and a single clear closing step or more gradual closing transition was observed during relaxation. Such transitions were seen in Figures 5.1 (a), 5.2 (a), and 5.5 (a). When the molecule displayed hopping or multiple-step transitions, however, this had to be taken into account in the analysis. For pulls with hopping, the first rip during the elongation of the molecule and the first zip during relaxation were treated separately from the remaining rips and zips (first rips and zips are here named according to the logic presented in Figure 5.3). This separation was made to allow analysis of uncorrelated rips and zips as indicated above. Likewise, two-step transitions were analyzed separately from one-step transitions since they were assumed to describe different types of molecular reaction.

One of the most illuminating ways to examine the data is to visually inspect the correlation of rip forces and rip lengths, zip forces and zip lengths in scatter plots. Such plots are shown in Figures 5.8 and 5.9. These plots are grouped according to the transitions that are observed during elongation and relaxation of the molecules. Different symbols are used for one-step and two-step rips and zips as well as for first "hop" rips or zips. Additionally the plots show curves predicting rip/zip length as a function of force for several possible structures of the single stranded RNA pseudoknot sequence. The predictions were made as described in Chapter 4.5.1 and will be discussed further in Chapter 6.

What stands out first of all is that the distribution of rip/zip forces and rip/zip lengths observed for PK 11/6 is radically different from the distribution of rip/zip forces and rip/zip lengths for PK 6/11. In particular, much higher rip forces and longer rip lengths are recorded for PK 11/6 than for PK 6/11. It is clear from Figure 5.8 (a) that these higher forces and lengths often, but not always, occur together, and that they relatively often occur as part of two-step opening transitions. The latter is not easily reconciled with the idea

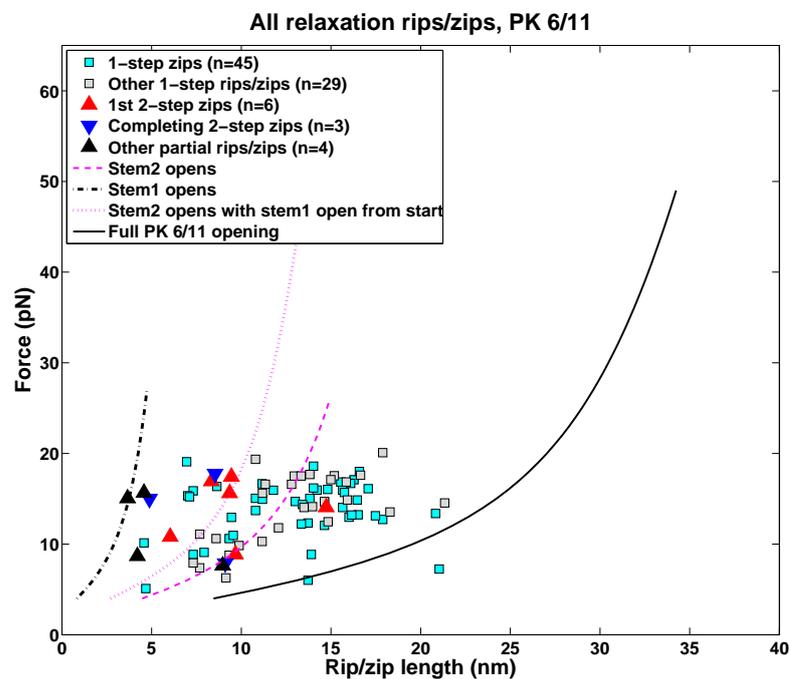
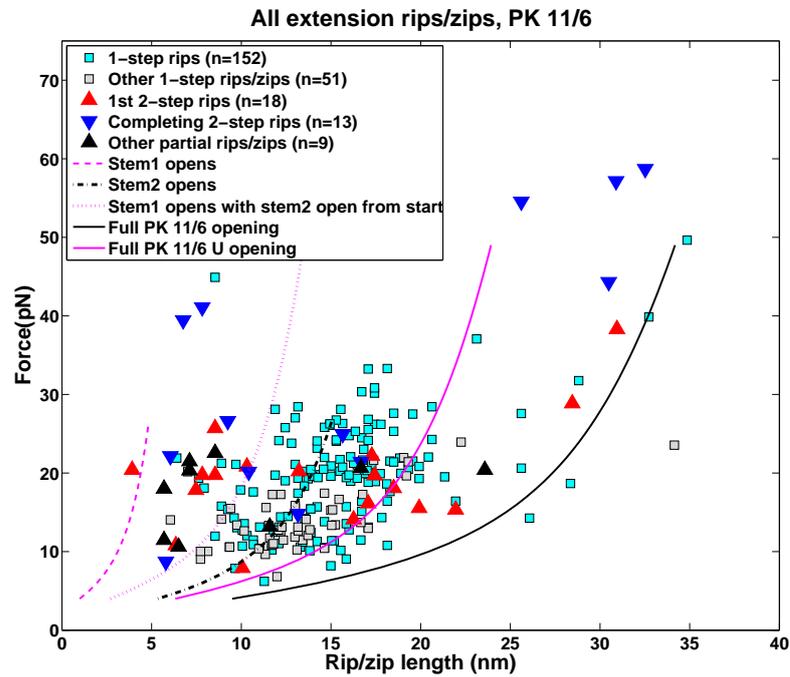
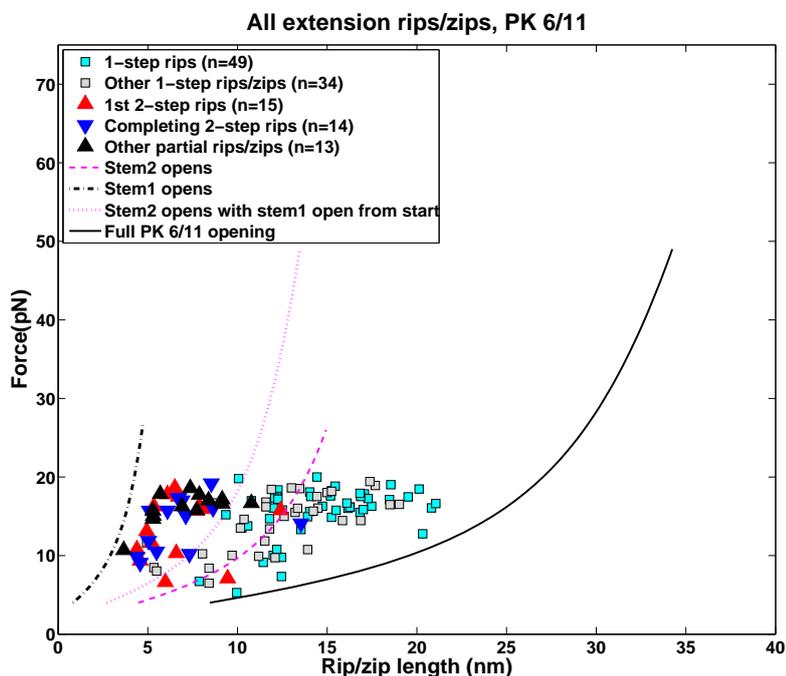
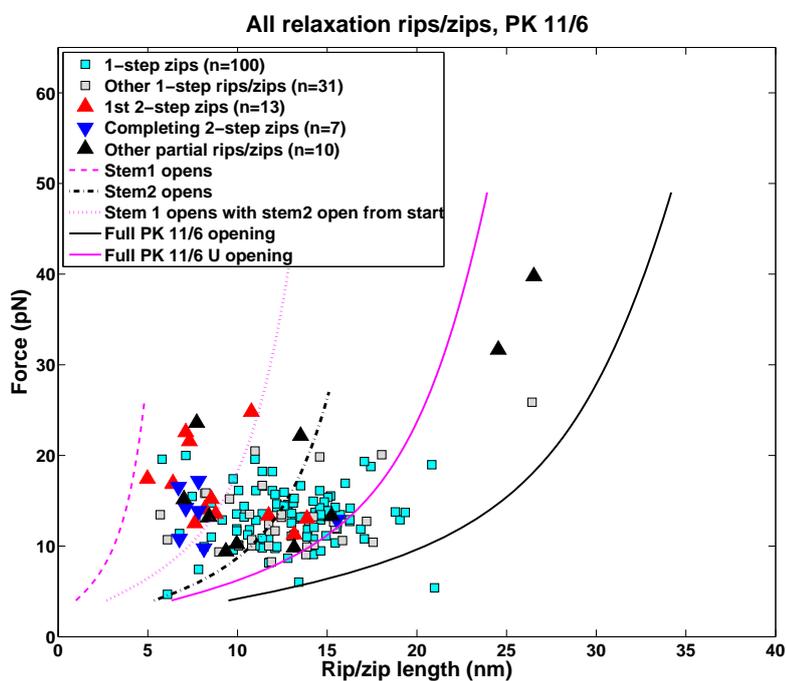


FIGURE 5.8: Scatter plots of rip/zip force versus rip/zip length for ALL *extension* rips and zips. (a) PK 11/6. (b) PK 6/11. Predicted curves made as described in Section 4.5.1. “Other” rips and zips refer to hop transitions.



(a)



(b)

FIGURE 5.9: Scatter plots of rip/zip force versus rip/zip length for ALL *relaxation* rips/zips. (a) PK 11/6, (b) PK 6/11. Predicted curves made as described in Section 4.5.1. “Other” rips and zips refer to hop transitions

that a two-step transition corresponds to one or the other of the stems of the pseudoknot unfolding. This will be discussed in Chapter 6.2.

Another striking difference between the data for PK 11/6 elongation rips and zips in Figure 5.8 (a) and the data in the other scatter plots is that while a main data cluster occurs in each plot, at least one extra cluster, possibly two, are observed for the PK 11/6 elongation rips and zips at higher forces and lengths. These additional clusters might well correspond to different molecular transitions. The main data cluster occurs at approximately the same forces and lengths in all four scatter plots - between about 5 and 25 pN (higher for PK 11/6 than PK 6/11) and between about 5 and 22 nm. This large cluster could again consist of sub-distributions of data. For instance it appears that the shorter rips and zips for PK 6/11 in Figure 5.8 (b) apparently occur at a wide range of forces while the longer rips all occur at fairly high forces. This might correspond to two different types of transitions.

It should here be noted that any rips and zips that might occur at lower forces than 5 pN and/or were shorter than about 5 nm could usually not be distinguished from noise using our rip/zip selection criteria. Also, the amount of data at higher forces may be limited by how often the molecular tethers broke at the digoxigenin/anti-digoxigenin link not far above 20 pN. However, some tethers were obtained for PK 6/11 that reached much higher force than 20 pN, but which displayed no rips or zips at these high forces. Lack of access to the high force regime is thus not the only reason that no rips or zips were recorded at high forces for PK 6/11.

An additional striking pattern seen clearly in Figures 5.8 and 5.9 is that for PK 6/11 elongation transitions and for all relaxation transitions, two-step transitions are generally (though not all) shorter than one-step transitions. This is emphatically not the case for the two-step rips and zips seen for PK 11/6 in Figure 5.8, where the two-step transitions are as widely scattered as the one-step transitions.

To a large degree the "hop" rip and zip data points (gray boxes, black triangles) lie in the same clusters and generally follow the same patterns as the uncorrelated first rips and zips observed during elongation and relaxation (blue boxes, red and blue triangles). Two observations stand out regarding "hop" rips and zips. The first is that the few long transitions observed during relaxation for PK 11/6 are all hop transitions. The second is that the hop rips and zips mostly occur at low forces, below about 20 pN. This could reflect that at lower force the single strand of the pseudoknot is under less tension and therefore more easily folds back into its original structure. It could also relate to the possibility that the lower force shorter rips and zips derive from a different type of molecular transition than the higher force longer rips.

All in all it appears from Figures 5.8 and 5.9 that the PK 11/6 transition from the closed to the open state occurs in a very different manner from the transition of PK 6/11. It is also clear that for PK 11/6, one-step opening transitions do not occur as simple sums of two-step openings. This contrasts to the pattern seen in the remaining scatter plots. The measured rip lengths will be compared to predicted rip lengths in Chapter 6.2, where an attempt will be made to explain the patterns described here. First, however, we will have a closer look at the distributions of rip lengths and use these to separate clusters of data which will in Chapter 7 be used to calculate the free energy of the transitions observed as well as kinetic parameters. We will also briefly look at some subsets of the data to determine how vulnerable they are to the observations made in individual experiments.

5.4.1 Exploring possible subdistributions of rips and zips

From the scatter plots in Figures 5.8 and 5.9 it is clear that the rip and zip data may derive from several distinct distributions rather than from a single distribution. Since distinct underlying distributions could well derive from different types of molecular transitions,

further analysis of the data required separating the distributions. This was only done for one-step transitions due to the low number of data for two-step transitions. The distributions of rip and zip *lengths* were considered most indicative of the distributions of the molecular transitions taking place, although rip force and rip length are correlated. The reason is that distinct rip lengths are expected to be associated with specific molecular transitions while rip forces are expected to display more random variation. Therefore clear evidence of several underlying rip/zip length distributions were used to divide the dataset for further analysis.

Distributions of rip and zip lengths with fits and residuals between data and cumulative fitted curves are shown in Figures 5.10 and 5.11. Note that the distributions are all quite similar except that a number of longer transitions are observed for PK 11/6, and relatively many short transitions are observed for PK 6/11, as can also be seen in the scatter plots in Figures 5.8 and 5.9. It is clear that the higher order fits result in smaller residuals, but also that they in some cases may overfit the data. For rips observed for PK 11/6 (Figure 5.10 (a), (c), and (e)), the second and third order fits appear much better than the first order fit, but for the other distributions this is not obvious, again much in tune with the impression gained from the scatter plots of force versus transition length in Figures 5.8 and 5.9.

Distributions of rip and zip forces are shown in Appendix E, Figures E.1 and E.2. Among these distributions, the rip forces for PK 6/11 most clearly stand out as multimodally distributed (compare Figure E.1 (b) to the other cumulative distributions). It is again difficult to evaluate by eye whether the other distributions are unimodal or multimodal, although the pattern of residuals for PK 11/6 rip forces does underscore that a single Gaussian does not describe this data well. Again, as seen in Figures 5.8 and 5.9, the forces observed for PK 11/6 elongation transitions are much higher than those observed in the other graphs.

Several methods were subsequently used to investigate the goodness-of-fit of the underlying distributions:

1. χ^2 -test of single Gaussian distribution: A simple χ^2 test was used to determine if a single Gaussian distribution could be rejected for the data. This χ^2 test relied on histogramming of the data using a built-in Matlab function.
2. Comparison of R^2 -values: χ^2 -testing of the double and tripple Gaussian fits to the overall distributions of data was difficult since the fits were made only using Maximum Likelihood Estimation, which does not give an individual data point variance needed to calculate the value of χ^2 . Instead the multiple correlation coefficient, R^2 , was calculated and used to evaluate the goodness of fit as described in Bevington and Robinson [70, pp. 205–207].
3. *F*-test of additional parameter validity: An *F*-test was used to test the inclusion of extra terms in the model. This test is also inspired by Bevington and Robinson [70, p. 207]; a more detailed description was found in [71].
4. *Aikaike Information Criterion*: Another standard method of comparing goodness-of-fit between several models, the *Aikaike Information Criterion* (AIC), was implemented following [72].
5. *Multivariate Gaussian Mixture Model fit*: A built-in Matlab function that fits a Gaussian mixture model to multivariate data was applied to the combined force-length data for rips and for zips to find underlying Gaussian distributions in two dimensions. This function also used the *Aikaike Information Criterion* to determine the number of underlying distributions needed to best fit the data.

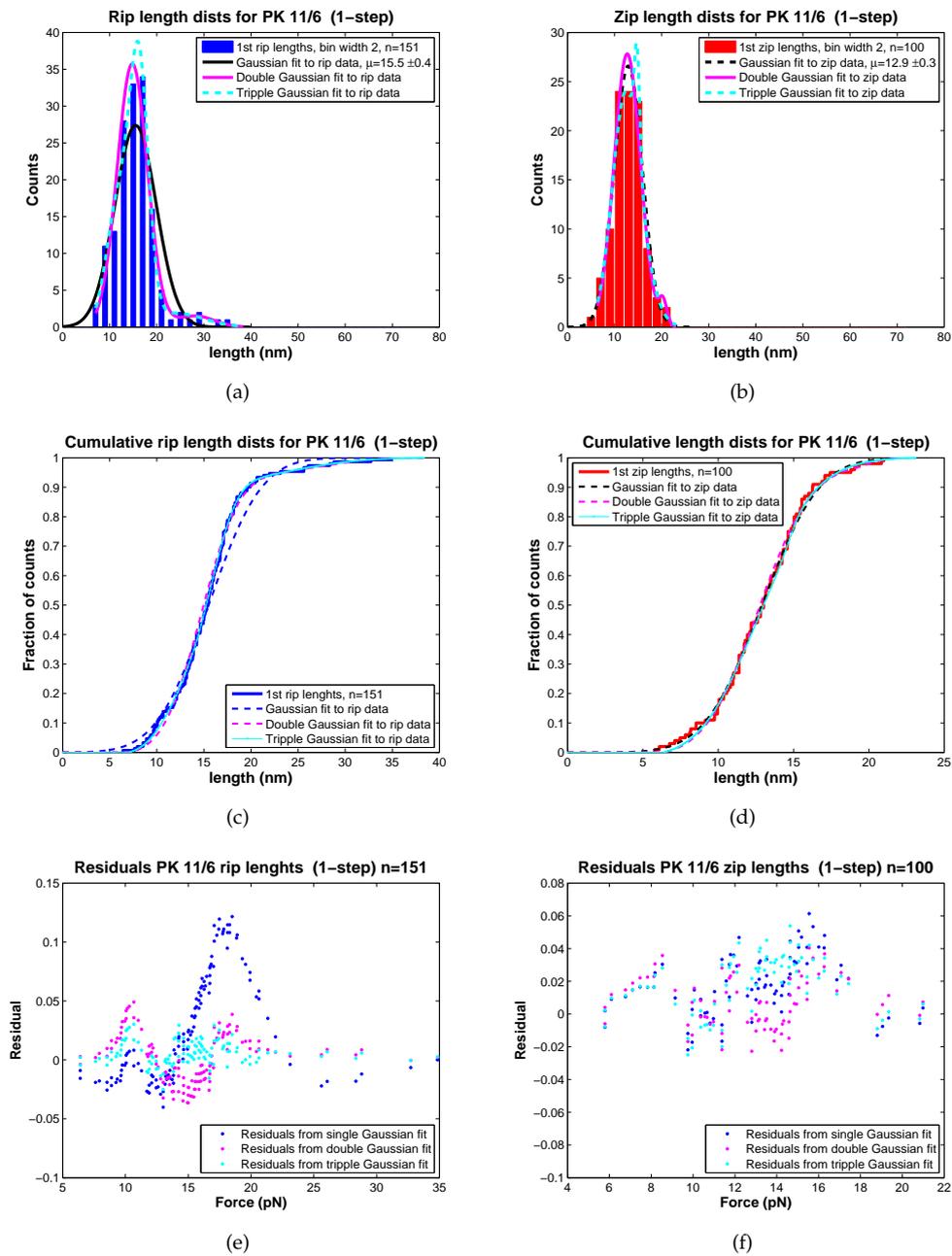


FIGURE 5.10: Rip (left) and zip (right) length distributions for PK 11/6: (a) Histogram of first one-step elongation rip lengths fit by single, double, and tripple Gaussians. (b) Histogram of first one-step relaxation zip lengths with single, double and tripple Gaussian fits. (c) and (d) Cumulative distributions of data and fits (note x-scales are adjusted so that maximum detail of the distribution can be discerned). (e) and (f) Residuals between fitted curves and experimental cumulative distribution.

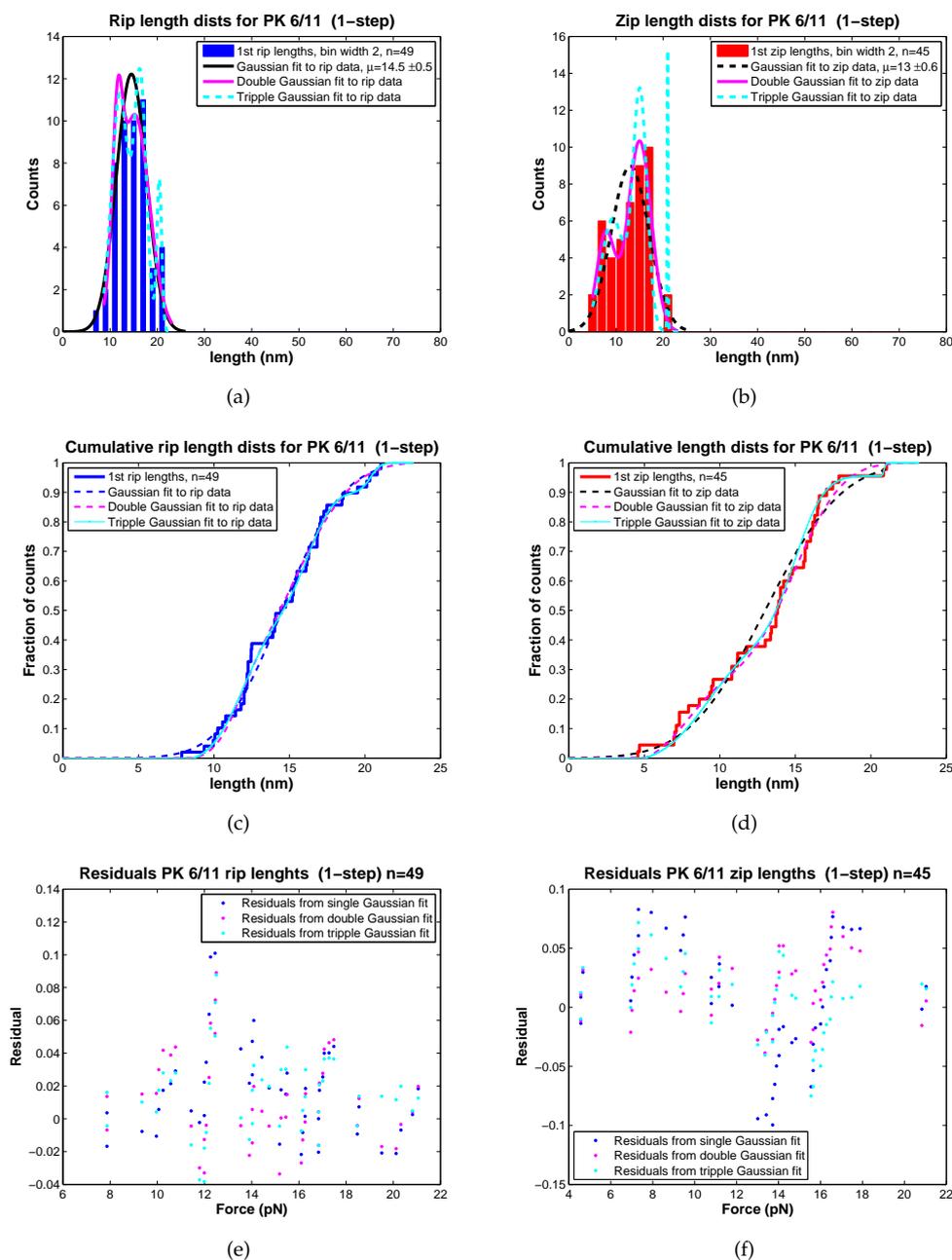


FIGURE 5.11: Rip (left) and zip (right) length distributions for PK 6/11: (a) Histogram of first one-step elongation rip lengths fit by single, double, and tripple Gaussians. (b) Histogram of first one-step relaxation zip lengths with single, double and tripple Gaussian fits. (c) and (d) Cumulative distributions of data and fits (note x-scales are adjusted so that maximum detail of the distribution can be discerned). (e) and (f) Residuals between fitted curves and experimental cumulative distribution.

6. *Visual inspection*: Visual inspection was used to compare the cumulative distributions of rip/zip lengths to the fitted distributions, not least through examination of plots of the residuals between data points and fitted cumulative distributions. The appearance of the scatter plots in Figures 5.8 and 5.9 was likewise an important guide.

In the end, visual inspection (Method 6) proved decisive in determining the number and locations of the underlying distributions used for further analysis. This was combined with the simple χ^2 test of whether a single Gaussian distribution could be rejected (Method 1) and calculation of the value of R^2 for each fit to evaluate the goodness of fit (Method 2). Though the other methods are less subjective, the problem with them is that they only compare goodness-of-fit but do not reveal whether any of the models are actually good descriptions of the data. As shown by e.g. Dudko et al [49], the expected distribution of rip forces for molecular transitions are not Gaussian in the first place. Thus although the statistical tests usually pointed to relatively high order Gaussian fits as the better fits (e.g. a quadruple Gaussian fit for the rip forces observed for PK 11/6), this might simply have to do with the non-Gaussian nature of the underlying distributions rather than the true number of underlying distributions. Method 5, multivariate fitting of Gaussian mixture models, was additionally problematic because the fits were not stable but depended on the random starting point values for the fits made by Matlab.

Tables E.1 and E.2 in Appendix E show the overall results of the various tests for different order fits for both lengths and forces. As stated, in forming the conclusions below, the two factors that weighed the most were the possible rejection of the single Gaussian distribution for rip *lengths* and the visual inspection of the scatter plots of force versus length (Figures 5.8 and 5.9). Detailed results of the tests are not included.

The conclusions are:

- *The distribution of first one-step elongation rips for PK 11/6 appears at least bimodal and possibly trimodal (or even higher order)*. This conclusion is drawn from the observation of the distribution of points in the scatter plot, Figure 5.8 (a), where at least two distributions of one-step rips (blue squares) are evident. A single Gaussian distribution is rejected for both rip lengths and rip forces, and the cumulative impression from the results of the other tests is that there is definitely more than one underlying distribution.
- *The distribution of first one-step relaxation zips for PK 11/6 appears unimodal*. This is based on the fact that a single Gaussian distribution is not rejected for the rip lengths (though it is for the rip forces, this is considered less important as discussed above). Also, crucially, visual inspection of the scatter plot reveals no more than one obvious distribution of one-step elongation zips. The other tests also point to no more than maximally two underlying distributions, indicating a simpler distribution than for PK 11/6 rips.
- *The distribution of first one-step elongation rips for PK 6/11 appears either unimodal or bimodal* Although a single Gaussian is *not* rejected for the distribution of rip lengths, the distribution of rip forces for these rips is unequivocally bimodal (Figure E.2). Visual inspection of the scatter plot in Figure 5.8 (b) likewise quite clearly hints at more than one underlying distribution of elongation transitions. Additionally, the alternative tests with only one exception point to a bimodal distribution.
- *The distribution of first one-step relaxation zips for PK 6/11 appears unimodal*. The reason is that a single Gaussian distribution cannot be rejected for either zip lengths or zip forces, and visual inspection also does not clearly indicate more than one distribution, partially due to the relatively low number of data points. The alternative tests do not give any clear picture of the number of number of underlying distributions.

5.4.2 Identification of rip subdistributions

The multimodal distributions identified for the one-step rip distributions above were used to divide the rip data into subdistributions. This was done through a very simple criterion: Is the rip length for this particular rip longer or shorter than the rip length corresponding to the intersection point of the estimated underlying Gaussian distributions? Clearly the subsets of data thus formed do not take into account the overlap that must exist between the distributions, but since there is no way of knowing exactly which points belong to which distribution in the overlap region, this simple division was deemed sufficient.

The resulting divisions of points are seen in Figures 5.12 (a) for PK 11/6 and 5.13 (a) for PK 6/11, as above presented together with predicted rip/zip lengths based on EWLC-modelling. The average force and extension for each subdivision of the data is shown in Tables 5.2 and 5.3 along with the predicted transition lengths at the experimentally found average forces. The measured average versus predicted rip lengths will be discussed further in Chapter 6.2. Figures 5.12 (b) and 5.13 (b) show scatter plots of the first zips on the relaxation curves recorded during the same pulls as the rips that are divided into subdistributions in Figures 5.12 (a) for PK 11/6 and 5.13 (a) (zips only displayed in cases where one-step zips actually occurred during the relaxation in the same pull). It is striking that the distributions of one-step zips deriving from the same pulls as the rip subdistributions are *not* divided into subdistributions of differing zip lengths but rather appear to overlap completely.

For PK 11/6, we will look in detail at the subdivision of points called distribution A (pink diamonds) in Figure 5.12 (a) and (b). Distribution A is slightly different in (a) and (c) due to the difference between fitting a double and a tripple Gauss to the same data, but still comprise basically the same data points. This subdistribution stands apart from the remaining data very distinctly and it seems intuitively reasonable to separate it out, especially since it also stands apart from the scatters of data points obtained for the PK 11/6 relaxation transitions and for all PK 6/11 transitions observed in Figures 5.8 and 5.9.

Further subdivision of the rip data for PK 11/6 as seen in Figure 5.12 (b) is less visually obvious. The subdistributions B and C could certainly belong to two subdistributions, and this idea will be explored further, but they could also be part of the same distribution.

Likewise the subdivision of the rip data for PK 6/11 in Figure 5.13 is not completely convincing, but not implausible either. Visually one is tempted to divide the data into a low force, short portion and a high force portion with a wider range of lengths, but due to the instability of the multivariate Gaussian mixture fit that was attempted (see above), it is hard to say if this approach would be better than the division based on rip length alone, which is shown here.

The subdistributions and their possible explanations will be further explored along with the promised discussion of the predicted rip lengths in Section 6.2. First, however, we will look at a different set of data subsets, namely those deriving from particular experiments which displayed "super tethers" (explanation will follow). These data subdivisions may not be as interesting as the ones explored above, but it is important to check how they influence the results.

5.5 Testing bias from "super tethers"

A few molecular tethers lasted much longer than the others and endured pulling to much higher forces with no breakage of the digoxigenin-antidigoxigenin link. These tethers are dubbed "super tethers" because they were so strikingly much more stable than the tethers that were usually caught. Three of these very durable molecular tethers also displayed

TABLE 5.2: Rip/zip forces and lengths for PK 11/6. The variable n is the numbers of rips or zips (forward or reverse transitions) analyzed. F is the rip/zip force, X is the rip/zip length. The error is the standard error of the mean (SEM) of the measured values. For multimodal distributions, n is the proportion of data points expected in that part of the distribution. This number is used to calculate the SEM for these distributions. The predicted rip/zip lengths corresponds to the average force for that distribution of rips/zips. Values closest to the measured average transition length is highlighted (dark for good correspondence, lighter for less clear correspondence or where there may be two possibilities).

	Data averages			Predictions (nm)			
	n	$\langle F \rangle$ (pN)	$\langle X \rangle$ (nm)	stem 1 ^a	stem 2 ^b	full PK 11/6 ^c	PK 11/6 U ^d
Mean of all 1st rips	169	19.8±0.5	13.5±0.3	4.3	13.9	28.3	17.6
1-step rips: 2G	145	19.4±0.5	14.8±0.3	4.2	13.7	28.0	18.1
	8	28.9±4.4	28.2±0.3	4.8	15.0	30.5	21.2
1-step rips: 3G	49	17.3±1.1	11.2±0.2	4.0	13.4	27.4	16.6
	89	20.1±0.6	16.1±0.2	4.3	14.0	28.4	19.7
	14	27.5±3.5	23.9±0.5	4.7	15.0	30.3	21.1
First of 2-step rips	21	19.8±1.4	16.0±1.2	4.25	13.9	28.3	17.6
Completing 2-step rips	13	33.4±4.7	13.6±0.3	5.0	15.5	31.4	21.9
Mean of all 1st zips	113	15.4±0.4	12.4±0.3	3.85	13.0	26.6	18.3
1-step zips	101	15.4±0.4	12.9±0.3	3.85	13.0	26.6	18.3
First of 2-step zips	16	14.4±0.7	9.3±0.8	3.8	12.8	26.2	18.5
Completing 2-step zips	7	16.2±2.8	8.6±1.2	3.9	13.2	27.1	18.6

^a Stem 1 of the originally expected pseudoknot structure opens/closes.

^b Stem 2 of the originally expected pseudoknot structure opens/closes.

^c All of the originally expected pseudoknot structure opens/closes.

^d All of the alternative pseudoknot structure, PK 11/6 U opens/closes. If only stem 1 or stem 2 of the alternative pseudoknot opens/closes, this is in practice indistinguishable from the predicted length of stem 1 of the originally expected pseudoknot opening. See discussion in Chapter 6.2.

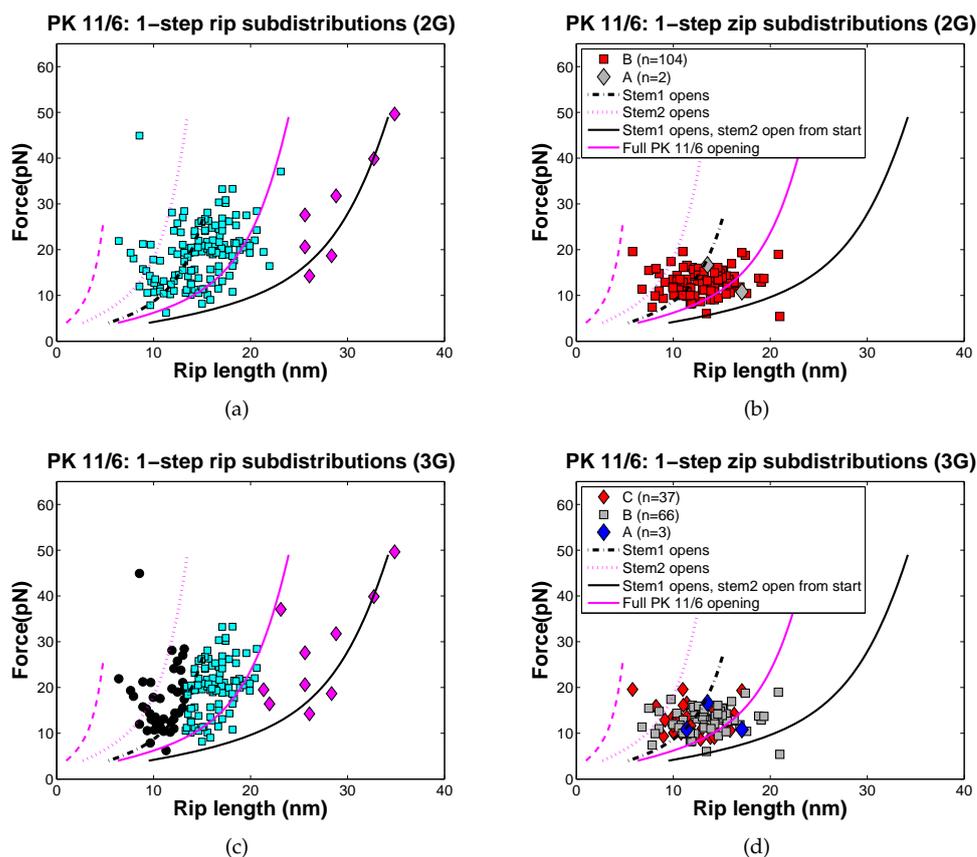


FIGURE 5.12: Subdistribution scatter plots for PK 11/6: (a) 2-Gaussian subdistributions of rip force versus rip length and (b) zip force versus zip length from same pulls as subdistributions in (a). (c) 3-Gaussian subdistributions of rip force versus rip length and (c) zip force versus zip length from same pulls as subdistributions in (a). Pink diamonds in (a) and (c) are from distribution A, blue squares are distribution B, and black dots are distribution (c). Curves representing predicted rip/zip lengths in (a) and (c) correspond to the labels used in (b) and (d). These are discussed in Chapter 6.2

TABLE 5.3: Unfolding/refolding forces and rip/zip lengths for PK 6/11. As for the previous table, the variable n is the numbers of rips or zips (forward or reverse transitions) analyzed. F is the rip/zip force, X is the rip/zip length. The error is the standard error of the mean (SEM) of the measured values. For multimodal distributions, n is the proportion of data points expected in that part of the distribution. This number is used to calculate the SEM for these distributions. The predicted rip/zip lengths corresponds to the average force for that distribution of rips/zips. Values closest to the measured average transition length is highlighted (dark for good correspondence, lighter for less clear correspondence or where there may be two possibilities).

	Data averages			Predictions (nm)		
	n	$\langle F \rangle$ (pN)	$\langle X \rangle$ (nm)	stem 1 ^a	stem 2 ^b	full ^c
Mean of all 1st rips	64	15.0±0.4	13.8±0.6	3.8	12.9	26.4
1-step rips: 1G	51	15.4±0.5	13.8±0.5	3.8	13.0	26.6
1-step rips: 2G	12	16.2±0.4	11.5±0.4	3.9	13.1	27.0
	39	12.8±1.3	15.2±0.2	3.5	12.2	25.3
First of 2-step rips	14	13.5±1	14.0±0.5	3.6	12.4	25.7
Completing 2-step rips	15	14.2±0.9	13.5±0.6	3.7	12.7	26.0
Mean of all 1st zips	51	12.7±0.4	12.6±0.6	3.5	12.3	25.1
1-step zips	46	14.5±0.5	13.0±0.6	3.7	12.8	26.2
First of 2-step rips	5	6.7±1.4	9.6±1.2	2.2	9.2	19.5
Completing 2-step rips	3	6.9±3	7.5±1	2.2	9.3	19.8

^a Stem 1 of the originally expected pseudoknot structure opens/closes.

^b Stem 2 of the originally expected pseudoknot structure opens/closes.

^c All of the originally expected pseudoknot structure opens/closes.

many more clear pseudoknot openings and closings than the others. These super tethers which survived many pulls with many clear rips or zips consequently somewhat dominate the data on rip and zip forces and lengths.

The possibility of bias from the conformation of particular molecules is a frequent issue in single molecule studies, where even such seminal articles as the experimental test of Jarzynski’s Equality by Liphardt et al. in 2002 [73] and the verification of the Crooks Fluctuation Theorem by Collin et al. in 2005 [46] rely on measurements of only seven and two to five molecules per experiment respectively. In the pseudoknot pulling study by Chen et al. from 2009 [21], the number of separate tethers used for data analysis hovers between 7-13 for each type of pseudoknot tested, and for one particular experiment data is presented which derives from only one experiment in which two tethers were caught (!).

Nonetheless it seems appropriate to examine the data deriving from the individual super tethers to see how they influence the results and whether we can learn something from the differences between the zip and rip patterns of individual molecules. This is done in detail in Appendix D. The most striking observation is that if one single super tether is left out of the data set for PK 11/6, the remaining data for PK 11/6 ends up looking very

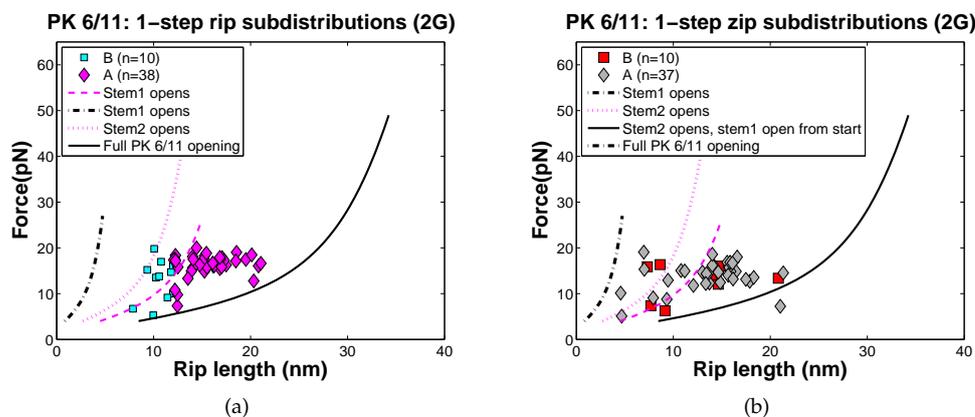


FIGURE 5.13: Subdistribution scatter plots, PK 6/11: (a) 2-Gaussian subdistributions of rip force versus rip length with predicted transition lengths and (b) zip force versus zip length from same pulls as subdistributions in (a).

similar to the data for PK 6/11: the distribution of rip forces is bimodal, the scatter plot of force versus extension is mostly clustered between 5 and 25 pN and between 5 and 25 nm with two longer exceptions, and two step rips are generally shorter than one-step rips. On average, rips still occur at higher force for PK 11/6 than for PK 6/11, but the cloud of long rips occurring at high forces observed in Figure 5.8 (a) are no longer present.

As might be expected, the calculated work energy, Gibbs free energy change, and kinetic parameters (Chapter 7) also change significantly if this one super tether is left out of the data set for PK 11/6. The values calculated for PK 11/6 then become much more similar to those calculated for PK 6/11. See Appendix D.

This does not mean that the full data set for PK 11/6 is not valid, but along with the knowledge that the number of rips and zips observed in total for PK 6/11 is much more sparse than for PK 11/6 and that only one super tether was observed for PK 6/11 compared to two for PK 11/6, these observations do imply that catching a few more super tethers for each molecule and thus getting much more data might elaborate and change the overall picture significantly. Note, however, that “catching a few more super tethers” for each pseudoknot would probably mean spending a month or two in the lab as their occurrence is so rare.

5.6 Summary of points made in this Chapter

So far we have seen that:

- The data representing the elongation rips/zips observed for PK 11/6 are radically different from those observed for PK 6/11, showing longer rips and zips occurring at higher forces for PK 11/6.
- All of the relaxation rip/zip data as well as the elongation rip and zip data for PK 6/11 and a large portion of the elongation rips/zips for PK 11/6 appear to cluster in a main region between about 5-25 pN and between about 5-22 nm.
- For PK 11/6 it is very difficult to see a pattern in the two-step rip and zip transitions observed. For the remaining data sets (both sets of relaxation data and the PK 6/11

elongation data) the two-step transitions appear generally shorter than the one-step transitions, corresponding to the notion that two two-step transitions may simply add up to a full one-step transition

- The one-step elongation rip data appears best fit by bi- or even trimodal subdistributions. The distinction between long rips and the remaining cluster of rips observed for PK 11/6 is especially clear.
- The difference between the rip data set for PK 11/6 and PK 6/11 largely stems from data from a single experiment. This means that the data are somewhat dependent on rare experimental conditions.

These facts will be investigated to some degree in the next chapter before we move on to looking at energy and kinetic parameter calculations for the two pseudoknots.

6 Discussion Part I: possible transitions and intermediates

In this chapter the distributions of rips and zips found in the previous chapter will be interpreted in more detail. First the observed rip forces will be discussed briefly and then the observed rip lengths. The observed rip lengths will be compared to predicted rip lengths that were shown already in Figures 5.8 and 5.9 and the possible subdistributions of transitions will be discussed.

6.1 Rip forces

Two obvious differences between the unfolding force distributions of the two pseudoknots stand out. The first is that the distribution of unfolding forces include much higher forces for PK 11/6 than for PK 6/11 and that the average unfolding force is also higher for PK 11/6 than for PK 6/11. The second is that the unfolding distribution is unimodal for PK 11/6 but bimodal for PK 6/11.

Both of these observations reveal that the two pseudoknots are different in ways that are not immediately apparent from the sequence of nucleotides they are made up of. The originally expected structures of the pseudoknots shown in Figure 4.1 are almost symmetric in their most prominent features, the stem and loop lengths. That reversal of the stems and loops would alter the way a ribosome interacts with the pseudoknot – and thus the frameshift efficiency – seems unsurprising, since the ribosome always encounters the pseudoknot from the 5' end. In contrast, it is not obvious that reversal of the stems and loops would alter the way that a pseudoknot is opened by a force that is applied blindly and presumably symmetrically to both ends of the pseudoknot by beads held by optical traps.

Similar to the observed correlation of unfolding force and frameshift efficiency observed in [17], the overall higher forces required to unfold PK 11/6 compared to PK 6/11 correlate with the difference in their frameshift efficiencies.

The differences in unfolding force distributions indicates that the reversal of loops and stems does not correspond to a simple reversal of the overall pseudoknot structure. This could be due to the stereochemistry of the bases that make up the pseudoknot, which could not be predicted in advance. It is probably important that not only has the longer loop been lengthened by one nucleotide in PK 6/11 compared to PK 11/6, both loops are also from a symmetry point of view reversed in sequence compared to the stems (i.e. the order of the nucleotides making up the stems but not the loops have been reversed in PK 6/11 compared to PK 11/6). This would clearly cause changes in the tertiary interactions of the loop with the remaining pseudoknot. The stereochemistry/tertiary interactions of the bases may thus significantly influence the mechanical properties of the pseudoknot, as was also observed by Chen et al. [21]. Indeed the structures formed by the pseudoknot nucleotide sequences may not correspond to the originally predicted structures in Figure

4.1 at all. This would mean that the differences between PK 11/6 and PK 6/11 is not only about the relative stability of their loop and stem interactions but about their overall structure.

A bimodal distribution of unfolding forces has been theoretically shown to arise in molecular transitions simply if several different conformations of the molecule are possible at a given force [74]. Different conformations could correspond to different chemical states or to different molecular orientation or degree of extension, all of which could arise randomly. Thus the bimodality of the distribution of unfolding forces for PK 6/11 indicates that several conformations of the molecule is encountered which respond differently to the force. The bimodality of the distribution of unfolding forces for the subset of data for PK 11/6 excluding one of the "super tether" molecules indicates the same for PK 11/6: the conformations that the molecule may randomly take on varies in a way that influences the unfolding force distribution.

The average magnitude of the rip forces found was as we saw about 20 pN for PK 11/6 with a high subdistribution averaging at about 28 pN (or 17, 20, and 28 pN when the rips were divided into three subdistributions) while the average rip force for PK 6/11 was only about 15 pN (13 and 16 pN when subdivided into two rip distributions). For comparison, Hansen et al. find average rip forces for their pseudoknots of 31 ± 1.9 pN and 39 ± 1.5 pN [17] at experimental buffer conditions and loading rates similar to the ones used here. They see refoldings in various patterns, for one pseudoknot often in two zip step at 19 ± 4 and 9 ± 4 pN. Green et al. [25] find an average rip force of 31.4 ± 3.8 and an average zip force of 15.1 ± 1.5 also with buffer and loading rate similar to the ones used here. Chen et al. found average pseudoknot rip forces at about 50 pN in a buffer which in contrast to the one used here contains no stabilizing Mg^{2+} ions. They use the same pulling speed as that used here (trap strength not indicated) [18]. In later experiments, also without Mg^{2+} ions, the same group found bimodal rip force distributions with one peak averaging at about 18 pN and the other varying between about 22 and 50 pN [21]. The average rip forces found by Liphardt et al. [4] found average unfolding forces for a range of RNA hairpins in the presence of Mg^{2+} ions between 8 and 22 pN at loading rates of 3 pN/s or about 1/3 of those used here.

The experiments without Mg^{2+} are difficult to compare to the results found here, but the others provide interesting contrast. It appears that only the high force long PK 11/6 rips correspond well in force to the pseudoknots investigated by Hansen et al. and Green et al. whereas the lower force PK 11/6 rips and the PK 6/11 rips correspond more in strength to the hairpins investigated by Liphardt et al.

6.2 Measured versus predicted rip and zip length

The completely unfolded single strand pseudoknot lengths of PK 11/6 and Pk 6/11 are expected to differ only by one base, which means that they should have almost the same unfolding length at the same force. If anything, PK 6/11 should unfold slightly further since it is the one that contains an extra nucleotide. However, as seen in Chapter 5.4, significantly longer rips are observed for PK 11/6 than for PK 6/11.

Part of the difference in unfolding length could derive from the difference in unfolding force, since the Extensible Worm-Like Chain model predicts that the molecule will be longer at higher forces, especially when it unfolds to a state with a longer single stranded portion. The higher average unfolding force for PK 11/6 could therefore perhaps explain its longer unfolding distance. To investigate whether this is the reason for the difference in rip lengths, the predicted change in length of the molecules at the measured average rip forces was calculated as described in Chapter 4.5.1.

The EWLC predictions of rip and zip length are in general a little shorter for PK 6/11 than for PK 11/6 (compare the Prediction columns for Stem1, Stem2 and the Full pseudoknot in Tables 5.2 and 5.3). This was expected due to the lower average rip forces observed for PK 6/11. However, what stands out is the discrepancy between the predicted rip lengths and the measured rip lengths. Two possible reasons are explored here. The first is that the EWLC parameters do not describe the molecular construct well, the second that the structure opening during the rip is different from what was originally expected. First we look at altering the parameters for the EWLC prediction.

6.2.1 EWLC parameter sensitivity

We start off with a quick back-of-the-envelope check of the expected rip length for the pseudoknots. This can be made by examining the difference between the expected contour lengths of the folded and the open pseudoknot. Calculation of the contour length is less prone to parameter uncertainty than the rest of the EWLC prediction, as the contour length of single and double stranded RNA is assumed to be fairly well established. The contour length of the single strand that is expected to open is about 40 nm ($40 \text{ nm} = (12 \text{ (stem1)} + 6 \text{ (loop1)} + 32 \text{ (loop2)} + 22 \text{ (stem2)}) \text{ nucleotides} * 0.56 \text{ nm/nucleotide}$), while the folded pseudoknot is expected to have a contour length of about 5 nm ($5 \text{ nm} = (6 \text{ (stem1)} + 11 \text{ (stem2)}) \text{ basepairs} * 0.28 \text{ nm/basepair}$). Thus the change in contour length is about 35 nm. If this were to correspond to an actual change in extension of only 15 nm, the single stranded portion of the molecular construct would have to be very flexible - much more flexible than we assume in the EWLC prediction shown in Chapter 5.4.2. This means that the persistence length, L_p , must be lowered, corresponding to more coiling of the molecule.

It turns out that the EWLC prediction can be made to correspond approximately to the measured value if the persistence length, L_p , of the single strand region is reduced from 1 nm to about 0.25 nm. Keeping constant values for L_p for the double stranded portion and for the elastic modulus K_0 for both single and double stranded regions, this gives a predicted unfolding length for PK 11/6 at 19.8 pN of about 16 nm (the mean rip force for all first PK 11/6 rips is 19.8 pN, while the mean rip length for all first PK 11/6 rips is 13.5 nm). In contrast, reducing L_p for the double stranded portion from 50 to 10 nm, as is assumed in [46], or increasing K_0 for the single stranded portion significantly (from 1 nN to 15 nN) does not change the predicted unfolding length much.

However, although it is possible that L_p is lower for short pieces of RNA than initially expected and that this could account for the short extensions measured, this does not seem like a plausible explanation. The value needed for L_p is quite a low one; in 2001 Liphardt et al. [4] cited the range 0.7-5 nm found in literature for L_p for single strand RNA and themselves found the value 1 nm; a more recent value of $L_p = 1.5 \pm 0.1 \text{ nm}$ is reported by Mangeol et al. in 2011 [75], which is an improvement upon a 2006 value of $1.1 \pm 0.6 \text{ nm}$ [76] by the same group. Also, apart from inconsistency with the literature the problem with simply explaining the discrepancy between measured and predicted extension with the need for a lower value of L_p in the EWLC is that this explanation is incompatible with the longest rip lengths measured for PK 11/6. Some of these rip lengths exceed 30 nm, and with a lower value of L_p , these rips would according to the EWLC model correspond to a change in contour length that is far longer than the actual total length of single stranded RNA expected in the construct. Even with the knowledge that noise in the position of the Nanotracker traps is significant, meaning that the measured change in extension might not correspond to the actual change in extension of the molecule, measuring rips of 30^+ nm for the pseudoknot opening seems impossible assuming single stranded L_p of only 0.25 nm. The Nanotracker noise occurs at about 1000 Hz, so with sampling at 5000 Hz and smoothing by a 200 data point moving average down to about 25 Hz, the noise really should not noticeably affect the change in extension measured. Additionally, altering the

parameters of the EWLC does not explain the observed lengths of two-step transition steps (see e.g. Figure 5.8 (a)).

We will therefore return to the alternative explanation for the discrepancy between the observed and predicted rip lengths: the majority of the transitions observed (even the one-step transitions) may not correspond to the full opening of the pseudoknot.

6.3 Subdistributions and intermediate states

As noted in Chapter 4.1, the single stranded RNA sequences of PK 11/6 and PK 6/11 may fold in numerous conformations other than the pseudoknot structures originally predicted by the program `pknobsRG`. Using other programs available online (`DotKnot` and `HotKnots` [63, 64]) as well as `pknobsRG` [61] with modified input, i.e., leaving out part of the nucleotide sequence, several other possible foldings of the single strand sequences have been identified. A variety of these structures are shown in Figures 6.1 and 6.2. They differ significantly in the rip lengths that they would yield upon unfolding, but their predicted energies of formation (ΔG) shown in the figure captions are surprisingly similar. The predictions of ΔG of formation made by these programmes is based only on the base pairings and leaves out tertiary interactions.

Probably the most stable alternative structure for PK 11/6 is the alternative pseudoknot, PK 11/6U, seen in Figure 6.1 (b). This structure was predicted from a truncated version of the nucleotide sequence for PK 11/6 and was suggested by Jesper Tholstrup. It is very similar to the PK 10/6 U structure predicted from a truncated version of the nucleotide sequence for PK 10/6, a pseudoknot investigated by Jesper [67]. PK 10/6 U was often observed in experiments that parallel the ones performed for PK 11/6 and PK 6/11 in this study.

For PK 6/11 no particularly stable alternative prediction stands out, except that in several of the alternative structural predictions a hairpin made up of stem2 and loop2 apparently remains stable (Figure 6.2 (b), (c), (e) and (g)).

If we hypothesize that the unexpectedly short rip lengths arise because the structure that opens does not correspond to the full pseudoknot, two possible explanations now present themselves:

1. The pseudoknot opens one stem at a time, and even when only one transition length is observed during a given pull, this corresponds only to one stem of the pseudoknot opening. This might mean that the other stem does not open at all during the pull, is already open at the start of the pull (the pseudoknot acts like a hairpin) or the opening of the other stem is lost in noise.
2. The folded structure that opens during a rip could itself be shorter than expected, as is observed for many of the possible alternative structures in Figures 6.1 and 6.2. This would give rise to shorter observed rip lengths.

To investigate these possibilities, predictions were made not only of the expected rip length for the full pseudoknot structure unfolding but also for some of the other possible structures that might unfold. These predictions were shown in Figures 5.8 and 5.8 and in Tables 5.2 and 5.3. Two of the possibilities investigated were that stem1 and loop1 might unfold, leaving behind (at least for a time) a hairpin consisting of stem2 and loop2 and vice versa. This idea, namely that the pseudoknot might open one stem at a time, is suggested by Hansen et al. [17] because of their observation of two-step zip transitions. While it is considered unlikely that a hairpin consisting of the shorter stem and longer loop (stem2 and loop2 for P11/6, stem1 and loop1 for PK 6/11) would be stable for long, the prediction

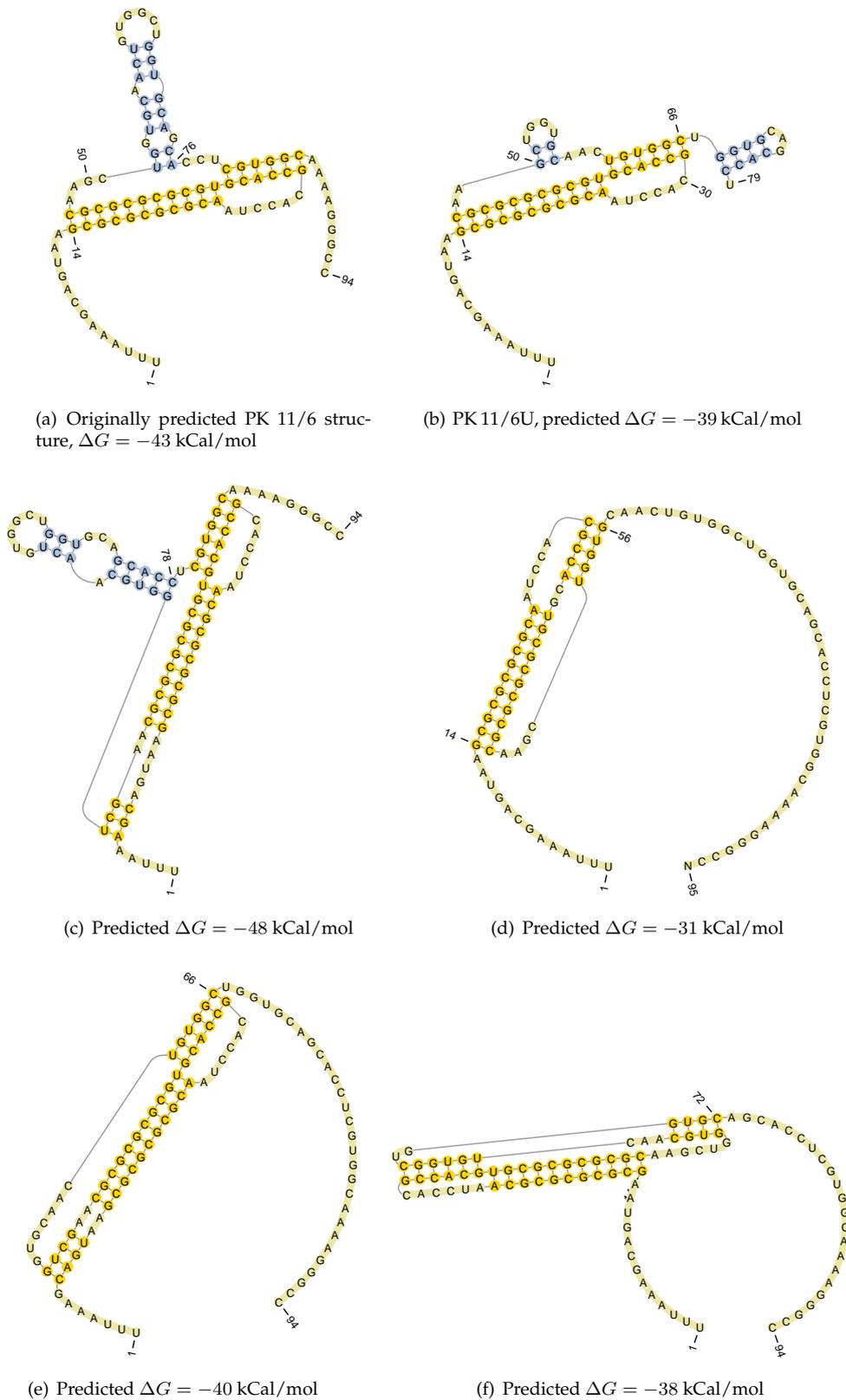


FIGURE 6.1: Possible structural conformations of PK 11/6: (a) Originally expected structure (predicted by pknotsRG, also shown in Figure 4.1). (b) PK 11/6 U structure predicted by pknotsRG [61] from PK 11/6 sequence with final part deleted. (c), (d), (e), (f): Alternative structures predicted from PK 11/6 sequence by the programme DotKnot [63]. Images rendered by PseudoViewer 3.0 [62]. Pseudoknot stems yellow, hairpins blue.

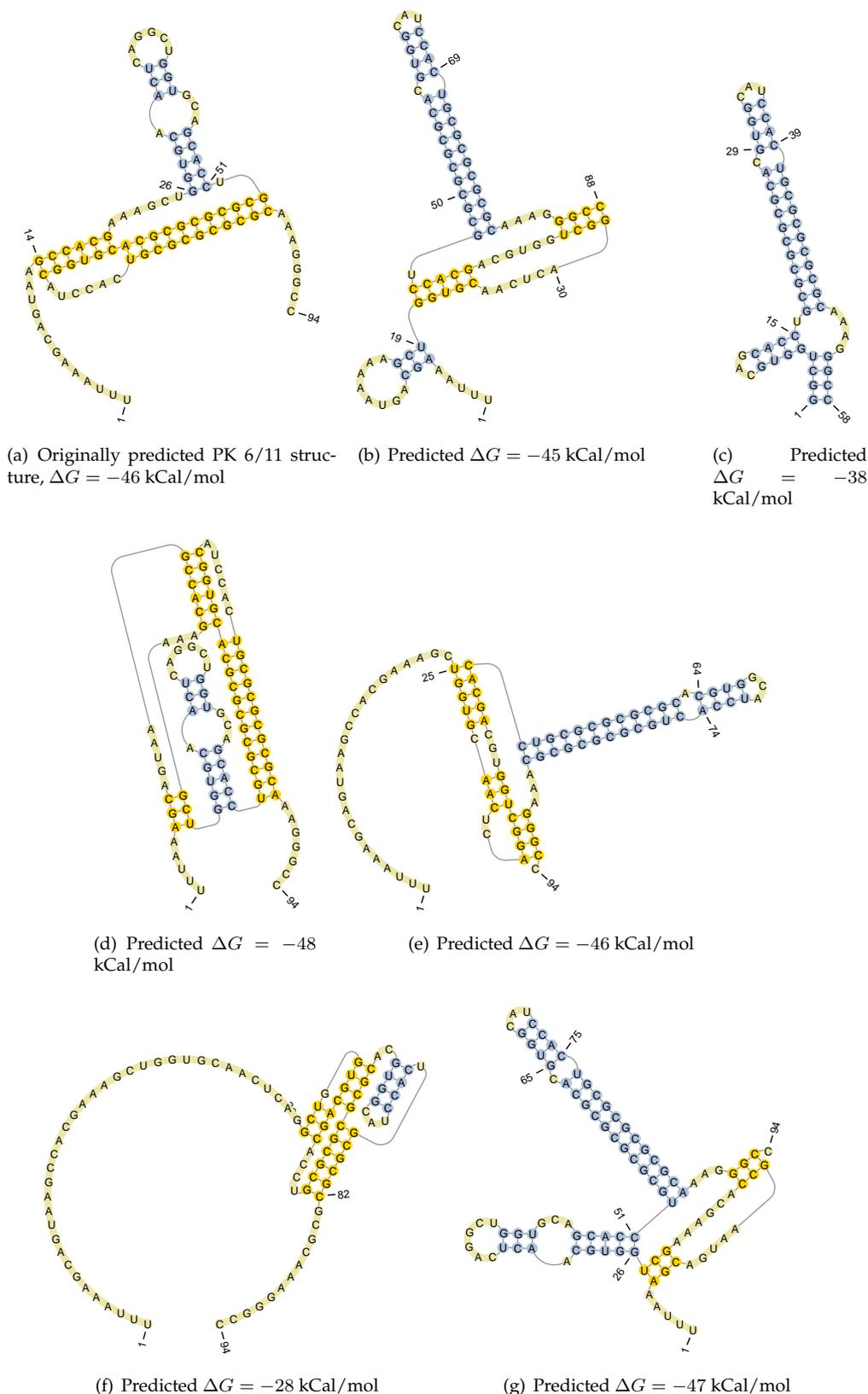


FIGURE 6.2: Possible structural conformations of PK 6/11: (a) Originally expected structure (also shown in Figure 4.1, predicted by pknotsRG). (b) Predicted folding when central part of structure is deleted (suggested deletion by Jesper Tholstrup, prediction made by pknotsRG). (c) Hairpin portion of (b) and a little surrounding sequence only (structure predicted with pknotsRG [61]). (d), (e), (f), (g): Alternative structures predicted from PK 6/11 sequence by the programme DotKnot [63]. Images rendered by PseudoViewer 3.0 [62]. Pseudoknot stems yellow, hairpins blue.

shows that if the structure unfolds one hairpin at a time with the unfolding of the longer stem and shorter loop first, this transition would be so short (about 4 nm) that it would most often be indistinguishable from noise in the force-extension curves. It thus appears possible either that only the short stem and long loop unfolded during pulling, leaving a relatively stable hairpin structure behind, or that the full pseudoknot unfolded one hairpin at a time with first the long stem and short loop, then the short loop and long stem opening in such a way that the opening of the first hairpin was obscured by noise (See Figures 6.1 (a) and 6.2 (a) for reference).

Another possibility considered was that only one hairpin made up of the longer stem and the shorter loop had formed at all and was opened during pulling. This is in the scatter plots called "stem2 opens, stem1 open from start" or "stem1 opens, stem2 open from start". Again it appears to be a possible - and plausible - explanation for the observed short rip lengths.

Finally, for the PK 11/6 data the predicted rip length of the shortened pseudoknot structure PK 11/6 U was investigated. This too appears to correspond to some of the observed rip and zip lengths for this construct. Note that if the PK 11/6 U structure formed or opened hairpin by hairpin, the individual rip lengths of these transitions would be so short that they would easily be obscured by noise.

To conclude, when the data for rips and zips shown e.g. in Figures 5.8 and 5.8 are shown together with predictions of rip/zip length based on the expected pseudoknot structure and stretch, it appears that only the very longest rips and zips for PK 11/6 could correspond to the full originally expected pseudoknot opening. What we see in the scatter plots for PK 11/6 is that the remaining rips and zips correspond much more closely to a rip length that would correspond to either:

- opening only stem2 and loop2 of the pseudoknot (assuming the opening of stem1 and loop1 are obscured by noise)
- opening only a hairpin consisting of stem1 and loop1 and no other structure
- opening the alternative pseudoknot structure of PK 11/6 U.

The observation in Figure 5.8 a that all one-step hop rips and zips during elongation of the molecule occur at fairly low forces, corresponding best to the various hairpins and not so well to the PK 11/6 U unfolding, lends credibility to the idea that the large cluster of data between 5 and 25 nm and 5 and 30 pN for the PK 11/6 construct is actually a mixture of points deriving from hairpin unfoldings and PK 11/6 unfolding. This is because it seems logical that hopping would more easily occur between the hairpin and single strand than between a fully formed pseudoknot and a single strand. However, it may also be that hopping simply generally is more likely to occur at lower forces where the molecular construct is under less tension. As PK 11/6 also at times opens in two step transitions that sometimes occur at extremely different forces (see Figure 5.1 and in which the individual steps of the two-step transitions are not necessarily shorter than the one-step rips, it appears that much more is going on than just opening of the pseudoknot one hairpin at a time. Clearly sometimes one intermediate structure changes midpull to a structure similar to the originally observed pseudoknot.

If it is true that PK 11/6 displays a mixture of hairpin, small pseudoknot and full pseudoknot conformations, this supports the observation that the rips are distributed in three different clusters, but the delineation between the two lower force, shorter rip distributions remains difficult to discern.

In contrast to PK 11/6, PK 6/11 is never observed to unfold in steps long enough to match the predicted rip length for the full pseudoknot. It appears that this pseudoknot instead mostly is observed either to open one hairpin at a time with the opening of the longer stem and shorter loop obscured by noise or to be mostly in the conformation of a single hairpin

that opens during the pull. It is also possible that the pseudoknot is switching between some of the states seen in Figures 6.2. When PK 6/11 opens in two step transitions, these are far shorter than the one-step transitions and on average appear add up to a full one-step transition. It is not obvious how these smaller partial transitions could add up to correspond to a longer transition that more or less corresponds to only stem2 and loop2 of the full pseudoknot opening.

All in all the results in Chapter 5 paint a convincing picture that the rips and zips observed for the two pseudoknot containing nucleotide sequences represent a mixture of transitions between a variety of structural conformations.

Figures 5.12 and 5.13 provide an interesting addition to the picture of how the single stranded RNA sequence passes from one structural conformation to another. They show subdistributions of one-step zips belonging to the same pulls as the one-step rips that were divided into subdistributions according to their different lengths. Clearly rips and zips occurring in the same pulls are not necessarily correlated in length. A long rip representing the unfolding of the full pseudoknot to the single strand may be followed by a short zip representing the folding of the single strand into a hairpin. If the structure reverts to a full pseudoknot, this must first occur when the molecule is at low force.

The next chapter will examine the energies and kinetic parameters associated with the observed structural transitions.

7 Results Part II: Gibbs free energy and kinetic parameters

This chapter will describe how the opening and closing transitions described in the previous chapters are used to estimate the Gibbs free energy change and kinetic parameters associated with each pseudoknot. Work calculations are based on both force and extension data for the transition while the calculations of the kinetic parameters, namely the distance to the transition state and the standard rate of reaction, are based only on the force measurements.

7.1 Work distributions and ΔG

The Gibbs free energy, $\Delta G_{transition}$, of the transition between two states of a molecule may be calculated from the distributions of work required for the transition. As described in Chapter 3.4, for irreversible transitions this calculation may be done either using the Crooks Fluctuation Theorem (CFT) or the Jarzynski Equality (JE). Both methods will be used below. Before we start out, however, we must note that this calculation requires that transitions used to estimate the work distributions all start and end in the same state. For this reason it is important to separate distinct molecular transitions in the analysis. This is not trivial but will be done as far as possible.

The work required for each transition was estimated in using two different approaches as described in Chapter 3.4.4, namely using the “simple” approach and the “Collin” approach.

In the “simple” approach, the areas directly beneath the rips and zips on the force-extension curves were used to estimate the work required for the transitions (see Figure 3.7). In cases where the pseudoknot opened and/or closed repeatedly only the first opening on the extension trace and the first closing on the relaxation trace were used. *The rips and zips used to find ΔG_{PK} using the simple approach were exactly the ones shown in Figures 5.12 and 5.13 in Chapter 5.4.* Two-step transitions were thus left out of the simple approach analysis.

For the “Collin” approach, the transitions included were slightly different. In this case the work for each transition was estimated as the area beneath the force-extension curve from the beginning to the end of the opening or closing transition. This meant that gradual and two-step transitions were also included as long as their beginning and end points were clearly defined at points where the elongation and relaxation curves coincided (see Chapter 3.4.4). Pulls were excluded from analysis if the rip and zip were significantly offset from each other due to e.g. noise or drift. On the other hand, even if the rip or zip was too gradual to be included as a clear rip or zip in the simple area analysis, the area under the full closing or opening could often be included in the Collin type analysis (for example, the openings and closings seen in Figure 5.1(f) and 5.1(c) were both included in this analysis but not in the simple area analysis). This explains the different numbers of pulls included in each analysis below.

As described in Chapter 3.4.3, from the work distributions for each pseudoknot, ΔG for the full transition was found using the CFT or the JE. $\Delta G_{transition}$ specifically for the opening/closing transition investigated was then found by subtracting the work, $W_{stretch}$, required to stretch/relax the handles and the pseudoknot. The latter was found theoretically as the integral under the Extensible Worm Like Chain model for how the construct stretched (described in Chapter 4.5.1). As in Chapter 3.4, the work for the full opening transition including handle and pseudoknot stretching is called W_f ("f" for forward transition) while the work for the full closing transition is called W_r ("r" for reverse transition).

7.1.1 Finding ΔG_{PK} using the simple area approach

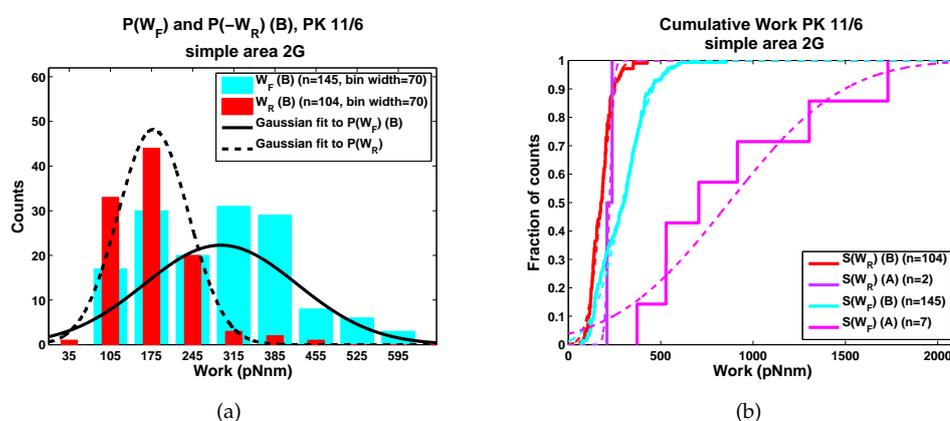


FIGURE 7.1: PK 11/6: Simple approach work distributions using double Gaussian (2G) rip subdivisions. Letter names of subdivisions correspond to the ones used in Figure 5.12 (a) and (b). (a): Histograms of $P(W_F)$ (B) and $P(W_R)$ (B) with single Gaussian fits to the work distributions, showing their point of intersection (although the number of data points in each distribution is different, the fitted probability distributions have been multiplied by the same constant in order to be comparable and intersect at the correct point). (b) Cumulative distributions of the work represented by the rips and zips in subdivisions A and B. Histograms for distribution A not shown due to low number of data points.

As described above, the subdivisions of rips and zips found in Chapter 5.4.2 were used to theoretically divide the rips and zips representing different types of molecular transitions before calculation of work energy for each transition. This meant that for PK 11/6 subdivision of the rip and zip data into two or three different types of transitions was explored while for PK 6/11 one single transition distribution or subdivision into two types of transitions were investigated. Figures 7.1 – 7.4 show all of the resulting work distributions as cumulative distributions and in the cases where there were enough data to justify it also histograms of the work distributions. It may be questioned whether the zip subdivisions identified in Figures 5.12 and 5.13 really matched the rip subdivisions well enough to merit this approach, but no other approach seemed immediately obvious.

Interestingly, although initial work investigations using the full distribution of rips and zips for each pseudoknot were all well fitted by Gaussian distributions (see Figure F.1 in Appendix F and Figure 7.3), the subdivisions of work energies also were fairly well fitted by Gaussians, at least when they contained enough data points for fitting. Especially

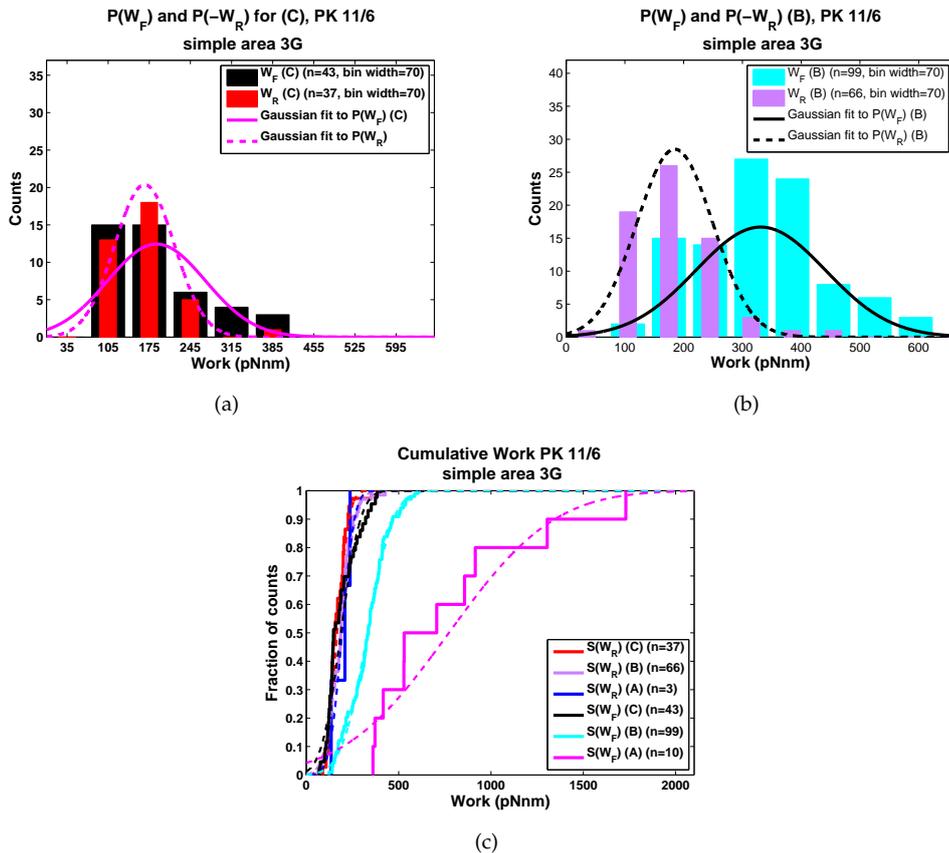


FIGURE 7.2: PK 11/6: Simple approach work distributions using tripple Gaussian (3G) rip subdistributions. Letter names of subdistributions correspond to the ones used in Figure 5.12 (c) and (d). (a): Histograms of $P(W_F)$ (C) and $P(W_R)$ (C) with single Gaussian fits, showing overlapping distributions. (b) $P(W_F)$ (B) and $P(W_R)$ (B) with single Gaussian fits showing distribution intersection. (c): Cumulative distributions of the work represented by the rips and zips in subdistributions A, B, and C. Histograms for subdistribution A not shown due to low number of data points. Note that although the number of data points in each distribution is different, the fitted probability distributions shown with the histograms have been multiplied by the same constant in order to be comparable and intersect at the correct points.

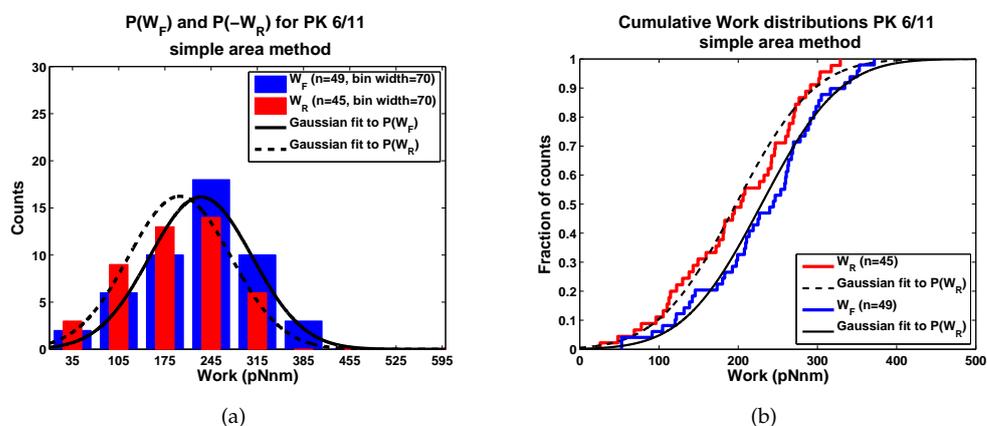


FIGURE 7.3: PK 6/11: Simple approach work distributions using single (i.e., full) Gaussian rip distribution. (a) Histograms and (b) cumulative distributions of work values for the full distribution of opening and closing transitions. Although the number of data points in each distribution is different, the fitted probability distributions shown with the histograms have been multiplied by the same constant in order to be comparable and intersect at the correct point.

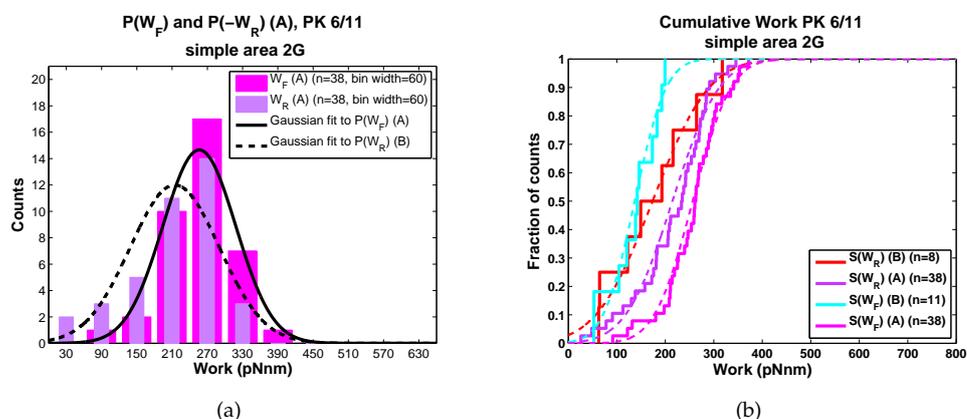


FIGURE 7.4: PK 6/11: Simple approach work distributions using rip tripple Gaussian (3G) subdistributions. Letter names of subdistributions correspond to the ones used in Figure 5.13. (a): Histograms of $P(W_F)$ (A) and $P(W_R)$ (A) with single Gaussian fits (although the number of data points in each distribution is different, the fitted probability distributions have been multiplied by the same constant in order to be comparable and intersect at the correct points). (b): Cumulative distributions of the work represented by the rips and zips in subdistributions A and B. Histograms for subdistribution A not shown due to low number of data points.

for PK 11/6, it looks as though the subdivision of the data according to rip length also helps separate out data points that belong to different work energy distributions.

A sample table with calculated values of work and ΔG for the double Gaussian subdistribution of transitions for PK 11/6 are summarized in Table 7.1. Similar tables for the tripple Gaussian subdistributions for PK 11/6 and the double Gaussian subdistributions for PK 6/11 as well as for the full distributions of work values for both pseudoknots may be found in Appendix F. A summary with the most important values from these tables and from the work calculations using the Collins approach is shown in Table 7.3

The variables included in the tables are defined as follows:

- n_f and n_r are the numbers of rips and zips included in the analysis.
- $\langle W_f \rangle$ and $\langle W_r \rangle$ are the averages work values for the forward and reverse work distributions.
- ΔG_{Crooks} is the work value at the intersection of Gaussian fits to the forward and reverse work distributions corresponding to ΔG for the full reversible transition (including stretching contributions) according to the Crooks Fluctuation Theorem.
- $\Delta G_{\text{JE}(f)}$ is the value of ΔG for the full transition found using the Jarzynski equality for the forward transitions only, $\Delta G_{\text{JE}(r)}$ is the same for the reverse transitions only and $\langle \Delta G_{\text{JE}} \rangle$ is their average.
- $\langle W_{\text{stretch}}(f) \rangle$ and $\langle W_{\text{stretch}}(r) \rangle$ are the average calculated contributions to ΔG from stretching the handles ($\langle W_{\text{stretchhandle}}(f) \rangle$) and pseudoknot ($\langle W_{\text{stretchPK}}(f) \rangle$) during the forward and the reverse transitions. $\langle W_{\text{stretch}} \rangle$ is an average of the stretching contributions weighted by the number of forward versus reverse transitions included in the analysis.
- $\Delta G_{\text{transition using CFT}} = \Delta G_{\text{Crooks}} - \langle W_{\text{stretch}} \rangle$.
- $\Delta G_{\text{transition using JE}} = \langle \Delta G_{\text{JE}} \rangle - \langle W_{\text{stretch}} \rangle$.
- $\Delta G_{\text{transition using } \langle W \rangle} = (\langle W_f \rangle + \langle W_r \rangle) / 2 - \langle W_{\text{stretch}} \rangle$.

The last item, “ $\Delta G_{\text{transition using } \langle W \rangle}$ ”, reflects that in addition to using the CFT and the JE, a very crude estimate $\Delta G_{\text{transition}}$ was calculated from $\langle W \rangle$, the average of the work energies for the forward and the reverse transitions. This was done to provide a sense of how the non-equilibrium conditions of the transitions affect the calculation of the work energy and also to enable estimation of $\Delta G_{\text{transition}}$ in cases where the data were too sparse for CFT or JE analysis and in cases where the forward and reverse transition work distributions overlap to such a degree that the CFT cannot be used as the distributions do not intersect between their maxima.

The errors shown in the table are the standard errors on the means except in the case of $\Delta G_{\text{transition using the CFT}}$ which was found using a bootstrapping script developed by Jesper Tholstrup. The error on $\Delta G_{\text{transition using the Jarzynski equality}}$ was not found.

An important point to notice in Table 7.1 is that the energy that goes into stretching the handles is generally negligible compared to the energy that goes into stretching (or relaxing) the single stranded portion of the structure that unfolds (or refolds). In this context it is necessary to emphasize that $\langle W_{\text{stretch}} \rangle$ was summed from the average work required to stretch the handles ($\langle W_{\text{stretchhandle}} \rangle$) and the average work required to stretch the full pseudoknot ($\langle W_{\text{stretchPK}} \rangle$) during the forward and the reverse transitions. Thus in the calculation of the stretch energy, it no allowance is made for the hypothesized shorter structures unfolding during shorter transitions. This means that the work that goes into stretching the single strand during each transition is overestimated and better estimates of $\Delta G_{\text{transition}}$ would be higher.

Another thing to note in the detailed work data table is that the estimate made of $\langle \Delta G_{\text{JE}} \rangle$ by averaging $\Delta G_{\text{JE}(f)}$ and $\Delta G_{\text{JE}(r)}$ is surprisingly close to the value found using the CFT

TABLE 7.1: PK 11/6, 2G subdistributions: Calculated ΔG from simple area approach. Calculations are based on the distributions shown in Figure 7.1. Variable names explained in text. No uncertainty is given for ΔG_{JE} since its calculation requires bootstrapping or a similar statistical technique (see [46]). No ΔG_{Crooks} or ΔG_{JE} was found for subdistribution A since the data for this subdistribution was so sparse.

	PK 11/6 2G	
	Subdistr A \pm SEM	Subdistr B \pm SEM
n_f	7	145
n_r	2	104
$\langle W_f \rangle$ (kCal/mol)	125.1 ± 28.8	42.1 ± 1.6
$\langle W_r \rangle$ (kCal/mol)	32.1 ± 2.0	25.6 ± 0.9
$(\langle W_f \rangle + \langle W_r \rangle)/2$ (kCal/mol)	79 ± 14	14.4 ± 0.9
ΔG_{Crooks} (kCal/mol)	NA (too sparse)	36.7 ± 0.9
$\Delta G_{JE(f)}$ (kCal/mol)	NA (too sparse)	12.9
$\Delta G_{JE(r)}$ (kCal/mol)	NA (too sparse)	58.8
$\langle \Delta G_{JE} \rangle$ (kCal/mol)	NA (too sparse)	35.8
$\langle W_{stretchhandle}(f) \rangle$ (kCal/mol)	-3.9 ± 0.5	-1.64 ± 0.04
$\langle W_{stretchPK}(f) \rangle$ (kCal/mol)	27.9 ± 3.8	21.0 ± 0.5
$\langle W_{stretch}(f) \rangle$ (kCal/mol)	24.0 ± 3.8	19.3 ± 0.5
$\langle -W_{stretchhandle}(r) \rangle$ (kCal/mol)	-1.70 ± 0.5	-1.29 ± 0.04
$\langle -W_{stretchPK}(r) \rangle$ (kCal/mol)	16.7 ± 3.1	15.7 ± 0.3
$\langle -W_{stretch}(r) \rangle$ (kCal/mol)	15.0 ± 3.1	14.4 ± 0.3
$\langle W_{stretch} \rangle$ (kCal/mol)	22 ± 3.6	17.3 ± 0.9
$\Delta G_{transition}$ using $\langle W \rangle$ (kCal/mol)	57 ± 14	16.6 ± 1.4
$\Delta G_{transition}$ using CFT (kCal/mol)	NA (too sparse)	19.4 ± 1.3
$\Delta G_{transition}$ using JE (kCal/mol)	NA (too sparse)	18.5

considering how far the estimates of ΔG_{JE} for either the forward or the reverse transition alone is. This pattern was generally observed in the work calculations where the number of work data points used in the Jarzynsky Equality to estimate ΔG was more than about 10.

Nonetheless, of the of the three estimates of $\Delta G_{transition}$ displayed, the one that is most reliable is probably the one found using the Crooks Fluctuation Theorem, since the data is too scarce for a reliable estimate using the Jarzynski Equality. Finding $\Delta G_{transition}$ using $\langle W \rangle$ is really only a crude approximation to the Crooks Fluctuation Theorem method which assumes that the distributions $P(W_F)$ and $P(W_R)$ intersect exactly halfway between their maxima. Note, however, that in the cases where the CFT cannot be used because the distributions overlap, the error will be small because the reaction is almost reversible anyway. This is of course not true in the cases where the CFT cannot be used because the number of data points in the work distributions is too sparse. The estimate of the energy of transition for subdistribution A for PK 11/6 found using this method is shown in the table but must be considered extremely uncertain, not only because of its large SEM but also because the few zips that were included in the analysis might not derive from the correct type of transition at all.

In the summary table, Table 7.3, the value for $\Delta G_{transition}$ considered most reliable for each

TABLE 7.2: Calculated ΔG (kCal/mol). Collin approach. Variables explained in text in Section 7.1.1. OBS Uncertainties in DeltaG CFT!

	PK 11/6 \pm SEM		PK 6/11 \pm SEM
	Bimodal		Unimodal
	Lower distr	Upper distr	
n	87	30	50
$\langle W_f \rangle$ (kCal/mol)	38.2 ± 1.7	146 ± 12	39.2 ± 1.9
$\langle W_r \rangle$ (kCal/mol)	31.2 ± 1.3	116 ± 8	37.7 ± 2.2
$(\langle W_f \rangle + \langle W_r \rangle)/2$ (kCal/mol)	34.7 ± 1.1	131 ± 7	38.5 ± 1.4
ΔG_{Crooks} (kCal/mol)	39.6	156	NA
$\Delta G_{\text{JE}(f)}$ (kCal/mol)	12.8	81.2	11.3
$\Delta G_{\text{JE}(r)}$ (kCal/mol)	54.8	262	96.0
$\langle \Delta G_{\text{JE}} \rangle$ (kCal/mol)	33.8	171	53.7
$\langle W_{\text{stretchhandle}} \rangle$ (kCal/mol)	3.1 ± 0.4	14.7 ± 1.9	0.3 ± 0.6
$\langle W_{\text{stretchPK}} \rangle$ (kCal/mol)	18.1 ± 0.5	26.4 ± 1.1	17.2 ± 0.5
$\langle W_{\text{stretch}} \rangle$ (kCal/mol)	21.2 ± 0.3	41.1 ± 1.1	17.5 ± 0.4
$\Delta G_{\text{transition}}$ using $\langle W \rangle$ (kCal/mol)	13.5 ± 1.1	90 ± 7	20.9 ± 1.5
$\Delta G_{\text{transition}}$ using CFT (kCal/mol)	$12.1 \pm$	$115 \pm$	NA
$\Delta G_{\text{transition}}$ using JE (kCal/mol)	12.6	130	36.1

transition is displayed.

7.1.2 Finding ΔG_{PK} using the Collin approach

The distributions of work energy found by using the Collin approach are seen Figure 7.5. With this approach (finding the entire area between zip and rip under the force extension curve) the distributions of work energies found are very broad. For the full data set for PK 11/6 the distributions were especially broad and appeared better approximated by the sum of two Gaussian distributions, one sharp with a mean at low work values and one very broad with a much higher mean, than by a single Gaussian (see Figures 7.5 (a) and (b)).

For $P(W_F)$ for PK 11/6 the bimodality probably derives from the high force long rips observed in Figure 5.8 (a), which are included with the Collin approach in contrast to the simple area approach, since two-step transitions were included by the Collin approach as long as they had clearly defined start and end points (points corresponding to D and B in Figure 3.7 in Chapter 3.4.4). The lower of the two work subdistributions thus may correspond to the main cluster in the scatter plot of relatively short low force rips while the upper of the two work subdistributions may correspond to the more diffuse cloud of higher force rips.

For $P(W_R)$ for PK 11/6 the bimodality is not readily predicted from the scatter plot in Figure 5.8 (b)), but is probably due to the gradual refoldings from the high force openings that are included in the Collin analysis but not in the scatter plot. This may mean that this approach captures a portion of the data - and a type of transition - that is not well captured by the simple approach. However, as noted in Chapter 3.4.4 these very large work value estimations are more vulnerable to errors in estimating the work that goes into stretching the handles and the single stranded pseudoknot than are the work values estimated by the simple area approach.

In order to extract free energy information from this bimodal data set, it was divided into two subdistributions as follows: The data for $P(W_F)$ was fitted by a double Gaussian using

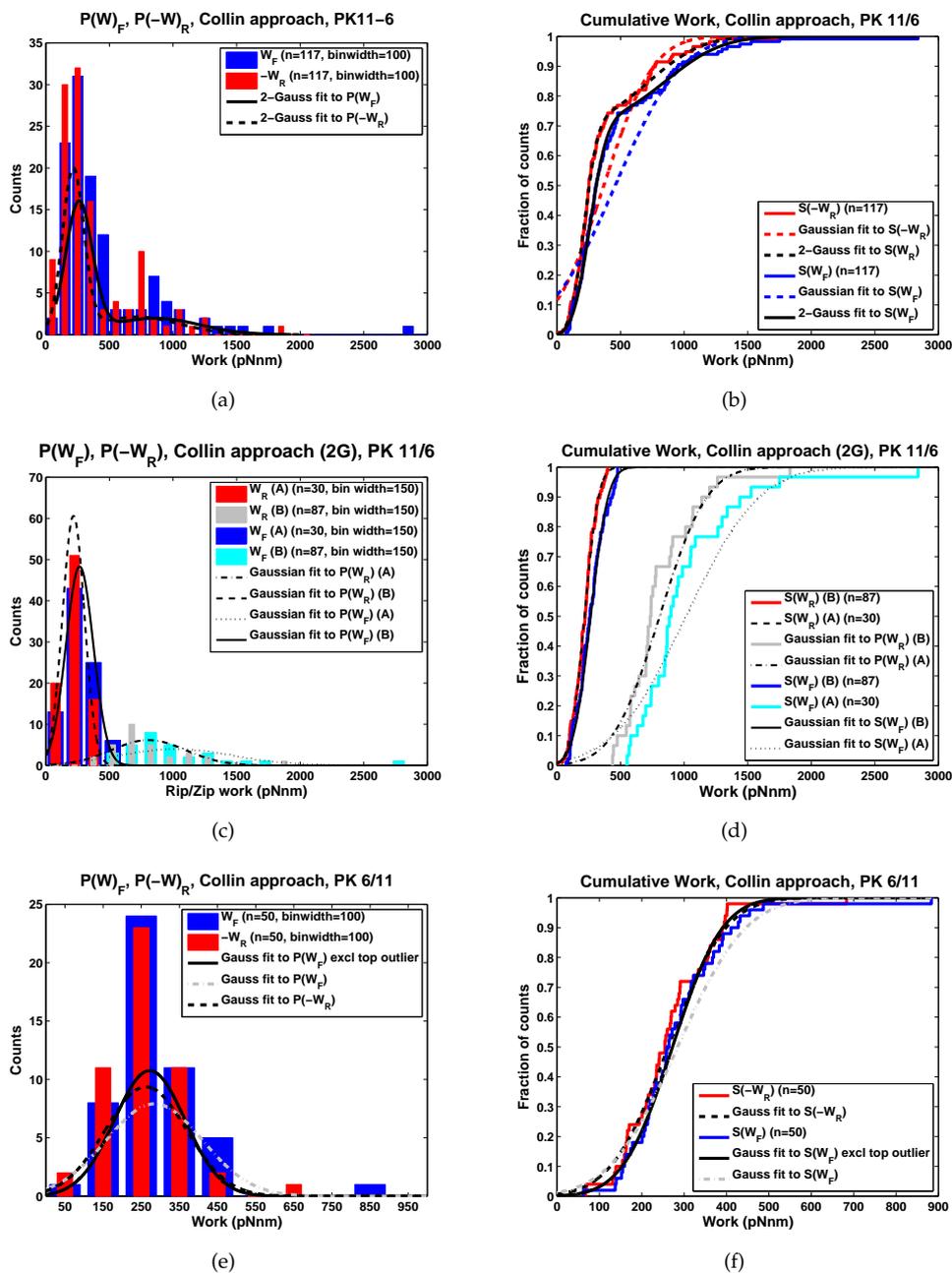


FIGURE 7.5: Collin approach to finding $P(W_F)$ and $P(-W_R)$ using the area beneath the force extension curve for the full opening and closing transitions. (a) PK 11/6: Histograms with Gaussian fits of $P(W_F)$ and $P(-W_R)$ as well as a double Gaussian fit to $P(W_F)$. Note that Gaussian fits are multiplied by a constant to be visible on the scale of the histograms. (b) Cumulative distributions and cumulative fits for PK 11/6. (c) PK 11/6: Histograms with Gaussian fits of $P(W_F)$ and $P(-W_R)$. The fit to $P(W_F)$ excludes the top outlier (fit to full dataset not shown). (d) Cumulative distributions and fits for PK 6/11. (e) PK 6/11: Histograms with Gaussian fits of $P(W_F)$ and $P(-W_R)$. (f) Cumulative distributions and fits for PK 6/11.

a Maximum Likelihood Estimation routine just as the rip length and rip force data was in Chapter 5.4.1. The intersection of the two Gaussians was found and used to distinguish the two distributions so that all opening transitions with higher energy were classified as belonging to one distribution and those with lower opening transition energy to the other. Since with this approach each opening transition was accompanied by a corresponding closing transition (see Chapter 3.4.4 and in particular Figure 3.7), distributions of closing transition work energies could automatically be determined. As may be seen in Figures 7.5 (c) and (d), these indirectly determined distributions of $P(W_R)$ neatly correspond to higher and lower work energy distributions exactly as the two distributions for $P(W_F)$. Therefore the Crooks Fluctuation Theorem and Jarzynski Equality are applied to each of these two sets of work distributions to find two different Gibbs free energies of transition that are assumed to correspond to two different molecular transitions.

The results of these calculations are shown in Table 7.2 together with the results for PK 6/11. The variable names in the table are the same as in Table F.1 (see explanation in text of Section 7.1.1) except that using the Collin approach the number of opening transitions, n , was the same as the number of closing transitions, so there was no need to distinguish n_f and n_r . Also, using the Collin approach $\langle W_{\text{stretch}}$ is identical for the opening and closing transitions, so there is no need to distinguish (or find the average of) $\langle W_{\text{stretch}}(f)$ and $\langle W_{\text{stretch}}(r)$.

For PK 6/11, $P(W_F)$ and $P(W_R)$ overlap completely when the Collin approach is used (see Figures 7.5 (e) and (f)). Therefore the CFT cannot be applied and ΔG for the full transition is calculated simply as the average of $\langle W_F \rangle$ and $\langle W_R \rangle$. Due to the overlap, the error introduced by this estimate should be very small. Since the work distributions appear very well approximated by a single Gaussian for both the forward and the reverse transitions, no attempt is made to divide the work data into subdistributions for PK 6/11 using the Collin approach.

Notice in Figures 7.5 (c) and (d) that $P(W_F)$ (B) and $P(W_R)$ (B) do not appear completely well fit by single Gaussians. Their shapes hint that the full Collin approach work distribution for PK 11/6 could be better fit by a tripple Gaussian, matching the distributions found in Chapter 5.4.1, with only very few data points in the highest energy work distribution. This idea is speculative and is not explored by an actual fit. However, the bad single Gaussian fit to these subdistributions means that the estimate of ΔG_{Crooks} made by finding the intersection of the Gaussian fits is probably not very accurate. By eye one would estimate from Figure 7.5 (c) that the two distributions intersect in between their maxima at about 700 pNnm or 100 kCal/mol rather than 155 kCal/mol. This would give an estimate of $\Delta G_{\text{transition}}$ after subtraction of $\langle W_{\text{stretch}}$ of about 60 kCal/mol. While this number still very different from the estimated transition energy for the lower distribution, it is not as strikingly high.

In the next section, the estimates of $\Delta G_{\text{transition}}$ using the Collin approach will be compared to the estimates using the simple area approach and to the values predicted by programmes that predict RNA folding energies from the basepairs expected in the structure.

Table 7.3 summarizes the estimated values of $\Delta G_{\text{transition}}$ using the different methods/approaches above. It also shows the theoretical predictions of $\Delta G_{\text{transition}}$ made by the programme pknotsRG (43 kCal/mol for PK 11/6 and 49 kCal/mol for PK 6/11, lower values for shorter the suggested structure of PK 11/6 U and the individual hairpins). For each approach, the estimated value that is assumed to be most accurate is shown - i.e. when available the value found using the CFT rather than the JE or a simple average of the work values. If a CFT estimate could not be made due to too much overlap between opening and closing transition work distributions, a simple average was judged to be sufficient to find a good estimate of $\Delta G_{\text{transition}}$. Therefore the JE estimates are not included in the table at

TABLE 7.3: Comparison of $\Delta G_{\text{transition}}$ obtained from various methods (kCal/mol). The values in the first four rows are theoretical predictions of the energy needed to open the originally expected pseudoknot structures, the alternative PK 11/6 U pseudoknot structure, and the individual pseudoknot hairpins. The predictions were made with the programme pknotsRG [61] and the mfold server [77]. The remaining values are the best estimates that could be made for the transition investigated; the method of determining the value is indicated by the subscript on ΔG .

	PK11-6 \pm SEM	PK6-11 \pm SEM		
Predictions	pknotsRG predict of ΔG_{fullPK}	43	46	
	pknotsRG predict of $\Delta G_{\text{PK11/6U}}$	39	NA	
	mfold predict $\Delta G_{\text{hairpin1}}$	22	~ 21.5	
	mfold $\Delta G_{\text{hairpin2}}$	~ 23	29	
Simple approach: 1G	ΔG_{CFT}	not relevant	15.5 ± 1.0	
	2G ΔG_{CFT}	19.4 ± 1.3	15.5 ± 1.9	
	$\Delta G_{\text{simple-}\langle W \rangle}$	57 ± 30	13.6 ± 2.3	
	3G	$\Delta G_{\text{simple-}\langle W \rangle}$	9.0 ± 1.8	
		ΔG_{CFT}	20.2 ± 2.8	
		$\Delta G_{\text{simple-}\langle W \rangle}$	48 ± 11	
Collin approach	ΔG_{CFT}	12.1		
	$\Delta G_{\text{visual-guess}}$	~ 60		
	$\Delta G_{\text{simple-}\langle W \rangle}$		20.9 ± 1.7	

all, as they appeared quite volatile (they are highly affected by the lowest work values of each distribution because the number of data points is so low).

The most important observations are:

- The opening and closing transition is nearly reversible for the full set of transitions for PK 6/11 and for the PK 11/6 low force distributions
- When all rip data are lumped together and treated according to the simple approach, the work to open PK 11/6 appears higher than that required to open PK 6/11; when subdistributions are used clearly one of the 11/6 subdistributions contains much higher energy transitions than the others.
- The simple approach possibly leaves out a portion of transitions requiring a large amount of work because these transitions do not take place in a single clear rip or zip. Such transitions appear better reflected by the Collin approach.
- The Collin approach indicates that the lower force transitions for PK 11/6 require relatively little work and the higher force transitions much more work than the PK 6/11 transitions
- Since the two approaches use different underlying sets of rips and zips, it is difficult to directly compare the work estimates made.
- The estimated values of $\Delta G_{\text{transition}}$ are all far lower than the predicted values except for the estimate from the high work value regime accessed for PK 11/6 with the Collin approach and to some extent with subdistribution A in the simple approach.
- The overall impression is that for PK 11/6 subdivision of the data into several underlying transitions with different associated transition energies is reasonable. There may be two or three transitions. The values in the table fall in three general ranges: The low distribution of the tripple Gaussian subdistribution used in the

TABLE 7.4: Calculated kinetic parameters with asymptotic standard error. Values from F-dependent fits (highlighted) are considered most reliable. Note that the asymptotic standard error underestimates the error because it assumes that k_0 and x^\ddagger are uncorrelated, which they are not.

	PK 11/6 \pm ASE	PK 11/6 \pm ASE
k_0 from $r \ln P$ (s^{-1})	0.21 \pm 0.01	0.0009 \pm 0.0002
- weighted fit to same	0.05 \pm 0.003	0.0004 \pm 0.0001
k_0 from F-dependent fit (s^{-1})	0.040 \pm 0.004	0.006 \pm 0.001
x^\ddagger from $r \ln P$ (nm)	0.27 \pm 0.01	2.25 \pm 0.07
- weighted fit to same	0.62 \pm 0.01	2.07 \pm 0.05
x^\ddagger from F-dependent fit (nm)	0.66 \pm 0.02	1.56 \pm 0.08
$\langle x^\ddagger \rangle$ from Dudko method (nm)	0.39 \pm 0.08	0.69 \pm 0.2

simple approach ($\Delta G_{\text{PK-simple-}\langle W \rangle-3G}$) and the lower work distribution using the Collin approach ($\Delta G_{\text{PK-Collin-Crooks-lower}}$) result in estimates of about 10 kCal/mol, the simple approach treatment of the full distribution and the double Gaussian subdistribution both result in estimates of about 20 kCal/mol, and the high work value distribution treated by the Collin approach results in an estimate of about 60 kCal/mol.

- For PK 6/11 it is harder to see evidence of the presence of different transition energies and it is difficult to tell whether the averages found using the simple or the Collin approach is more accurate; both have weaknesses due to the way they were calculated.

For comparison of the range of values found, values of 18.6 and 22.5 kCal/mol were found for similar pseudoknots PK400 and PK401 by Hansen et al. [17] by the Jarzynski-method and 24.8 kJ/mol was found for PK401 using the Crooks Fluctuation Theorem. These values appear to agree well with those found in the present work. Green et al. [25] find $\Delta G_{\text{transition}}$ for four different pseudoknots also derived from Infectious Bronchitis Virus in the range of 19–34 kCal/mol, also in the same range as the values found here. Like the values presented here, the values from [17] and [25] are based on experiments in the presence of Mg²⁺ ions; another pseudoknot study by Chen et al [21] does not use Mg ions, which is expected to have a large stabilizing effect, so this study cannot be easily compared to the values here.

The work distributions and energies and the implications for possible different transitions observed will be discussed further in Chapter 6.2.

7.2 Kinetics: distance to transition state

The distance to the transition state, x^\ddagger , and the standard rate of reaction, k_0 , were investigated following the approaches described in Theory 3.3: first using the "rlnP" method used in Hansen et al.[17], then attempting to use the method described in Dudko et al.[52]. Results are summarized in table 7.4.

The fits performed to calculate the values in the table using the rlnP-method are seen in Figure 7.6. As may be seen in the figures, three different fitting methods were used. Initially, a weighted and an unweighted fit was made to Equation (3.7). The difference is that in the weighted fit, the data for rlnP are weighted by $1/P$. Thus the uncertainty is assumed to be the same for all the values of P used to calculate the values of rlnP, making the uncertainty in rlnP proportional to $1/P$. If no weights are used, the larger values of P corresponding to high values of F will disproportionately determine the fit. However, it is

difficult to determine what the uncertainties on values of P are as P is found directly from the cumulative probability distribution for the data. It is thus not $\ln P$ that is uncertain but the corresponding value of F . Fitting to the equation

$$F = \frac{k_B T}{x^\ddagger} \ln\left(1 - \frac{x^\ddagger * r \ln P}{k_0 k_b T}\right) \quad (7.1)$$

with F as the dependent variable rather than $r \ln(P) = \frac{-k_0}{x^\ddagger k_b T} \exp\left(\frac{x^\ddagger * F}{k_b T} - 1\right)$, where P is the dependent variable avoids this problem. No weighting of the data is necessary in this fit; the fits found from fitting to $r \ln P$ with and without weights are shown for comparison. This appears to be the most reliable of the fits and thus these values are probably the most dependable.

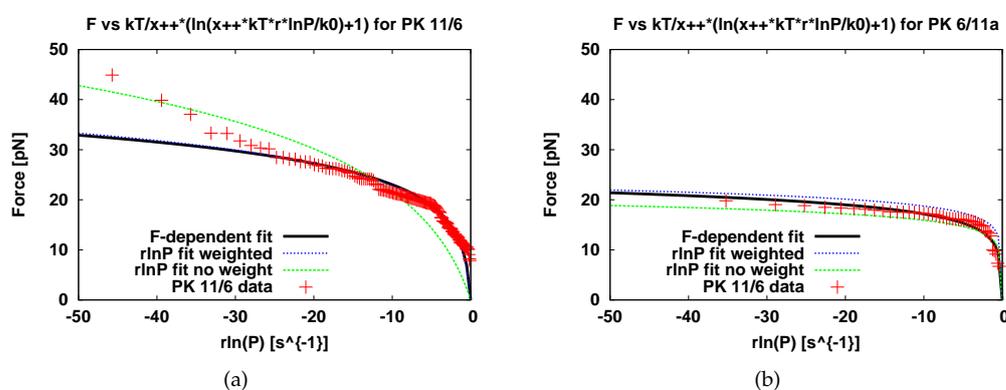


FIGURE 7.6: Fits to Equation (7.1) to find the parameters x^\ddagger and k_0 (called F-dependent fit) plus weighted and unweighted fits to Equation (3.7) for: (a) PK 11/6; (b) PK 6/11.

Note that the uncertainty given in table 7.4 is the asymptotic standard error of the parameters found by the fitting programme Gnuplot. The asymptotic standard error assumes that the parameters are linearly independent and therefore underestimates the uncertainty of x^\ddagger and k_0 which are quite strongly dependent on each other.

7.2.1 Kinetic parameters: Subdistributions and reverse transitions

Beyond calculation of the kinetic parameters for the full distribution of rip forces for each pseudoknot, the kinetic parameters for the subdistributions found in Chapter 5.4.2 were also found. This appeared especially relevant for PK 11/6, as the data fitted in e.g. Figure 7.6 (a) do not appear to lie on a single line corresponding to a single one-step reaction. Additionally, the full distributions of zip forces were also used to find kinetic parameters for the reverse distributions. All of these additional calculations were made using F-dependent fitting since this was the most reliable.

The resulting fits are shown in Figures 7.7 and 7.8 and the kinetic parameters found are shown in Table 7.5. For ease of comparison with the values found above for the full rip datasets using F-dependent fitting are also shown again.

The subdistributions made for PK 11/6 do not appear to capture completely the underlying transition distinctions; the data points still do not lie on a curve that corresponds to the theoretical curve for a one-step reaction. However, both for the tripple and double Gaussian subdistributions, the high-force subdistribution does appear to lie on a distinct

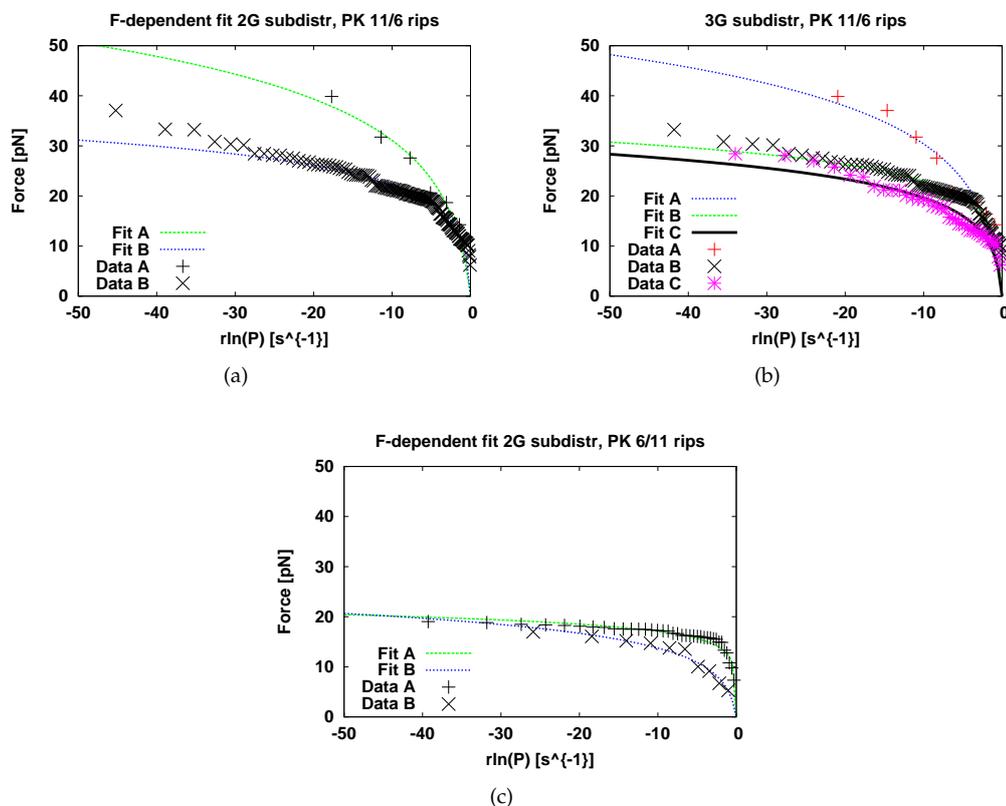


FIGURE 7.7: Subdistribution fits to $r\ln(P) = \frac{-k_0}{x^\ddagger kT} \exp\left(\frac{x^\ddagger * F}{kT} - 1\right)$ to find the parameters x^\ddagger and k_0 for (a) PK 11/6 double Gaussian subdistributions; (b) PK 11/6 tripple Gaussian subdistributions; and (c) PK 6/11 double Gaussian subdistributions. A, B, and C refer to the distributions of transitions shown in Figures 5.8 and 5.9

curve. Interestingly, *the fit to this curve gives quite different parameters than the other curves, indicating that the high force transition could involve a more brittle structure than the remaining transitions.* If the high force data that does not appear to lie on the portion of the curve matching the theoretical shape were excluded, the value found for x^\ddagger would possibly be even higher.

The subdistributions for PK 6/11 also appear to lie on distinct curves. Here the lower force transitions are the ones that appear more brittle, though they are still less brittle than any of the PK 11/6 subdistributions.

The fits made to the zip force data appear very well fitted by the shape of the theoretical curve. Strikingly, for PK 6/11 the parameter value found for x^\ddagger is almost the same as for the full rip force data set, matching nicely the image of a reversible transition as observed in the work distribution estimation and Gibbs free energy calculation. In contrast, the parameter value for x^\ddagger for the zip transitions PK 11/6 corresponds to a much less brittle structure than the rip transition x^\ddagger .

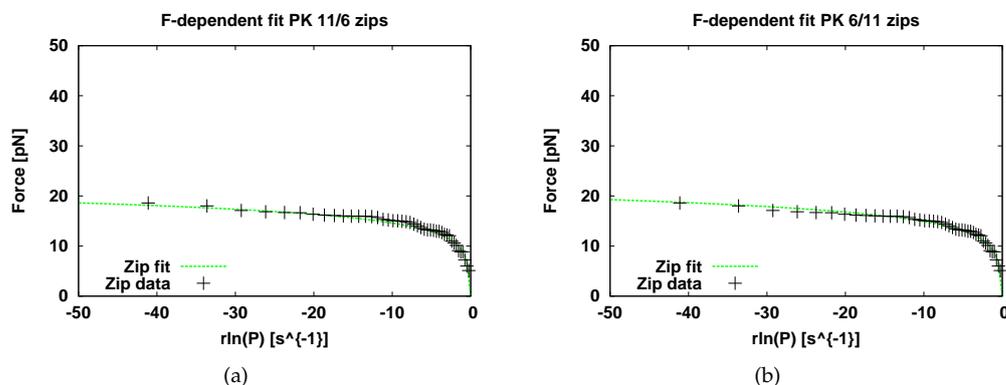


FIGURE 7.8: Zip transition parameter fits: F-dependent fits to $r\ln(P) = \frac{-k_0}{x^\ddagger kT} \exp\left(\frac{x^\ddagger * F}{kT} - 1\right)$ for finding the parameters x^\ddagger and k_0 for the reverse (zip) transitions. (a) PK 11/6. (b) PK 6/11.

TABLE 7.5: Calculated kinetic parameters with asymptotic standard error: full and subdivided rip datasets, zip dataset. Note that the asymptotic standard error underestimates the error because it assumes that k_0 and x^\ddagger are uncorrelated, which they are not.

	PK 11/6 \pm ASE	PK 6/11 \pm ASE
k_0 : all rip data (s^{-1})	0.040 ± 0.004	0.006 ± 0.001
k_0 : 2G subdistr A (s^{-1})	0.07 ± 0.02	0.001 ± 0.0004
2G subdistr B (s^{-1})	0.03 ± 0.002	0.10 ± 0.02
k_0 : 3G subdistr A (s^{-1})	0.007 ± 0.02	
3G subdistr B (s^{-1})	0.018 ± 0.002	
3G subdistr C (s^{-1})	0.052 ± 0.008	
x^\ddagger : all rip data (nm)	0.66 ± 0.02	1.56 ± 0.08
x^\ddagger : 2G subdistr A (s^{-1})	0.32 ± 0.06	1.98 ± 0.1
2G subdistr B (s^{-1})	0.75 ± 0.02	0.94 ± 0.09
x^\ddagger : 3G subdistr A (s^{-1})	0.36 ± 0.07	
3G subdistr B (s^{-1})	0.85 ± 0.02	
3G subdistr C (s^{-1})	0.75 ± 0.05	
k_0 : all zip data (s^{-1})	0.012 ± 0.001	0.015 ± 0.001
x^\ddagger : all zip data (nm)	1.64 ± 0.04	1.51 ± 0.03

7.2.2 Comparison of kinetic parameters to literature values

The values found indicate that PK 11/6 is more brittle (lower x^\ddagger) and less stable (higher k_0) than PK 6/11. In experiments very similar to the ones presented here, Hansen et al. [17] found $k_0 = 0.16 \pm 0.08 s^{-1}$ and $x^\ddagger = 0.18 \text{ nm} \pm 0.06$ for one pseudoknot and $k_0 = 0.074 \pm 0.007 s^{-1}$, and $x^\ddagger = 0.19 \text{ nm} \pm 0.01$ for another. Green et al. [25] found $k_0 = 0.00012 s^{-1}$ for one of their pseudoknots and x^\ddagger in the range of about 2-6 nm for four different

pseudoknots. Chen et al. 2007 [18] find $x^\ddagger = 0.8$ nm for the pseudoknot they investigate with force-jump and force-drop experiments. The experiments by Chen et al. were done without the presence of Mg ions in contrast to the two other studies, so they may not be comparable. The others are quite similar to the values found here, with the numbers for PK 11/6 more closely resembling the brittle pseudoknots observed by Hansen et al. and PK 6/11 apparently more similar to the more stable and less brittle pseudoknots investigated in Green et al.

For comparison with RNA hairpins, the value for x^\ddagger found by Liphardt et al. for a simple RNA hairpin 23 basepairs long (an A-form helix) was 11.5 nm and that of a hairpin with an additional helix (a three-helix junction) was 12 nm while they found that the distance to transition state for a slightly modified three-helix junction with an extra A-rich bulge was only 1.6 ± 0.1 nm [4]. Woodside et al. found a range from about 3-20 nm for x^\ddagger for the unfolding of a series of DNA hairpins of varying stem and loop lengths [78]. For an RNA hairpin of stem length 21 with a few bulges, Li et al. found x^\ddagger for the unfolding transition of 7 ± 5 nm [79].

These values will be examined more closely in Section 6.2. For now let us conclude that the values found for x^\ddagger for PK 11/6 and PK 6/11 certainly do not indicate that the structures unfolded were simple hairpins rather than pseudoknots. However they also do not rule out the possibility that at least PK 6/11 could be folded as a complex type of hairpin rather than a pseudoknot.

As a check of the validity of the values found for x^\ddagger and k_0 (though only for the full subdistributions of data), Equation (3.8) is used to predict the distributions of unfolding forces for each pseudoknot at the experimental value of r . See predicted distributions in Figure 7.9. The shapes of these predicted distributions are quite different from the observed distributions, especially in the case of PK 11/6 which extraordinarily closely resembles a Gaussian distribution. This confirms the picture seen already in the scatter plots of force versus extension in Figure 5.8, namely that the transition observed for PK 11/6 is not a simple two-state transition. It might be that the single Gaussian appearance of the distribution of unfolding forces for PK 11/6 is in fact hiding a sum of (at least) two distributions as hinted by Figure 5.8 (a).

7.2.2.1 Dudko method

An alternative equation that can be used to find the kinetic parameters for the molecular transition was shown in Equation (3.9). This equation by Dudko, Hummer and Szabo ([49], here called the Dudko method, relies on estimating $\tau(F) = \frac{1}{k(F)}$ from the histogram distribution of unfolding forces, $p(F)$. The method tries to extract more information about the energy landscape than the "rlnP" method based on Bell's formula that is used above. It was attempted here, but because of the low number of data, the results are not very conclusive.

No value for k_0 was found. The data was simply too sparse as with only very few histogram bins the values obtained for $\tau_0 = \frac{1}{k_0}$ (and ΔG_0^\ddagger) were more or less arbitrary (from fitting to Equation (3.9) using Gnuplot). However, using $\nu = 1$ for a range of histogram bin widths and starting point values for $\tau_0 = \frac{1}{k_0}$ and ΔG_0^\ddagger , a range of values were obtained for x^\ddagger which appeared consistent enough to be averaged and included in Table 7.4. The parameter ν describes the assumptions made about the underlying energy landscape and $\nu = 1$ corresponds to making no assumptions about the shape of the energy landscape just as when the "rlnP" method is used. The values of $\tau(F)$ used for the fit were found using histograms with 5 to 9 bins. No weighting was performed for the fit, although this might be possible based on the uncertainty of the values found for $\tau(F)$. Figure 7.10 shows the fit made in Gnuplot to Equation (3.9) with $\nu = 1$ for selected histogram estimates of $\tau(F)$.

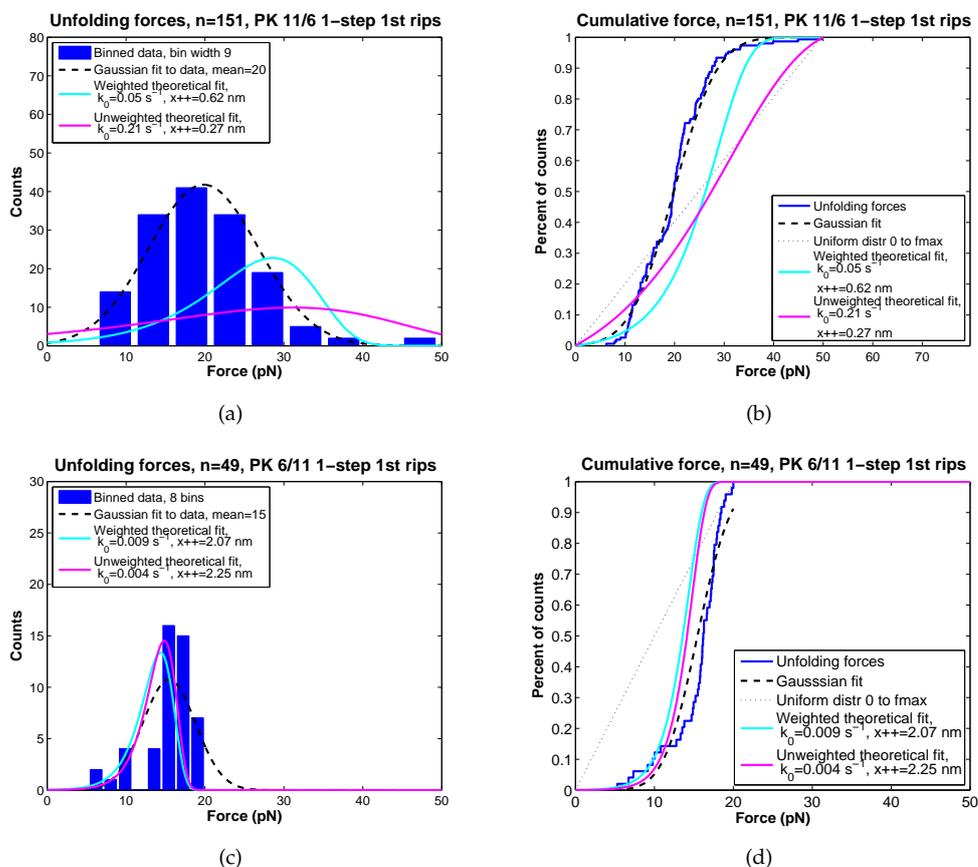


FIGURE 7.9: Prediction of the distribution of unfolding forces using the values of x^\ddagger and k_0 found with the rlnP method for PK 11/6 and PK 6/11.

It is worth noting that part of the difficulty in fitting to the sparse data for $\tau(F)$ might derive from the fact that the distribution of unfolding forces in fact was not that of a two-state transition and the underlying assumptions of Equation (3.9) are not fulfilled. Without weights the fit is highly sensitive to variation of especially the largest values of $\tau(F)$ which occur at low forces. If the distribution does not correspond to theory in this region whether due to random variation or due to a fundamentally different underlying distribution, the values found for $\tau_0 = \frac{1}{k_0}$ and ΔG_0^\ddagger will not be reliable in the least.

Using the Dudko method with different values of ν to find x^\ddagger and k_0 for differently shaped energy landscapes failed entirely - the fitting was too unstable even to estimate x^\ddagger . The point of the method is to investigate the dependence of the parameters on the shape of the energy landscape, so this was unfortunate. However, the effect of varying ν versus the effect of varying x^\ddagger and τ_0 is investigated in Figure 7.11 and it is seen that the effect of varying ν is insignificant compared to the effect of varying τ_0 and to a lesser extent x^\ddagger . The extreme variation in the fit with a relatively small change in τ_0 illustrates why it was possible to find a fairly stable ballpark estimate of x^\ddagger by fitting sets $\tau(F)$ from different histogram estimates to Equation (3.9), whereas it was not possible to find a stable range of values for τ_0 .

All in all, however, this theoretically highly interesting method of extracting extra kinetic information from the data cannot be said to have provided extra insight into the data set presented here.

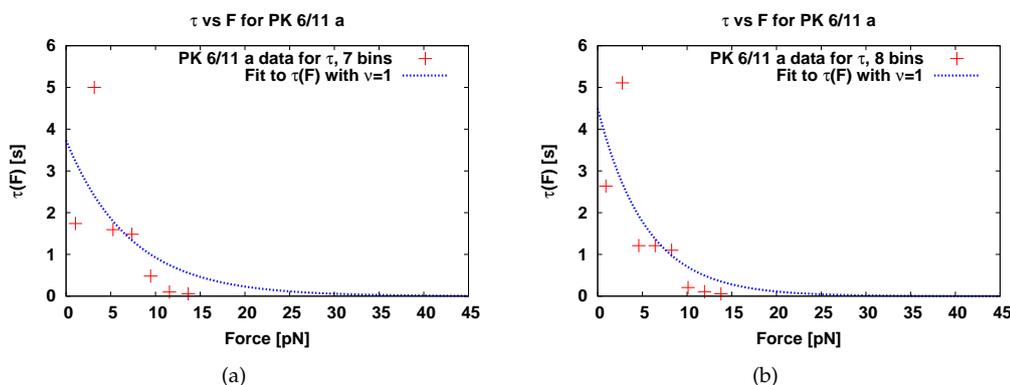
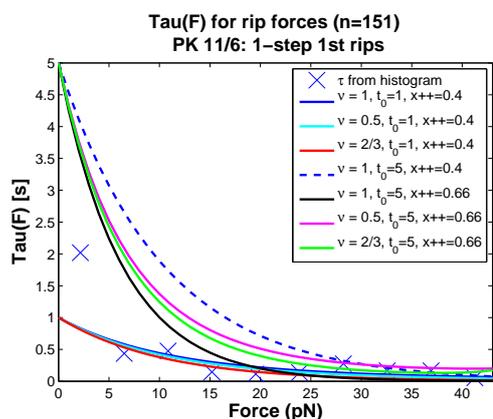


FIGURE 7.10: Selected fits using Dudko method to $\tau(F) = \frac{1}{k(F)}$: (a) Fit for PK 11/6 using values for $\tau(F)$ found using 7 bins of force data. (b) Fit for PK 6/11 using values for $\tau(F)$ found using 8 bins of force data.

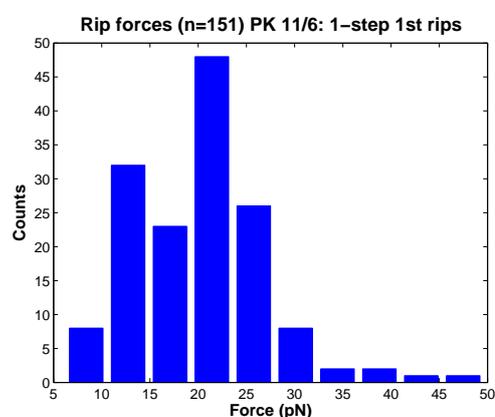
7.3 Summary of points made in this chapter

In this chapter we have seen that

- The work energy for the transitions represented by the main distributions of data points are unexpectedly low for both pseudoknots - much lower than the predicted ΔG for the full transition from open to closed pseudoknot.
- The transitions observed for the lower value work distributions for PK 11/6 and for the full dataset for PK 6/11 were close to reversible with apparently very similar unfolding and refolding transitions.
- The Gibbs free energy change calculated using the full set of rip data for PK 11/6 was significantly larger than that calculated using the full rip data set for PK 6/11
- The found Gibbs free energy changes for the various possible transitions clustered in three size ranges: about 50-60 kCal/mol (higher than the prediction even for the full pseudoknots), about 20 kCal/mol (corresponding well to the values of opening energy for hairpins, especially when it is considered that the energy subtraction for stretching the single strand is too high), and a possible distribution all the way down at about 10 kCal/mol, possibly also corresponding to a hairpin when it is considered that the energy subtraction for W_{stretch} is too high.
- The rate of pseudoknot opening at zero force was higher for PK 11/6 than for PK 6/11, though not very much
- The distance to transition state for both pseudoknots was much lower than the values found in literature for RNA hairpins. This means that the structures are more brittle, i.e., they cannot be deformed significantly without breaking.
- The distance to transition state was lower for PK 11/6 than for PK 6/11, indicating that PK 11/6 is more brittle than PK 6/11
- The difference in brittleness might indicate that PK 6/11 is more hairpin-like than PK 11/6.



(a)



(b)

FIGURE 7.11: Experimental $\tau(F) = \frac{1}{k(F)}$ found using Dudko method for 10 bins for PK 11/6 full data set. (a) plots of theoretical $\tau(F)$ for various variable values. The differences between the curves due to using different shape parameters are much smaller than the differences from using different values of τ_0 . Variation in values for x^\ddagger and ΔG^\ddagger has less influence within the range of values where they give real function outputs. (b) Histogram bins used to find values of $\tau(F)$.

- If data is divided into subdistributions, it becomes apparent that for PK 11/6 transitions occur associated with very different Gibbs free energies - a few apparently requiring somewhat more energy than the full opening of the originally predicted pseudoknot.
- The high-force (and high work energy requiring) transitions also appear associated with a more brittle structure than the remaining transitions for PK 11/6.

In the next chapter we will interpret these results.

8 Discussion Part II: energy and kinetics

Like the unfolding force distributions, the Gibbs free energy changes and kinetic parameters calculated for the pseudoknots in Chapters 7.1 and 7.2 point to some important differences between the pseudoknots. These points were summarized at the end of Chapter 7. The most important points are revisited and drawn together here.

As mentioned in Chapter 7.1, the calculated values of the Gibbs Free energy is not too far from the values found by Hansen et al. [17] and Green et al. [25] for similar pseudoknots. The values are lower than expected from the calculations made with computer programs that can predict the folding free energy of RNA pseudoknots based on their primary sequence and the difference in values was also not expected from these programmes.

As also mentioned, the reason the measured energy is lower than the predicted energy - not only for the pseudoknot structures but even for hairpin predictions meant to match the shorter low force transitions - could derive from an overestimate of the amount of energy going into stretching the molecule. This would match the hypothesis that the structure being opened and closed is usually not the expected full pseudoknot structure but one that has a lower energy of formation - i.e. mostly hairpins.

The different degrees of brittleness found for the different subdistributions of transitions for PK 11/6 match the idea that the structure requiring high force to unfold and displaying long rips and zips is indeed a pseudoknot as it is very brittle. The other transitions match the brittleness cited in literature not for simple hairpin structures but for other pseudoknots and complex hairpin structures, so it may be that these are the types of conformations seen.

All in all the force and kinetic data lend support to the idea that only the high force long rips for PK 11/6 actually represent the originally expected pseudoknot structure whereas the other rips could well represent only one stem opening/closing (a hairpin-like structure) or an alternative structure altogether.

8.1 Energy Landscapes and intermediate states

The rip force, rip length and free energy change and kinetic data presented all indicate that more than just a simple transition from closed to open state and back is taking place during the pseudoknot unfolding and refoldings. As in [18], a schematic energy landscape may be drawn reminding us that the transition from closed to open pseudoknot is expected to be at least a three state transition. See Figure 8.1. While the energy landscapes of RNA hairpin helices have been described as relatively smooth, allowing easy structural transitions through unzipping and re-zipping of single basepairs, the energy landscape of pseudoknots are expected to be more "rugged" with sharply favorable conformations in some areas along the reaction coordinates and smooth transitions between different structural conformations in others [51, 18]. This certainly may appear to be the case for the PK 11/6 pseudoknot which apparently only rarely takes on the conformation of the full pseudoknot that has a relatively high $\Delta G_{transition}$ but apparently may more easily take

on and transfer in and out of the conformation of both an alternative, shorter pseudoknot (perhaps) as well as different hairpin-like structures. In contrast, the energy landscape for the PK 6/11 sequence may be smoother - or with a very high barrier to pseudoknot formation, so high that it is only very rarely overcome at all and the structure is not encountered experimentally.

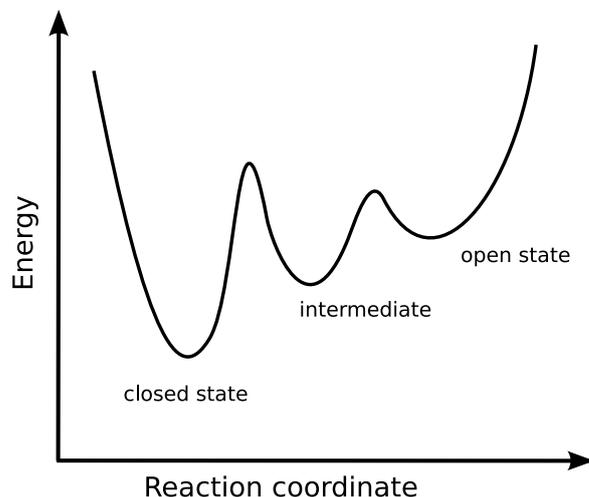


FIGURE 8.1: Energy landscape for a hypothetical three-state transition from closed to open pseudoknot. The reaction coordinate is an abstract coordinate along which the transition takes place. There will probably be several different possible paths with different intermediate states with different free energy and different barrier heights between states. The exact path the pseudoknot takes will be partly determined by chance. Thus it may sometimes get “stuck” in intermediate states.

Much research is currently taking place on the possible tertiary structure of pseudoknots, both on how it may be predicted and how it may affect their physiological roles. Chen et al. [21] suggest that triplex formation between the stems and loops of the pseudoknot may stabilize their structure. These triplexes consisted of UA-U, GC-C and GU-C stem base pairs interacting with loop bases. In Figures 6.1 (a) and 6.2 (b), possibilities for such triplex formation in the expected pseudoknot structures may be examined. It appears that there are many such possibilities for both originally suggested pseudoknot structures, but that there are most for PK 11/6. Thus this may be part of the reason that the pseudoknot structure is encountered for PK 11/6 but not for PK 6/11.

See e.g. [15] for further discussion of the possible stabilizing effect of triplexes and quadruplexes at the bases of the helix junctions where the stems of pseudoknot structures come together.

9 Conclusion and perspectives

This chapter will draw up conclusions based on the results and discussions presented in this thesis. Additionally, a few perspectives and ideas for further investigations will be presented.

9.1 Background for the conclusion

This thesis originally set out to investigate how single stranded RNA pseudoknot structures unfold when pulled by optical tweezers, whether the different RNA structures of PK 11/6 and PK 6/11 would show different unfolding patterns, and whether it would be possible to relate any differences in the unfolding patterns to the differences in ribosomal frameshifting efficiency of the two structures.

9.1.1 Basic differences

The first observation made from the experiments was that for both pseudoknot structures, sudden unfolding and refolding events (rips and zips) were observed during pulling experiments, corresponding to changes in conformation and length of the single stranded RNA being investigated. Enough unfolding and refolding events were observed for each structure that they could be compared quantitatively and it became clear that PK 11/6 and PK 6/11 are different in strength and folding pattern.

From the outset of the investigation it was known that PK 11/6 causes ribosomal frameshifting while PK 6/11 does not. It is also known that a relatively long stem1 is important for a pseudoknot to be able to cause efficient frameshifting [59], and it was therefore hypothesized that the short stem1 of PK 6/11 was the reason it did not cause frameshifting. Based on structural prediction, the two pseudoknots were expected to be equally resistant to pulling, however, since the expected pseudoknot structures were almost symmetric. The pulling experiments in this thesis were therefore carried out to test whether PK 11/6 would nonetheless turn out to be stronger than PK 6/11 and that pseudoknot strength would therefore correlate with ribosomal frameshifting efficiency as proposed by Hansen et al. [17]. Alternatively it was conjectured that PK 11/6 and PK 6/11 might be equally strong and the difference in frameshifting efficiency might arise from the reversal of the structure encountered by the ribosome.

The results show that the two structures are not structurally symmetric as expected. Instead, the structural conformations encountered for the two pseudoknots appear quite different from each other and each RNA sequence apparently is able to pass among several possible conformations.

9.1.2 Pseudoknot versus hairpin

A structure that displays the full rip or zip length of the originally expected pseudoknot was only observed for PK 11/6. In contrast, PK 6/11 never unfolded or refolded in the manner that would be expected for the full pseudoknot structure. Instead, the unfoldings and refoldings observed for PK 6/11 corresponded well to the changes in length that would be expected if stem2 and loop2 of the originally expected pseudoknot were to unfold or refold suddenly. Thus under the experimental conditions encountered here, PK 6/11 did not behave as a pseudoknot but rather as an RNA hairpin.

Although PK 11/6 at times unfolded in the manner predicted for the originally expected pseudoknot, it too may sometimes have occurred as only a single hairpin made up of stem1 and loop1 of the originally predicted structure. Additionally it may have occurred as an alternative, shorter pseudoknot structure (PK 11/6 U), formed by the originally expected stem1 and loop1 plus a second stem and loop made up of the originally predicted loop2. This alternative pseudoknot structure was observed to be stronger than the hairpin but still much weaker than the originally predicted pseudoknot, even though predictions of the free energy of the two pseudoknots were relatively similar.

These observations were supported by subdivision of the unfolding and refolding data for both pseudoknots into long and short transitions which yielded different sets of estimates of the Gibbs free energy of formation and of the kinetic parameters characterizing the structural transitions.

9.1.3 Unexpected structures

Overall the conclusion is that frameshift efficiency and structural strength as characterized by the force required for unfolding do appear to correlate for PK 11/6 and PK 6/11. However, the reason for the correlation appears to be that PK 6/11 does not form a pseudoknot structure at all (or at least it was not observed to do so under the experimental conditions here and it may therefore be expected that PK 6/11 forms a pseudoknot structure much more rarely than PK 11/6 does). If it is true that PK 11/6 can form a pseudoknot while PK 6/11 cannot, this may explain the difference in their frameshifting efficiency: PK 6/11 does not cause frameshifting simply because it is not a pseudoknot and therefore does not block ribosomal progress enough to cause the ribosome to slip backwards.

The occurrence of shorter two-step transitions and frequent hopping back and forth between different states for PK 6/11 indicates that it may not always fold or unfold as a simple hairpin but sometimes behaves as two hairpins that are opened and closed independently during pulling, or that it occurs in other relatively "soft" conformations that are easily altered.

Likewise the many two-step transitions observed for PK 11/6 indicate that it may sometimes behave as two hairpins being opened independently or - perhaps more likely - that it moves from one conformation to another, occasionally even transitioning into the full pseudoknot structure while being pulled, only to transition to the fully open single strand when the force is increased enough.

Taken together with the work undertaken by Jesper Tholstrup on related pseudoknot structures, this thesis shows that many different conformations of the same RNA single strand sequence may be encountered in a buffer that resembles physiological buffer conditions. These different conformations may display very different resistance to mechanical stress and may therefore affect translating ribosomes differently. Perhaps these structures sometimes cause frameshifting and sometimes not depending simply whether they are folded as pseudoknots when encountered by a ribosome.

These observations correspond very well to the conclusions of other RNA pseudoknot pulling studies proposing the observation of intermediate folding states [18, 21]. It also matches the conclusion of RNA modeling efforts which suggest that the energy landscape of different RNA folding conformations may allow easy transition from a hairpin folding state to a pseudoknot folding state [16].

9.1.4 Predicting tertiary structure

It is clear from this thesis work and the work of Jesper Tholstrup that currently existing programs for predicting pseudoknot structure cannot tell us which of the many possible structural conformations we will encounter under physiological conditions nor how efficiently a potential pseudoknot-forming RNA sequence will promote frameshifting. Tertiary interactions between the bases that make up the structure cannot yet be accurately predicted but are apparently very important, as these must cause the main differences between PK 11/6 and PK 6/11.

As noted, Chen et al. [21] suggest that differences in the strength and frameshifting ability in a series of pseudoknots were caused by triplex formation between bases, a type of tertiary interaction. The formation of this hypothesis was aided by the knowledge of the crystal structure of the basic pseudoknot that this investigation was based on, as the crystal structure suggested the presence of such triplexes. The crystal structure of the Infectious Bronchitis Virus (IBV) pseudoknot (the wild-type origin of the pseudoknots studied here) has not been elucidated, but it is still possible to speculate that triplex formation influences the stability of IBV pseudoknot variants.

9.1.5 Conclusion summary

1. PK 11/6 and PK 6/11 *are* different in strength and structure as well as in frameshifting ability.
2. As hypothesized based on frameshift efficiency, PK 11/6 on average is stronger than PK 6/11.
3. PK 11/6 was observed to open at least occasionally as the originally predicted pseudoknot - a brittle but strong structure. The PK 11/6 nucleotide sequence also appeared to occur relatively often in the conformation of a shorter pseudoknot, PK 11/6 U. Finally, the sequence sometimes appeared to occur only as a single hairpin or to open one hairpin at a time.
4. PK 6/11 was always observed to open hairpin by hairpin or to occur simply as a hairpin structure made out of the expected stem2 and loop2 of the originally expected structure. Thus it may not usually occur as a pseudoknot at all. The structures formed by the PK 6/11 sequence generally appear weaker and more elastic than the PK 11/6 structures.
5. The structures formed by the pseudoknot-containing RNA single strands vary widely and the most common conformations are not easily predicted.
6. The PK 11/6 nucleotide sequence may be more likely to form a pseudoknot than the PK 6/11 sequence because it contains more possibilities for base triplex formation between loops and stems that would contribute to its structural stability.
7. The observation of multiple folding states match the conclusions drawn by Jesper Tholstrup in his investigations of related pseudoknots and the observations made in two pseudoknot pulling studies by Chen et al. [18, 21].

9.2 Perspectives

The results shown in this thesis underscore how uncertain structural predictions of RNA folding patterns still are and in particular they draw attention to the dynamics of molecular folding patterns. At least in the case of single stranded RNA, structural prediction is not only about finding one particular most likely molecular conformation but about finding a range of likely structures and being able to predict how easily the molecule moves from one conformation to another.

A better understanding of the dynamics of mRNA conformations may in a larger context help shape our knowledge of how ribosomal translation takes place. Unending possibilities abound for expanding the investigations made in this thesis to better understand these dynamics. One strategy is to dive further into the detailed dynamics of the RNA nucleotide sequences investigated here, confirming and expanding the conclusions made here by investigating the folding patterns of the individual hairpins expected in each sequence. Another strategy is to continue the investigation with a wider range of pseudoknots with different systematic variation in structure. Finally a third option would be to expand the investigation to directly probe the interaction between ribosomes and pseudoknots.

9.2.1 Further pseudoknot pulling investigations

A logical continuation of the investigations made here would be to make molecular constructs whose single strand portion consists only of the truncated pseudoknot forming part of the sequence of PK 11/6, the sequence of stem1 and loop1 of PK 11/6, and the sequence of stem2 and loop2 of PK 6/11. Such experiments might verify whether the folding and unfolding observed here corresponds to the opening and closing of these predicted substructures.

An interesting additional investigation would be to perform pulling experiments on a construct whose single strand portion corresponds to the wild-type IBV frameshift-inducing pseudoknot to see whether this highly frameshift-inducing pseudoknot more often in fact occurs as a pseudoknot than the structures investigated here. Such experiments should perhaps also be carried out on the individual hairpins expected to make up the wild-type pseudoknot.

Finally it may be possible in future experiments to try to design IBV pseudoknots with systematic variation in the number of expected base triplexes between loops and stems. Examining the frameshift and stall efficiencies as well as the mechanical strength of such a sequence of pseudoknots in the manner used in the present study could help illuminate the importance of this particular type of tertiary interaction for the overall most favorable structures of RNA single strands. If investigations are undertaken to explore tertiary interactions more systematically, it would be worthwhile to further study existing models attempting to predict structure by including tertiary interactions, including the Vfold model developed by Cao et al. [80].

9.2.2 Thermodynamic investigations - diving further into the details of each molecular transition

An interesting bonus of performing pulling experiments with simpler structures than the pseudoknots investigated here, i.e., RNA hairpins, is that the results of such experiments might better lend themselves to thermodynamic analysis along the lines proposed by Dudko et al [52] and by Hummer and Szabo [81]. In this type of analysis the underlying energy landscape of the molecular transition is investigated and mapped, and the apparent

energy of activation is found. This was difficult in the present study because so many different transitions were taking place which could not easily be separated. Investigating a structure that forms fewer intermediates might make the analysis easier.

If investigations are made of the hairpins and the smaller pseudoknot that might constitute the intermediate states observed here for PK 11/6, they might yield reaction rates and kinetic parameters characterizing some of the subtransitions observed. This again might enable a much clearer understanding of the possible folding pathways of PK 11/6 - and perhaps also of its possible interactions with a ribosome.

If these investigations were also undertaken for PK 10/6 (examined in parallel to PK 11/6 by Jesper Tholstrup), these investigations together might yield a much better understanding of the folding dynamics of the nucleotide sequences of PK 10/6 and PK 11/6 since they are so closely related.

9.2.3 Experimental improvements and ideas for expanded pulling experiments

If further pulling experiments were to be carried out, it would be a great advantage to eliminate some of the experimental problems experienced with the NanoTracker setup. In particular, it would be desirable to eliminate the noise of the galvanic mirrors controlling the trap positions by implementing a different type of trap movement control, perhaps using one stationary trap and one trap controlled by an Acousto-Optical Deflector. Eliminating this noise would first of all improve calibration, but it might also make it easier to obtain stable molecular tethers because the beads would be moving much less in the traps, allowing slower molecular interactions. Such a change in the experimental setup, however, would require rebuilding the NanoTracker or building a new dual trap setup.

If rebuilding of the setup were undertaken, another improvement would be better control of the z-height and trap strength through possibilities for aligning the lasers and adjusting their focus. Additionally, a simple improvement can be made by precise quantification of the intensity loss of the photo diode at high intensities so that it may be corrected for during calibration as suggested by [39]. Improved data analysis would also be achieved if the identification of rips and zips on the force-extension curves were done by a formal filtering routine

Finally, to expand experimental possibilities it would also be extremely useful to have a feedback system built into the trap control so that the force on the beads could be maintained at a particular level. This would allow constant force and force-drop/force-jump experiments, greatly expanding the range of experiments for investigating transition kinetics. Together with a flow chamber system that allows flushing in of new reactants (possible to buy for the NanoTracker), a feedback system enabling constant force experiments could even make it possible to attempt the observation of real time *in vitro* frameshifting of translating ribosomes. To do this, a molecular construct containing an RNA pseudoknot would have to be held between two beads at constant force. One could then observe the rate at which the pseudoknot opens and closes passively and contrast it to the rate at which it opens and closes in the presence of translating ribosomes. To allow ribosomes to translate the structure, however, much work would have to be done, as the molecular construct would need modification so that the ribosome would attach and start translating, and the buffer would have to contain tRNAs, amino acids, and other components needed for translation. Such an experiment would be similar to the ones carried out by Wen et al. and Qu et al. [13, 14], who observed opening of RNA hairpins by translating ribosomes.

Bibliography

- [1] L. B. Oddershede. Optical tweezers techniques. In Henrik G. Bohr, editor, *Handbook of Molecular Biophysics. Methods and Applications*, pages 197–213. WILEY-VCH Verlag GmbH & Co. KGaA, 2009.
- [2] S.M. Block. Making light work with optical tweezers. *Nature*, 360(6403):493–5, 1992.
- [3] T T Perkins. Optical traps for single molecule biophysics: a primer. *Laser & Photonics Review*, 3(1-2):203–220, 2009.
- [4] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr., and C. Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292(5517):733–737, 2001.
- [5] M.C. Williams and I. Rouzina. Force spectroscopy of single DNA and RNA molecules. *Current opinion in structural biology*, 12(3):330–336, 2002.
- [6] M.D. Wang, M.J. Schnitzer, H. Yin, R. Landick, J. Gelles, and S.M. Block. Force and velocity measured for single molecules of RNA polymerase. *Science*, 282(5390):902, 1998.
- [7] A. Ashkin. Optical trapping and manipulation of neutral particles using lasers. *Proceedings of the National Academy of Sciences*, 94(10):4853, 1997.
- [8] A. Ashkin, JM Dziedzic, JE Bjorkholm, and S. Chu. Observation of a single-beam gradient force optical trap for dielectric particles. *Optics letters*, 11(5):288–290, 1986.
- [9] A. Ashkin, J. M. Dziedzic, and T. Yamane. Optical trapping and manipulation of single cells using infrared laser beams. *Nature*, 330(6150):769–771, 1987.
- [10] S.B. Smith, Y. Cui, and C. Bustamante. Overstretching b-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271(5250):795, 1996.
- [11] M.D. Wang, H. Yin, R. Landick, J. Gelles, and S.M. Block. Stretching DNA with optical tweezers. *Biophysical Journal*, 72(3):1335–1346, 1997.
- [12] S. Uemura, M. Dorywalska, T.H. Lee, H.D. Kim, J.D. Puglisi, and S. Chu. Peptide bond formation destabilizes shine–dalgarno interaction on the ribosome. *Nature*, 446(7134):454–457, 2007.
- [13] J.D. Wen, L. Lancaster, C. Hodges, A.C. Zeri, S.H. Yoshimura, H.F. Noller, C. Bustamante, and I. Tinoco. Following translation by single ribosomes one codon at a time. *Nature*, 452(7187):598–603, 2008.
- [14] X. Qu, J.D. Wen, L. Lancaster, H.F. Noller, C. Bustamante, and I. Tinoco. The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature*, 475(7354):118–121, 2011.

- [15] D. P. Giedroc, C. A. Theimer, and P. L. Nixon. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *Physical Review E*, 298:167–185, 2000.
- [16] Song Cao and Shi-Jie Chen. Biphasic folding kinetics of RNA pseudoknots and telomerase RNA activity. *Journal of Molecular Biology*, 367(3):909–924, 2007.
- [17] T. M. Hansen, S. N. S. Reihani, L. B. Oddershede, and M. A. Sørensen. Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proceedings of the National Academy of Sciences USA*, 104(14):5830–5835, 2007.
- [18] G. Chen, J.D. Wen, and I. Tinoco Jr. Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *Rna*, 13(12):2175–2188, 2007.
- [19] D.P. Giedroc and P.V. Cornish. Frameshifting RNA pseudoknots: structure and mechanism. *Virus research*, 139(2):193–208, 2009.
- [20] J Tholstrup, L B Oddershede, and M A Sorensen. mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Research*, (2):1–11, 2011.
- [21] G. Chen, K.Y. Chang, M.Y. Chou, C. Bustamante, and I. Tinoco, Jr. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proceedings of the National Academy of Sciences USA*, 106(31):12706–12711, 2009.
- [22] E. P. Plant, K. L. M. Jacobs, J. W. Harger, A. Meskaukas, J. L. Jacobs, J. L. Baxter, A.N. Petrov, and J.D. Dinman. The 9-Å solution: How mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA - a publication of the RNA society*, 9(2):168–174, 2003.
- [23] Biao Liu, David H. Mathews, and Douglas H. Turner. RNA pseudoknots: folding and finding. *F1000 biology reports*, 2(January):8, 2010.
- [24] Michela Taufer, Abel Licon, Roberto Araiza, David Mireles, F H D Van Batenburg, Alexander P Gulyaev, and Ming-Ying Leung. Pseudobase++: an extension of pseudobase for easy searching, formatting and visualization of pseudoknots. *Nucleic Acids Research*, 37(Database issue):D127–D135, 2009.
- [25] L. Green, C.H. Kim, C. Bustamante, and I. Tinoco Jr. Characterization of the mechanical unfolding of RNA pseudoknots. *Journal of molecular biology*, 375(2):511–528, 2008.
- [26] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):231 ff, 1999.
- [27] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, Apr 1997.
- [28] Nanotracker product brochure. Accessed at www.jpk.com/jpk-nanotracker-1.download.035194e1bf93318bb0239ea5d.pdf found at www.jpk.com/jpk-nanotracker-tm.387.html on Feb. 21, 2011.
- [29] K. Berg-Sørensen and H. Flyvbjerg. Power spectrum analysis for optical tweezers. *Review of Scientific Instruments*, 75(3):594–612, 2004.
- [30] J.K. Dreyer, K. Berg-Sørensen, and L. Oddershede. Improved axial position detection in optical tweezers measurements. *Applied optics*, 43(10):1991–1995, 2004.

- [31] S. C. Kuo and M. P. Sheetz. Force of single kinesin molecules measured with optical tweezers. *Science*, 260:232–234, 1993.
- [32] A. C. Richardson, S. N. S. Reihani, and L. B. Oddershede. Non-harmonic potential of a single beam optical trap. *Optics Express*, 16(20):15709–15717, 2008.
- [33] A. Ashkin. Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophys. J.*, 61:569–582, 1992.
- [34] F. Gittes and C. H. Schmidt. Signals and noise in micromechanical measurements. In Michael P. Sheetz, editor, *Methods in Cell Biology, Volume 55, Laser Tweezers in Cell Biology*, pages 129–156. Academic Press, 1998.
- [35] K. Svoboda and S.M. Block. Biological applications of optical forces. *Annual review of biophysics and biomolecular structure*, 23(1):247–285, 1994.
- [36] K.C. Neuman and S.M. Block. Optical trapping. *Review of Scientific Instruments*, 75:2787, 2004.
- [37] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes – The Art of Scientific Computing*. Cambridge: Cambridge University Press, third edition, 2007.
- [38] John Bechhoefer and Scott Wilson. Faster, cheaper, safer optical tweezers for the undergraduate laboratory. *American Journal of Physics*, 70(4):393, 2002.
- [39] K. Berg-Sørensen, L. Oddershede, E.L. Florin, and H. Flyvbjerg. Unintended filtering in a typical photodiode detection system for optical tweezers. *Journal of Applied Physics*, 93:3167, 2003.
- [40] E.J.G. Peterman, F. Gittes, and C.F. Schmidt. Laser-induced heating in optical traps. *Biophysical journal*, 84(2):1308–1316, 2003.
- [41] L. Corson, W. Drost-Hansen, and F. J. Millero. Viscosity of water at various temperatures. *Journal of Physical Chemistry*, 73(1):34–39, 1969.
- [42] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of λ -phage DNA. *Science*, 265(5178):1599–1600, 1994.
- [43] L. Oddershede, February 2012. Oral communication.
- [44] P. Gross, N. Laurens, L.B. Oddershede, U. Bockelmann, E.J.G. Peterman, and G.J.L. Wuite. Quantifying how DNA stretches, melts and changes twist under tension. *Nature Physics*, 7(9):731–736, 2011.
- [45] J. van Mameren, P. Gross, G. Farge, P. Hooijman, M. Modesti, M. Falkenberg, G. J. L. Wuite, and E. J. G. Peterman. Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proceedings of the National Academy of Sciences*, 106(43):18231–18236, 2009.
- [46] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr., and C. Bustamante. Verification of the crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437:231–234, 2005.
- [47] I. Tinoco, P.T.X. Li, and C. Bustamante. Determination of thermodynamics and kinetics of RNA reactions by force. *Quarterly reviews of biophysics*, 39(4):325–360, 2006.
- [48] G. Hummer and A. Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical journal*, 85(1):5–15, 2003.

- [49] O.K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical review letters*, 96(10):108101, 2006.
- [50] E. Evans, K. Halvorsen, K. Kinoshita, and W. P. Wong. A new approach to analysis of single molecule force experiments. In *Handbook of Single-Molecule Biophysics*.
- [51] I. Tinoco Jr. Force as a useful variable in reactions: unfolding RNA. *Annu. Rev. Biophys. Biomol. Struct.*, 33:363–385, 2004.
- [52] Olga K Dudko, Gerhard Hummer, and Attila Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41):15755–15760, 2008.
- [53] G. E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90(5):1481–1487, 1998.
- [54] C. Jarzynski. Nonequilibrium work relations: foundations and applications. *The European Physical Journal B-Condensed Matter and Complex Systems*, 64(3):331–340, 2008.
- [55] Thomas Heimburg, November 2010. Course notes on Non-Equilibrium Thermodynamics and Evolution, taught at the Niels Bohr Institute.
- [56] C.H. Bennett. Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics*, 22(2):245–268, 1976.
- [57] A. Mossa, J. M. Huguet, and F. Ritort. Investigating the thermodynamics of small biosystems with optical tweezers. *Physica E: Low-dimensional Systems and Nanostructures*, 42(3):666–671, 2010.
- [58] A. Mossa, S. de Lorenzo, J. M. Huguet, and F. Ritort. Measurement of work in single-molecule pulling experiments. *The Journal of chemical physics*, 130:234116, 2009.
- [59] S. Naphthine, J. Liphardt, A. Bloys, S. Routledge, and I. Brierley. The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *Journal of Molecular Biology*, 288(3):305–320, 1999.
- [60] J. Tholstrup and M. Sørensen, 2012. Unpublished data.
- [61] J. Reeder, P. Steffen, and R. Giegerich. pknotsrg: RNA pseudoknot folding including near-optimal structures and sliding window. *Nucleic Acids Research*, 35(Issue Suppl 2):W320–W324, 2007.
- [62] Pseudoviewer. An online programme that generates visualizations of predicted pseudoknot structures found at <http://pseudoviewer.inha.ac.kr/>. It is described in [82].
- [63] J. Sperschneider, A. Datta, and M. J. Wise. Heuristic RNA pseudoknot prediction including intramolecular kissing hairpins. *RNA*, 17(1):27–38, 2011.
- [64] Hotknots software. An online programme that predicts pseudoknot structures. Found at <http://dotknot.csse.uwa.edu.au/>. Described originally in [83] and improved by parameters estimated in [84].
- [65] Pan T X Li, Carlos Bustamante, and Ignacio Tinoco. Real-time control of the energy landscape by force directs the folding of RNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17):7039–44, 2007.
- [66] M. Sørensen, April 2012. Oral communication.

- [67] J. Tholstrup. *PhD thesis, in preparation*. PhD thesis, University of Copenhagen, 2012. Kindly provided ahead of publication.
- [68] I. Brierley, N.J. Rolley, A.J. Jenner, and S.C. Inglis. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *Journal of molecular biology*, 220(4):889–902, 1991.
- [69] T. M. Hansen. *Studies of translation, on the ensemble level using in vivo techniques and on the single molecule level using optical tweezers*. PhD thesis, University of Copenhagen, 2007. Accessed at www.nbi.dk/tweezer/ during 2011 and 2012.
- [70] P. R. Bevington and D. K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. Springer, second edition, 2002.
- [71] Petrutza Caragea. The f-test as a comparison of full and reduced models, 2007. Course notes from the Iowa State University course STAT 401, Statistical methods for Research Workers, accessed April 2012 at www.public.iastate.edu/~pcaragea/S401F07/Handouts/Full_vs_Reduced.pdf.
- [72] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference - A Practical Information-Theoretic Approach*. McGraw Hill, third edition, 2003.
- [73] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, 296:1832–1835, 2002.
- [74] O.K. Dudko, A.E. Filippov, J. Klafter, and M. Urbakh. Beyond the conventional description of dynamic force spectroscopy of adhesion bonds. *Proceedings of the National Academy of Sciences*, 100(20):11378, 2003.
- [75] P. Mangeol, T. Bizebard, C. Chiaruttini, M. Dreyfus, M. Springer, and U. Bockelmann. Probing ribosomal protein–RNA interactions with an external force. *Proceedings of the National Academy of Sciences*, 108(45):18272–18276, 2011.
- [76] P. Mangeol, D. Côte, T. Bizebard, O. Legrand, and U. Bockelmann. Probing DNA and RNA single molecules with a double optical tweezer. *The European Physical Journal E: Soft Matter and Biological Physics*, 19(3):311–317, 2006.
- [77] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415, 2003.
- [78] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciences*, 103(16):6190–6195, 2006.
- [79] P. T. X. Li, D. Collin, S. B. Smith, C. Bustamante, and I. Tinoco Jr. Probing the mechanical folding kinetics of tar RNA by hopping, force-jump, and force-ramp methods. *Biophysical journal*, 90(1):250–260, 2006.
- [80] S. Cao, D.P. Giedroc, and S.J. Chen. Predicting loop–helix tertiary structural contacts in RNA pseudoknots. *RNA*, 16(3):538–552, 2010.
- [81] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Sciences*, 98(7):3658, 2001.

-
- [82] Y. Byun and K. Han. Pseudoviewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics*, 25(11):1435, 2009.
- [83] J. Ren, B. Rastegari, A. Condon, and H.H. Hoos. Hotknots: heuristic prediction of RNA secondary structures including pseudoknots. *Rna*, 11(10):1494–1504, 2005.
- [84] M.S. Andronescu, C. Pop, and A.E. Condon. Improved free energy parameters for RNA pseudoknotted secondary structure prediction. *RNA*, 16(1):26–42, 2010.
- [85] D.S. Lemons and P. Langevin. *An introduction to stochastic processes in physics: containing "On the theory of Brownian motion" by Paul Langevin, translated by Anthony Gythiel*. Johns Hopkins University Press, 2002.

A Power spectrum and other calibration treatment details

This appendix will show how the expectation value of the power spectrum is derived and how the best routine was selected for fitting the theoretical power spectrum to the experimental power spectrum. First of all, however, the Fourier transformation which is used to calculate the power spectrum from data will be described.

A.1 Fourier transforms and power spectrum construction

Performing a Fourier transformation can be seen as dissolving a signal measured over time into a sum of cosine and sine functions with different frequencies. Thus Fourier transforming takes a measure of a signal in the time domain and makes it into a measure in the frequency domain. For a signal $x(t)$ that is continuous in time, the Fourier transform is:

$$\tilde{x}(f) = \int_{-\infty}^{\infty} x(t) e^{2i\pi ft} dt$$

while the inverse transform is:

$$x(t) = \int_{-\infty}^{\infty} \tilde{x}(f) e^{-2i\pi ft} dt$$

The transformation can be normalized in several slightly different ways; this one follows the recommendation in [37]. Note that both time and frequency are defined from $-\infty$ to ∞ . For a series of N discrete measurements x_n made at times $t = n\delta t$ with n running from one to N , the Fourier transformation becomes

$$\tilde{x}(f_k) = \sum_{n=1}^N x_n e^{-2\pi i f_k t} \delta t$$

with frequency f_k defined for $-\frac{N}{2} \leq k \leq \frac{N}{2}$.

As mentioned in the main text, Gittes and Schmidt [34] note that the Fourier transform contains exactly the same amount of information as the original function, it just presents it differently. The power spectrum is derived from it with a loss of half of the information but a gain in usefulness. The one-sided power spectral density for a signal that is continuous in time has already been shown in the main text but is repeated here for clarity:

$$S_x(f) = |\tilde{x}(f)|^2 + |\tilde{x}(-f)|^2 = 2 |\tilde{x}(f)|^2$$

Note that $S_x(f)$ is defined only for $f \geq 0$ (which seems to make physical sense when we are dealing with a time series of distance measurements) and the last equality holds only for functions $x(t)$ that are real. This is the power spectrum definition that is used for calibrating optical traps in [34] as well as [35] and [36] (the two-sided power spectrum can be used as well as long as the treatment is internally consistent). See [34] for the definition of the power spectrum for discrete data as is needed for actual experimental calibration.

A.2 Solving the Langevin Equation

Now we go on to derive the expectation value for the power spectrum, $S_x(f)$. This is done by solving the Langevin equation of Brownian motion for a particle in a harmonic potential [34]:

$$F_{tot} = F_{therm} - \gamma v(t) - \kappa x(t). \quad (\text{A.1})$$

This equation says that the total force on the particle is the sum of the thermal forces acting on it, a drag force caused by its movement through the medium, and a force from the harmonic potential. The variable γ as above is the drag coefficient and F_{therm} is the collective effect of the thermal forces from the surrounding medium. We assume that these thermal forces are exactly balanced by the drag forces and the trap potential at any given moment so that F_{tot} can be set equal to zero. This assumption is not entirely trivial, as discussed in [85], but it works quite well. We then have:

$$F_{therm} = \gamma \frac{dx}{dt} + \kappa x(t) \quad (\text{A.2})$$

To solve this equation for the power spectrum, we need to know that $|\tilde{F}_{therm}|^2 = 2\gamma k_B T$ for all frequencies. The derivation of this entity is also not trivial; one demonstration can be found in the appendix of [38]. It relies on the fluctuation-dissipation theorem and the Wiener-Khintchine theorem.

First we find the time derivative of the position measurement series in terms of its Fourier transform:

$$x(t) = \int_{-\infty}^{\infty} \tilde{x}(f) e^{-2\pi f t} dt \implies \frac{dx}{dt} = \int_{-\infty}^{\infty} \tilde{x}(f) (-2\pi i f) e^{-2\pi i f t} dt.$$

Knowing $|F_{therm}|^2$ we can then write the Langevin equation in terms of its Fourier components:

$$\begin{aligned} \int_{-\infty}^{\infty} \tilde{F}_{therm}(f) e^{-2i\pi f t} dt &= \kappa \int_{-\infty}^{\infty} \tilde{x}(f) e^{-2i\pi f t} dt + \gamma \int_{-\infty}^{\infty} \tilde{x}(f) (-2\pi i f) e^{-2\pi i f t} dt \\ &= \int_{-\infty}^{\infty} (\kappa - 2\gamma\pi i f) \tilde{x}(f) e^{-2\pi i f t} dt. \end{aligned}$$

We remove the summation and find that the Fourier transform of the Langevin equation can be written as:

$$\tilde{F}_{therm}(f) = (\kappa - 2\gamma\pi i f) \tilde{x}(f).$$

Now we can square the modulus of the equation above and plug in our knowledge of $|\tilde{F}_{therm}|^2$ from the fluctuation-dissipation theorem:

$$\begin{aligned} |\tilde{F}_{therm}(f)|^2 &= 2\gamma k_B T = |[(\kappa - 2\gamma\pi i f) \tilde{x}(f)]|^2 \\ &= (\kappa^2 + 4\gamma^2 \pi^2 f^2) |\tilde{x}(f)|^2 \end{aligned}$$

Finally we find:

$$\begin{aligned} S_x(f) &= 2 |\tilde{x}(f)|^2 \\ &= \frac{4\gamma k_B T}{\kappa^2 + 4\gamma^2 \pi^2 f^2} \\ &= \frac{k_B T}{\gamma \pi^2 (f_c^2 + f^2)} \end{aligned}$$

where in the last line we have substituted in $f_c = \frac{\kappa}{2\pi\gamma}$.

As described in the main text we can use this theoretical expectation value for the power spectrum to find the calibration factor β from volts to distance. Ideally (but not with the Nanotracker noise) we would also be able to find the trap stiffness κ and the calibration factor $\alpha = \beta\kappa$ from volts to force. To find β we need to use theoretical predictions of the temperature, T , and viscosity, η , of the medium surrounding the bead, as well as the radius, r , of the bead:

$$\beta = \text{sqr}t\left(\frac{k_B T}{6\pi^2 \eta r S_{QPD}}\right).$$

where S_{QPD} is the experimentally measured power spectrum which has units of

A.3 Selection of the best power spectrum fitting routine

Figure A.1 shows fits to the experimental power spectrum using three different fitting routines together with the test plots of force-extension curves to evaluate which routine was most accurate:

- Figure A.1 (a) shows calibration performed by a weighted least squares fit to the power spectrum region of high frequency only, $f \gg f_c$, where $S_x(f) \approx \frac{k_B T}{\gamma \pi^2 f^2}$.
- Figure A.1 (b) shows the corresponding test plot of a force-extension curve for the pulling of an inelastic polystyrene tether, which should be a vertical line that increases in force but not extension.
- Figure A.1 (c) shows calibration performed by a weighted least squares fit of the low and high frequency data regions to the full equation for the power spectrum. This fit excludes only the noisiest data region. Like the fit in Figure A.1 (a), it was performed in Gnuplot.
- Figure A.1 (e) shows *unweighted* fits of the low and high frequency portions of the experimental power spectrum to the theoretical one. This fit was carried out using a fitting routine implemented in IgorPro by Jesper Tholstrup.
- Figures A.1 (d) and (f) show the test force-extension curve plots of the pulling of the same inelastic tether as in Figure A.1 (b). It is seen that the force-extension curve is closest to vertical in Figure A.1 (f) corresponding to the unweighted fit.

A.3.1 Quantification of Nanotracker noise

For the experiment shown in Figure 2.3.2, we can calculate the “extra” noise caused by the system beyond the expected thermal motions in the trap. Using Stokes and Power Spectrum calibration we find the conversion factors α and β for the traps which yield κ and thus the predicted standard deviation, $\sigma_{xPredict}^2$, for Trap 1 and Trap 2 from thermal motion:

$$\begin{aligned} \kappa_1 = 133\text{pN/um} & \implies \sigma_{x1Predict}^2 = \frac{k_B T}{\kappa_1} = 31\text{nm}^2 \\ \kappa_2 = 229\text{pN/um} & \implies \sigma_{x2Predict}^2 = \frac{k_B T}{\kappa_2} = 18\text{nm}^2 \end{aligned}$$

and

$$\sigma_{x1Predict} = 5.6\text{nm}, \quad \sigma_{x2Predict} = 4.2\text{nm}$$

The actual σ_x^2 is found from the measured σ_{QPD}^2 , using β found using the power spectrum calibration:

$$\sigma_{x1}^2 = \beta_1^2 \sigma_{QPD1}^2 = (80\text{nm/V})^2 \cdot (0.21\text{V})^2 = 282\text{nm}^2$$

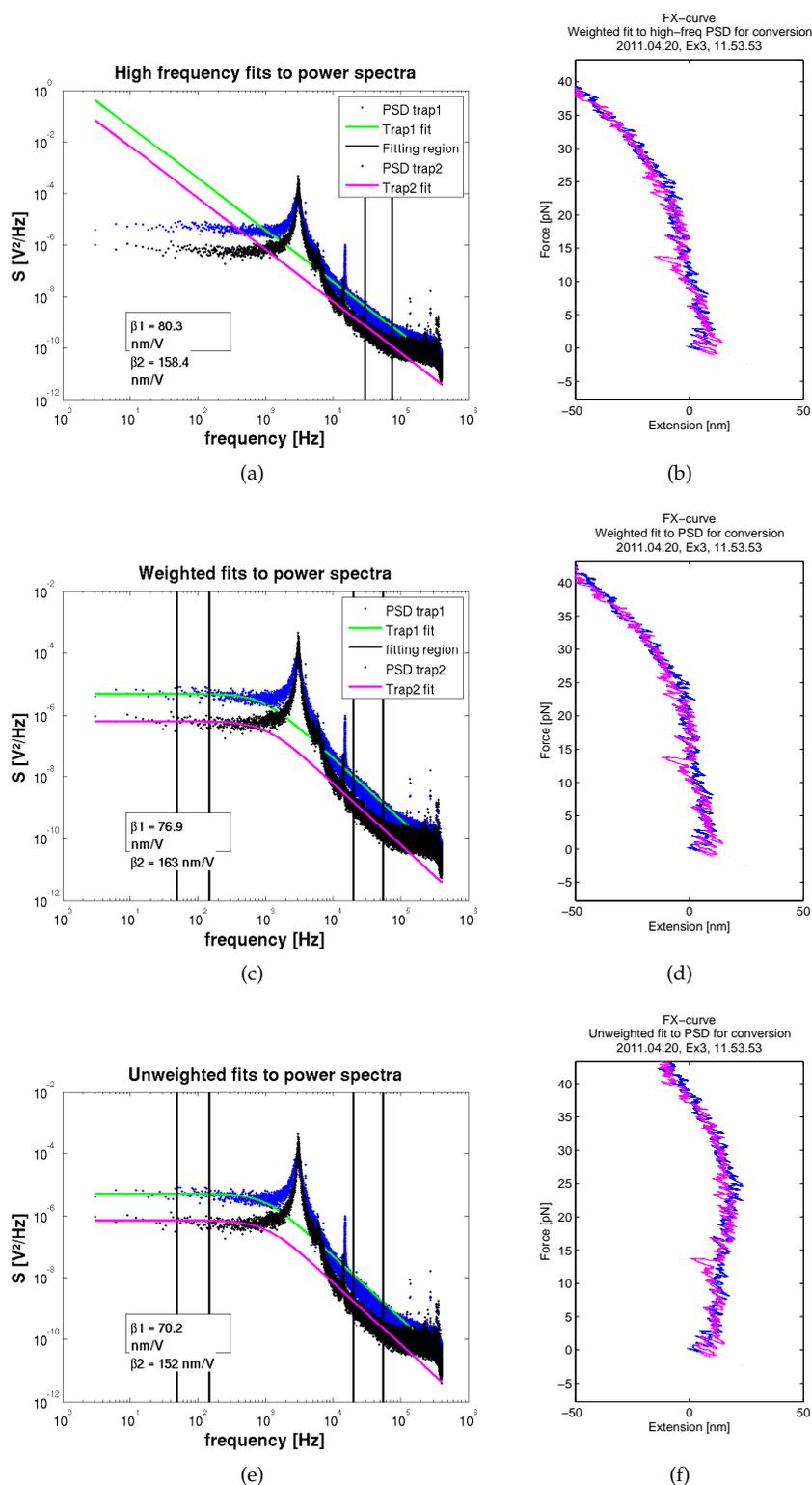


FIGURE A.1: Left: Fits to the power spectra of the two NanoTracker traps. Right: Tests of the calibration factors through plots of the force-extension curve for an inelastic polystyrene tether joining two beads. The three curves on the right should show a vertical line of zero extension for all forces. The fitting regions for the power spectra on the left are demarcated by black lines and the values found for β^2 for each trap is shown on the graphs.

$$\sigma_{x2}^2 = \beta_1^2 \sigma_{QPD1}^2 = (155\text{nm/V})^2 \cdot (0.22\text{V})^2 = 1162\text{nm}^2$$

and

$$\sigma_{x1} = 17\text{nm}, \quad \sigma_{x2} = 34\text{nm}$$

Thus the noise from the system increases the standard deviation of the bead positions in Trap 1, σ_{x1} , by a factor of three compared to that predicted from thermal motion alone, $\sigma_{x1Predict}$, while the standard deviation of the bead motion in trap 2, σ_{x2} , is increased by a factor of eight compared to $\sigma_{x2Predict}$. The magnitude of the standard deviation due to noise is thus quite significant. The total noise experienced by the molecule attached to both beads is the combination of noise in both traps and has a standard deviation of $\sqrt{(17^2 + 34^2)} = 38$ nm. This is on the order of the length of the unfolding of the pseudoknots.

B Frameshift and stall efficiency investigation

Frameshifting is the defining characteristic of mRNA pseudoknots that this thesis attempts to correlate with their mechanical properties. It is therefore important to understand how the frameshift efficiencies of the pseudoknots are quantified even though this analysis was not performed by the author.

Images from a gel analysis done by Michael Sørensen to evaluate the frameshift and stalling efficiencies of PK 6/11 are shown in Figure B.1. This figure illustrates the general procedure used by Jesper Tholstrup and Michael Sørensen to investigate the frameshifting of the pseudoknots studied here. Similar images for PK 11/6 are shown in Figure S5 in Tholstrup et al. [20]. For clearer understanding of the figure, this is the conceptual outline of the frameshifting assays performed by Jesper and Michael: The pseudoknots were expressed *in vivo* in *E. coli*, meaning that a gene coding for the pseudoknot and a frameshifted or non-frameshifted protein product was transcribed into mRNA by the *E. coli*'s cellular machinery and subsequently translated into protein products by the *E. coli*'s ribosomes. When the ribosome was not frameshifted, a relatively short protein was produced before the ribosome encountered a stop codon in the original reading frame. If frameshifting occurred, either the ribosome might get stuck within the pseudoknot, producing a very short protein product, or it would proceed through the opened pseudoknot, producing a large protein product before reaching a stop codon in the -1 reading frame. The protein products were labelled radioactively so that they could be extracted from the cells. Then they were separated on SDS-gels by size and pH, i.e., in two dimensions in contrast to the conventional frameshift measurement assay, which separates the products in one dimension only, namely by size. This procedure allows determination of the relative amounts of each product type. See Figure 2A in [20] for a clear theoretical illustration of the locations of the expected products on the gel and see the same source for a complete description of how the frameshift assays were carried out.

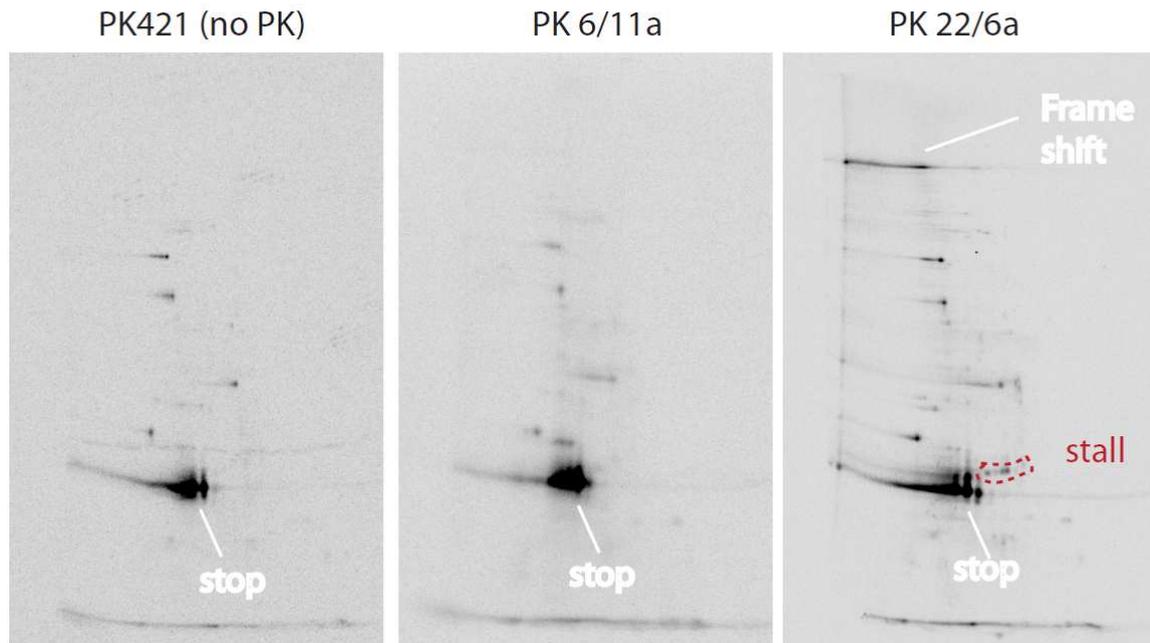


FIGURE B.1: Comparison of 2D SDS-PAGE frameshift assays for protein products from an mRNA with no pseudoknot (PK421, not otherwise discussed here), PK 6/11, and a pseudoknot which frameshifts very much (PK 22/6a, also not otherwise discussed here). The different types of protein product are separated on the SDS gel according to size (top: larger, bottom: smaller) and isoelectric point (increasing pH from left to right). “Stop” indicates non-frameshifted ribosomes which encountered a stop codon downstream from the pseudoknot. “Stall” marks products from stalled frameshifted ribosomes and “Frameshift” marks full protein products from frameshifted ribosomes. Also seen on the gel are short polypeptide fragments (left part of bottom blobs) and markers of known peptide length (middle dots). Assays and illustration kindly made by Michael Sørensen.

C Biochemical methods

As described briefly in Chapter 4.3, for this thesis, the author participated in synthesizing the downstream DNA handles and the mRNA pseudoknot strand, annealing handles to RNA and preparing RNA-handle constructs with beads for pulling. Below the routine for preparing RNA-handle constructs will be described in detail. As mentioned, all of the protocols were suggested by and carried out with the guidance of Jesper Tholstrup, who also synthesized plasmids, finished upstream handles and DNA oligos used for downstream handle synthesis which were used in the processes described below.

C.1 Preparing RNA-handle constructs: downstream handle preparation

Downstream handles were prepared from DNA oligos which had digoxigenin attached to the 5' end of the strand that would become the handle strand.

Double-stranded (ds) DNA oligos had been generated by PCR by Jesper Tholstrup. They were extracted with phenol and digested with λ -exonuclease which breaks down DNA strands from the 5' end. Since digoxigenin protected the handle strand, the exonuclease only removes the opposite single strand, leaving single stranded handle. After digestion, the handles are washed with EDTA and sarcosyl to remove the exonuclease, then again carefully extracted with phenol, washed with salt, measured and frozen for later use.

Detailed protocol starting from purified handle oligos:

- DNA pellet from PCR is diluted with 40 μ m Millipore water
- measure concentration of ds handle oligos in NanoDrop
- save 0.5 μ l ds oligo solution ("pre- λ ") for gel testing
- Add water (first!), 10 μ l 10x λ -exonuclease buffer (Fermentas), and 4 μ l λ -exonuclease (Fermentas) to a total of 100 μ l
- Incubate for 24 minutes at 37. The resulting solution is called " λ -exo"
- Prepare an agarose gel and run marker (GeneRuler), " λ -exo" and "pre- λ " solution to check that the handles have been produced satisfactorily
- The remaining " λ -exo" handle solution is labelled "5' dig" handle solution plus name and date and frozen at -20

C.2 DNA plasmids and transcription of DNA to make RNA

The DNA template for the pseudoknot is an artificial sequence that has been designed, purchased and inserted in a plasmid. A plasmid is a relatively small double stranded circular piece of DNA often found in bacteria. It can be replicated independently of the rest of the DNA in the cell in which it is found and can also be picked up from solution by bacteria which then express the genes found on the plasmids. This plasmid property was used for the *in vivo* frameshifting assays made by Jesper Tholstrup.

Before the DNA template is transcribed into RNA it is amplified by PCR. During this process lots of linear copies are made of the pseudoknot-encoding section of the plasmid that is needed. This is then used to synthesize RNA. The process is as follows:

- We start with solutions of circular plasmid DNA with PK sequence + primers + T7 DNA polymerase
- → PCR is performed to copy DNA
- → Linear DNA template is extracted and purified
- → We add linear DNA template + T7 RNA polymerase
- → RNA is extracted, purified and frozen

C.2.1 Preparing RNA-handle constructs: template amplification

PCR is a simple sequence of heating and cooling reactions that create many copies of a DNA template. The principles of the PCR routine used here are: Heating to 95 denatures the circular plasmid DNA, separating the double strand into single strands. Cooling to 58 allows primer to bind. Reheating to 72 allows Taq-DNA polymerase to copy the desired strand (Taq polymerase was originally purified from *Thermophilus aquaticus* bacteria who live at very high temperatures). The cycle is performed automatically and is continued until the DNA has been copied sufficiently.

Protocol for template amplification by PCR:

- A plasmid solution is prepared for PCR: water, Fermentas buffer (Dream Taq), nucleotides, primer, plasmid and polymerase are added in correct amounts in that order
- PCR is done in an incubator which can be programmed to turn the temperature up and down at specific intervals so that the recommendations from Fermentas for the PCR reaction are followed
- The resulting linear template DNA is run on an agarose gel with a marker to check that the reaction worked as expected
- The template is extracted and washed with the buffers provided by Fermentas
- The concentration of the resulting purified DNA template is measured by NanoDrop.
- The remaining DNA template is labelled and frozen at -20

C.2.2 Preparing RNA-handle constructs: RNA synthesis

RNA is transcribed from the DNA template using T7 RNA polymerase (this is a polymerase originally found in T7 phage). This process requires great care, since RNA is easily destroyed by RNA-ase enzymes that are everywhere. To counteract this, RNA-ase inhibitor, RNA-sin, is added to the reaction mixture.

The protocol for transcription is:

- Water, transcription buffer (Fermentas), nucleotides, template DNA, RNA-sin and T7 DNA polymerase are mixed
- The mixture is incubated for 1.5-2 hours at 37
- The reaction is stopped by cooling to 15
- The resulting RNA mixture is run on an agarose gel together with a marker and with DNA template to check that the reaction has succeeded
- The RNA is extracted with phenol, chloroform, salt (NaCl) and ethanol:
 - * phenol is added (under a fume hood!) to the solution and mixed on a whirl mixer, then the supernatant is removed (that's where the RNA is) and the phenol is properly disposed of

- * chloroform is added (also under a fume hood) and mixed, the supernatant including the RNA is again removed and the chloroform properly disposed of
- * 96 % ethanol is added along with NaCl to 200 μ M in 1 μ l volume
- * freeze at -20 for 30 minutes
- * centrifuge at 0 for 20 minutes
- * remove the solution, taking care not to disturb the pellet of salt and RNA at the bottom of the Eppendorf tube
- * dry under exhaust and mix with TE-buffer
- The resulting RNA is again measured on an agarose gel against a marker (to see if the reaction worked). Some template DNA will also be left behind in solution, but should not be a problem

C.3 Preparing RNA-handle constructs: annealing handles to RNA

RNA is annealed to DNA handles by heating. The process is not always successful and the relative amounts of RNA solution and either handle solution may need to be adjusted, so careful checking using an agarose gel to examine the partial products of the process is necessary. See Figure 4.4 for an example.

Protocol:

- * Mix 10 μ l control samples of RNA, RNA plus biotin handle, and RNA plus digoxigenin handle, and 40 μ l RNA plus both handles in buffer R. Order of mixing: Buffer R, water to dilute if necessary, RNA, handle
- * Use very little RNA, 1 % of the total volume may be enough - or even too much. The amount of handle needed depends on concentration
- * Heat RNA-handle solution to 75 for 5 minutes
- * Cool slowly to 15 over the course of about 1 hour
- * Test the outcome on an agarose gel
- * Store the RNA-handle constructs at -20 or even -70

C.4 Annealing RNA constructs to beads for pulling experiments

The basic protocol for sample preparation for the pulling experiments was:

- * 5 μ l 1x buffer R was mixed with 0.3 μ l 3.05 μ diameter streptavidin coated polystyrene beads (Spherotech, SVP-30-5) and 0.1-0.5 μ l RNA construct
- * Mix by tapping with a finger and leave for 10-25 minutes at room temperature
- * Dilute with 250 μ l 1x buffer R
- * Add 0.5-1.5 μ l 3.05 μ diameter anti-digoxigenin-coated polystyrene beads (Spherotech, DIGP-20-2)

All components were kept at 5 except for the RNA construct, which was kept at -20 . It is important to be aware that anti-digoxigenin degrades over the time of a few months at this temperature. Again, the RNA construct also may degrade over the course of a few months, though the exact time scale is not known.

D Testing bias from super tethers

The data for unfolding and refolding forces and distances rely heavily on data from only a few experiments. This is simply because the molecular tethers caught in some experiments survived pulling without breaking far better and at much higher forces than the others. A few of the very resistant molecular tethers also displayed many more clear pseudoknot openings and closings than the others, and these "super tethers" which survived many pulls with many clear rips or zips consequently dominate the results.

For pseudoknot PK 11/6 molecular tethers were caught in 14 experiments, but only eight displayed clear molecular transitions (rips or zips). Of these eight, two super tethers gave rise to respectively 108 and 50 out of 171 pulls with rips and/or zips recorded. In the case of pseudoknot PK 6/11a, a single experiment gave rise to the majority of the data: 51 out of altogether 64 pulls with rips and zips. The remaining thirteen pulls with rips and/or zips occurred in five different experiments.

To examine how the individual super tether characteristics might have influenced the overall results, the unfolding and refolding forces and rip/zip lengths for the subsets of the data deriving from super tethers and non-super tethers were examined. The average values are shown in tables D.1 and D.2. Unfortunately, for both pseudoknots so few data derived from non-super tethers that it was very difficult to determine whether their distribution differed significantly from the super-tether distributions. Though the average rip and zip forces observed for the non-super tethers are lower than those of the super tethers, this might simply be a consequence of the higher forces applied to the super tethers. The non-super tethers usually broke before such high force could be applied; if not, they did not display any large number of clear rips or zips.

The most striking observation in the super tether data is that the distribution of unfolding forces deriving from the data for the PK 11/6 super tether "2011-05-05-Ex4" is bimodal. This stands out in contrast to the data for the other PK 11/6 super tether, "2011-04-26-Ex2", and the overall unfolding force data, which appear unimodal. If the data for the super tether "2011-04-26-Ex2" were ignored, the overall unfolding force data for PK 11/6 would appear bimodal just as the unfolding force data for PK 6/11, which is dominated by a single super tether. It must thus be kept in mind that more data for PK 6/11 a might have changed the observed distributions somewhat.

To provide a visual comparison to the distributions for the overall unfolding force data, the distributions of unfolding and refolding forces for super tether "2011-04-26-Ex2" and for the remaining tethers (super tether "2011-05-05-Ex4" plus non-super tethers) are shown in Figure D.2. They may be compared to the entire force distributions for PK 11/6 in Figure E.1 and the force distributions for PK 6/11 a in Figure E.2.

It did not make sense to examine the subsets of the data from super tethers and non-super tethers which occurred in one or two steps in detail since the number of two-step transitions was so low. Note however that all high force and most long rips for PK 11/6 were observed in "2011-04-26-Ex2". Thus the force-extension scatter plot for the subset of data excluding "2011-04-26-Ex2" is quite similar to that deriving from PK 6/11 a. See figure D.1

TABLE D.1: Super tether unfolding/refolding forces and rip/zip lengths: PK 11/6

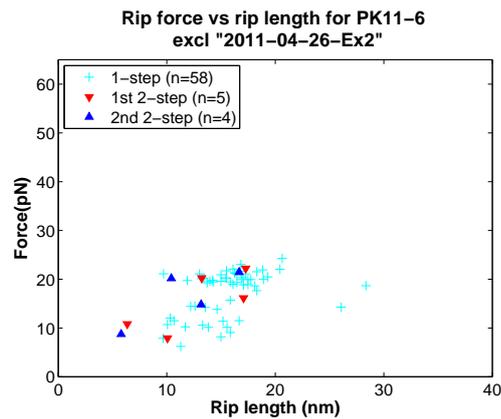
	n_u	n_r	$F_u \pm \text{sem}(\text{pN})$	$F_r \pm \text{sem}(\text{pN})$	$X_u \pm \text{sem}(\text{nm})$	$X_r \pm \text{sem}(\text{nm})$
All 1st rips	169	113	19.8±0.5	13.5±0.3	15.4±0.4	12.4±0.3
No super tethers	13	8	13.9±1.3	9.7±1.7	16±2	13±2
"2011-04-26-Ex2"	106	63	21.3±0.7	13.6±0.4	15.4±0.4	11.3±0.4
"2011-05-05-Ex4"	50	42	20.3±1.1 10.6±2.5	14.0±0.4	15.3±0.5	14.0±0.4
"2011-05-05-Ex4" and non-super tethers	63	50	20.4±1.3 11.5±2.8	13.3±0.5	15.4±0.5	13.8±0.5

TABLE D.2: Super tether unfolding/refolding forces and rip/zip lengths: PK 6/11 a

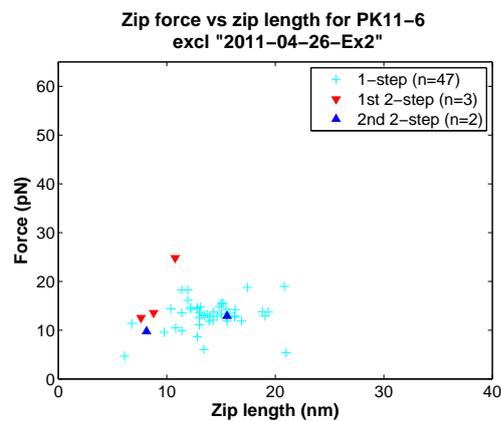
	n_u	n_r	$F_u \pm \text{std}(\text{pN})$	$F_r \pm \text{sem}(\text{pN})$	$X_u \pm \text{sem}(\text{nm})$	$X_r \pm \text{sem}(\text{nm})$
All 1st rips	64	51	16.8±0.5 10.1±2.8	13.8±0.5	12.7±0.4	12.6±0.6
No super tethers	14	7	15.5±2.9 6.6±0.8	9±1	14±1	11±2
"2011-11-03-Ex6"	50	44	16.9±1.1 11.0±2.2	14.6±0.6	12.5±0.4	12.7±0.6

The reason the super tethers did not break is not known.

In the case of "2011-05-05-Ex4", a PK 11/6 super tether, the stiffness of trap 1 was significantly lower than usual at about 106 pN/ μm compared to an average of about 153 pN/ μm (there was large variation from experiment to experiment, however; see description of trap stiffness in Methods and Materials). It could be that the lower trap stiffness, which would reduce the trap-induced vibrations, strained the molecule less than usual. Certainly the lower effective trap stiffness experienced by the molecule, 75 pN/ μm versus an average of 91 ± 13 pN/ μm , would lower the expected force at which the molecule unfolds. For the other PK 11/6 super tether, "2011-04-26-Ex2", the trap stiffness experienced by the molecule, 93 pN/ μm , was very close to the average. The same was true for the PK 6/11 super tether, "2011-11-03-Ex6"; the effective trap stiffness was 107 pN/ μm versus an average of 108 ± 14 pN/ μm .



(a)



(b)

FIGURE D.1: PK 11/6 excluding data for "2011-04-26-Ex2": Scatter plots of a) unfolding force versus rip length and b) refolding force versus zip length. Black dots represent rips/zips classified as occurring in one step, red triangles represent the first step of two-step transitions and blue triangles represent the completing step of two-step transitions.

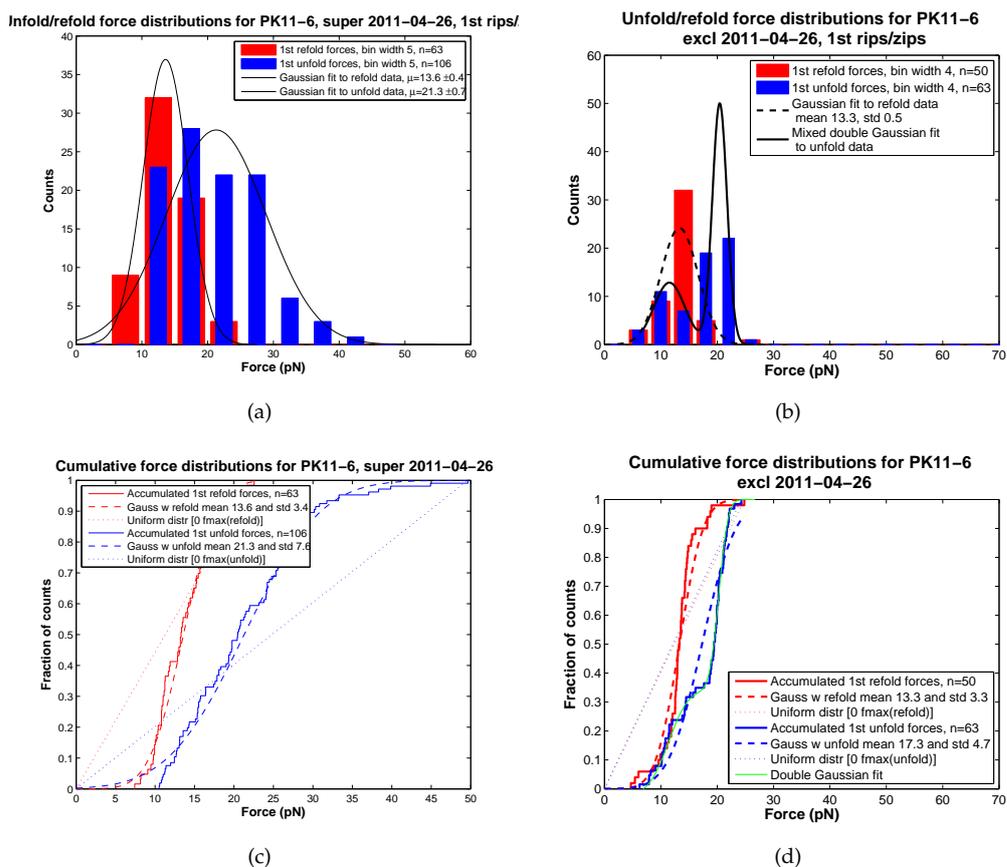


FIGURE D.2: Super tether influence on unfolding force distributions for PK 11/6: Histogram (a) and cumulative functions (c) for "2011-04-26-Ex2" as well as histogram (b) and cumulative functions (d) for the remaining data for the pseudoknot. The latter are dominated by data from another super tether, "2011-05-05-Ex4". Compare to figure ?? a) and c).

E Additional data for separating Rip and Zip subdistributions

E.1 Force distributions of rips and zips

The distributions shown in Figures E.1 and E.2 are not used directly to separate rip subdistributions but do add some insight into the distributions that are also seen in two dimensions in the scatter plots in Figures 5.8 and 5.9.

E.2 Results of statistical tests

TABLE E.1: Results of model comparison testing for PK 11/6 one-step rips and zips. 1G, 2G, 3G, 4G stand for single, double, tripple and quadruple Gaussian. The method numbers refer to the test methods listed above. Gray highlights the criteria that were used most in final evaluation.

Method	rips		zips	
	lengths	forces	lengths	forces
1 Simple χ^2 test ^a	1G rejected	1G rejected	1G <i>not</i> rejected	1G rejected
2 Best R^2 value ^b	2G best	1G best	1G best	1G best
3 F-test of extra model term	3G best	4G best	2G best	1G best
4 AIC test	3G best	4G best	2G best	2G best
5 Multivariate Gaussian Mixture fit	2G/3G (unstable)		usually 2G (unstable)	
6 Visual inspection (scatter plots)	2G or 3G		1G	
Conclusion	2G or 3G		1G	

^a Rejection evaluated at 5 % level.

^b The best R^2 value was the one closest to 1. Most of the fits had R^2 values within a few percent of unity.

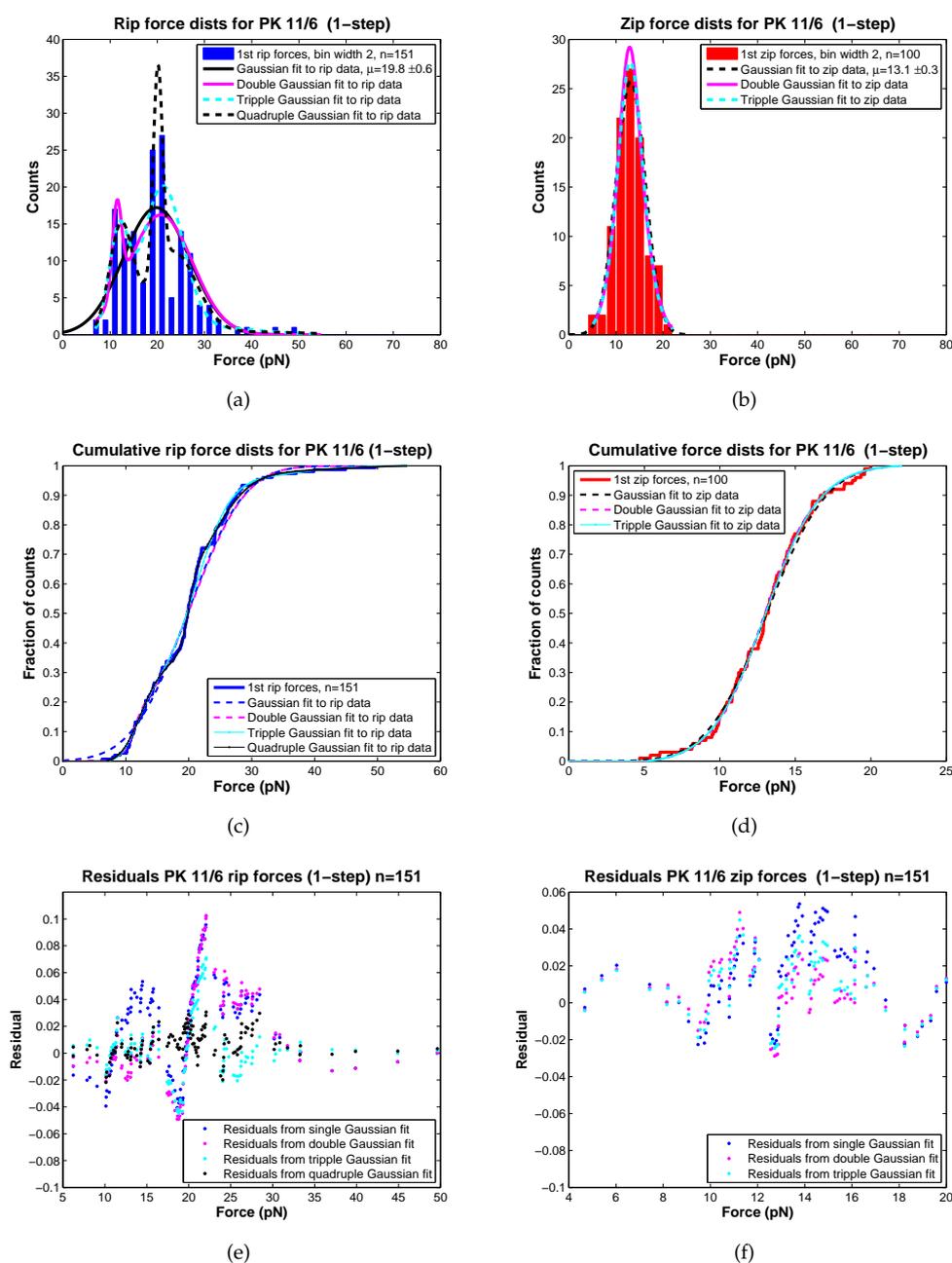


FIGURE E.1: Rip (left) and zip (right) force distributions for PK 11/6: (a) Histogram of first one-step elongation rip forces fit by single, double, and tripple Gaussians. (b) Histogram of first one-step relaxation zip forces with single, double and tripple Gaussian fits. (c) and (d) Cumulative distributions of data and fits (note x-scales are adjusted so that maximum detail of the distribution can be discerned). (e) and (f) Residuals between fitted curves and experimental cumulative distribution.

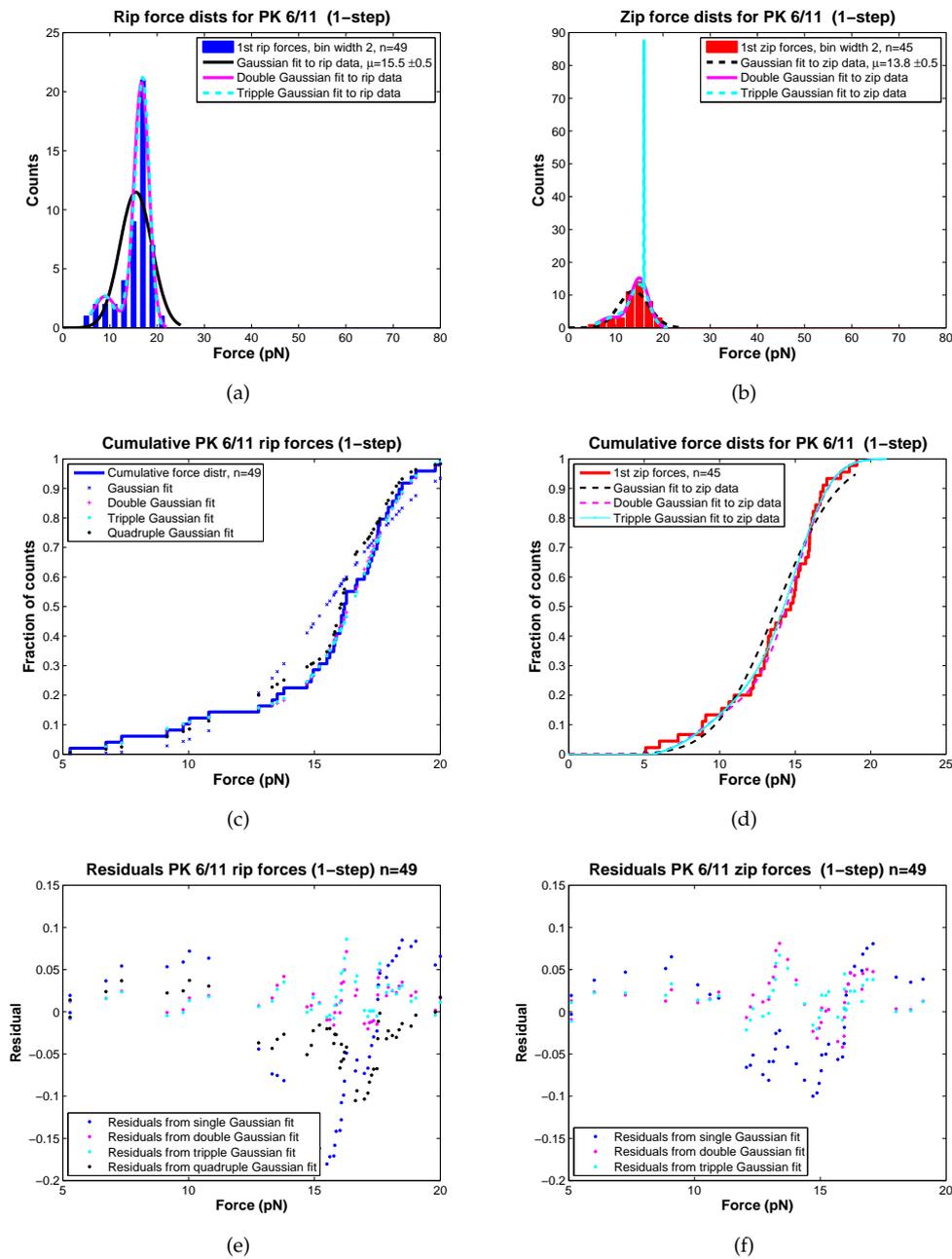


FIGURE E.2: Rip (left) and zip (right) force distributions for PK 6/11: (a) Histogram of first one-step elongation rip forces fit by single, double, and tripple Gaussians. (b) Histogram of first one-step relaxation zip forces with single, double and tripple Gaussian fits. (c) and (d) Cumulative distributions of data and fits (note x-scales are adjusted so that maximum detail of the distribution can be discerned). (e) and (f) Residuals between fitted curves and experimental cumulative distribution.

TABLE E.2: Results of model comparison testing for PK 6/11 one-step rips and zips. 1G, 2G, 3G, 4G stand for single, double, tripple and quadruple Gaussian. The method numbers refer to the test methods listed above. Gray highlights the criteria that were used most in final evaluation.

Method	rips		zips	
	lengths	forces	lengths	forces
1 Simple χ^2 test ^a	1G <i>not</i> rejected	1G rejected	1G <i>not</i> rejected	1G <i>not</i> rejected
2 Best R^2 value ^b	1G best	2G best	1G best	2G best
3 F-test of extra model term	2G best	2G best	2G best	1G best
4 AIC test	2G best	2G best	2G best	3G best
5 Multivariate Gaussian Mixture fit	usually 2G (unstable)		usually 3G (unstable)	
6 Visual inspection (scatter plots)	2G		1G	
Conclusion	2G		1G	

^a Rejection evaluated at 5 % level.

^b The best R^2 value was the one closest to 1. Most of the fits had R^2 values within a few percent of unity.

F Additional data on work and kinetics

F.1 Work distributions - single distribution for all work data treated with simple approach

The simple approach work energy analysis and Gibbs Free Energy calculation was first done for the full set of one-step rips and zips observed for each pseudoknot, i.e., *not* divided according to the subdistributions found in Chapter 5.4.2. Subsequently, the analysis was extended to individual subdistributions

Using the simple area-under-rip/zip method, distributions including *all* unfolding and refolding work values in $P(W_F)$ and $P(-W_R)$ were obtained which were fairly well approximated by Gaussian distributions as may be seen in figures F.1 and 7.3. For $P(W_F)$, the Gaussian fit was improved by excluding the top two work value data points for PK 11/6, so this fit is used for the remaining analysis (compare the full black curve to the dash-dotted in Figure F.1 (a)). These two high outlier data points probably represent the two one-step data points (light blue crosses) furthest to the top and right in the scatter plot in Figure 5.8 (a). Since most of the other data points at this high force and rip length in the scatter plot are part of two-step transitions, they are not included in this version of the work analysis. The simple approach used here therefore quite possibly "misses" part of the dataset in the work analysis.

It was realized that it made more sense to use rip subdistributions for the analysis of the work for PK 11/6 since a single distribution of transitions was effectively rejected for this pseudoknot sequence. The original figure is still shown below.

For both pseudoknots the Gibbs free energy of the equilibrium full transition was found using the Crooks Fluctuation Theorem (CFT) as the intersection of the Gaussian approximations to the full opening and closing transition work distributions (see Chapter 3.4.3). From ΔG for the full transition was subtracted an average of $\langle W_{\text{stretch}}(f) \rangle$ and $\langle W_{\text{stretch}}(r) \rangle$ (for the simple area method, the average is weighted by the number of data points for the forward and the reverse transition). How W_{stretch} is found is described briefly below and in detail in Chapter 3.4.3. An estimate is thus found for ΔG_{PK} , the Gibbs free energy change for opening/closing the pseudoknot without force.

Using the same work distributions, the Jarzinsky Equality (JE) method of finding the equilibrium ΔG of the full transition was also attempted. This was done for both the unfolding and the refolding work distributions, yielding starkly different estimates of the equilibrium ΔG , as might be expected due to the low number of data points and the vulnerability of the Jarzinsky method to low value outliers. Following Collin et al. [46] the average of the two Jarzinsky estimates for ΔG was then found for the full transition. It was for both pseudoknots surprisingly close to the average found using the Crooks and averaging methods described above.

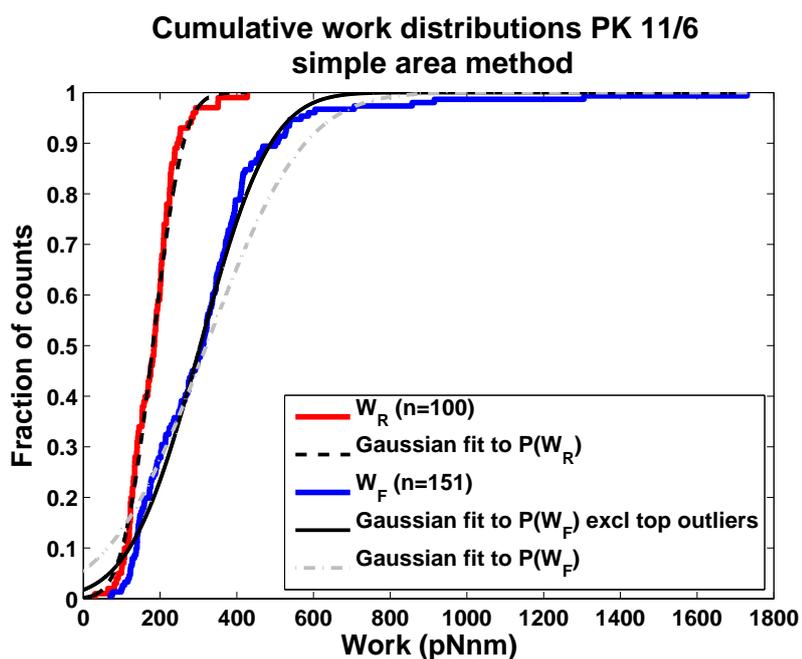
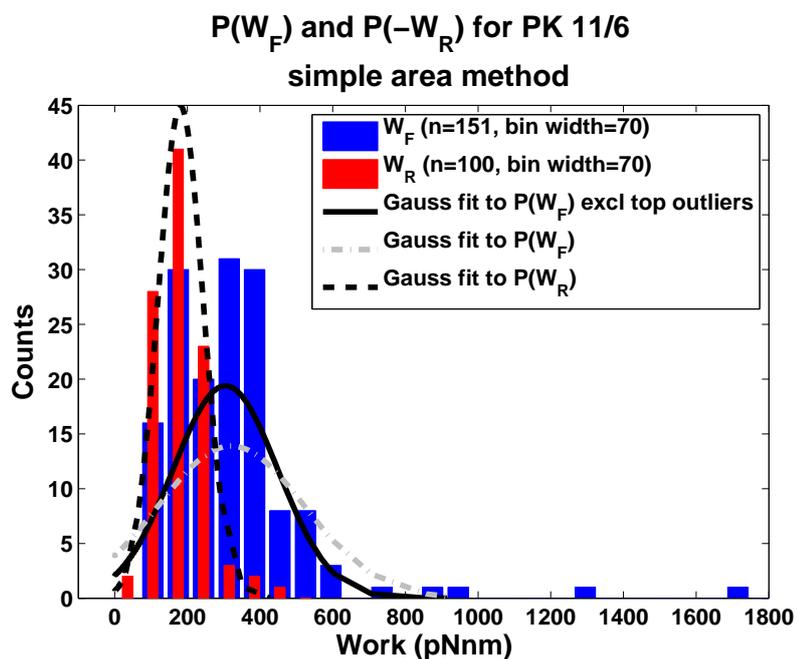


FIGURE F.1: PK 11/6: Simple approach histograms (a) and cumulative distributions (b) of raw works values for opening and closing transitions measured as the areas under the rips/zips on the force-extension curves. For opening transition work, two Gaussian approximations are made, one which excludes the top two outliers and one which includes all data points. The fit that excludes the top two outliers appears much better and is the one used subsequently. Note that although the number of data points in each distribution is different, the fitted probability distributions have been multiplied by the same constant in order to be comparable and intersect at the correct point, which means that they do not completely match the corresponding data histograms in height.

TABLE F.1: Calculated ΔG using 1st rips during molecule extension and 1st zips on molecule relaxation. Simple area method. Variable names explained in text. No uncertainty is given for ΔG_{JE} since its calculation requires bootstrapping or a similar statistical technique (see [46]).

	PK 11/6 \pm SEM	PK 6/11 \pm SEM
n_f	151	49
n_r	100	45
$\langle W_f \rangle$ (kCal/mol)	44.8 ± 1.9	32.6 ± 1.6
$\langle W_r \rangle$ (kCal/mol)	26.1 ± 0.9	28.4 ± 1.7
$(\langle W_f \rangle + \langle W_r \rangle)/2$ (kCal/mol)	35.4 ± 1.0	30.5 ± 1.2
ΔG_{Crooks} (kCal/mol)	38.0 ± 1.0	30.8 ± 1.9
$\Delta G_{JE(f)}$ (kCal/mol)	13.0	9.4
$\Delta G_{JE(r)}$ (kCal/mol)	58.9	45.0
$\langle \Delta G_{JE} \rangle$ (kCal/mol)	35.8	27.3
$\langle W_{\text{stretchhandle}}(f) \rangle$ (kCal/mol)	-1.76 ± 0.06	-1.84 ± 0.09
$\langle W_{\text{stretchPK}}(f) \rangle$ (kCal/mol)	21.4 ± 0.5	17.2 ± 0.5
$\langle W_{\text{stretch}}(f) \rangle$ (kCal/mol)	19.6 ± 0.5	15.4 ± 0.5
$\langle -W_{\text{stretchhandle}}(r) \rangle$ (kCal/mol)	-1.31 ± 0.04	-1.72 ± 0.09
$\langle -W_{\text{stretchPK}}(r) \rangle$ (kCal/mol)	16.1 ± 0.4	16.7 ± 0.6
$\langle -W_{\text{stretch}}(r) \rangle$ (kCal/mol)	14.8 ± 0.4	15.0 ± 0.5
$\langle W_{\text{stretch}} \rangle$ (kCal/mol)	17.7 ± 0.3	15.3 ± 0.4
ΔG_{PK} using $\langle W \rangle$ (kCal/mol)	17.7 ± 1.2	15.3 ± 1
ΔG_{PK} using CFT (kCal/mol)	20.3 ± 0.5	15.5 ± 1
ΔG_{PK} using JE (kCal/mol)	18.1	12.1

TABLE F.2: Calculated ΔG using subdistributions, simple area approach for the PK 11/6 tripple Gaussian distribution subdivision. Calculations are based on the distributions in Figure 7.2. Variable names explained in text in section 7.1.1. No uncertainty is given for ΔG_{JE} since its calculation requires bootstrapping or a similar statistical technique (see [46]). No ΔG_{Crooms} was found for the 3G PK 11/6 distributions since they overlapped so much that the method was meaningless.

	PK 11/6 3G		
	Subdistr A \pm SEM	Subdistr B \pm SEM	Subdistr C \pm SEM
n_f	10	99	43
n_r	3	66	37
$\langle W_f \rangle$ (kCal/mol)	111 \pm 21	47.6 \pm 1.6	27.0 \pm 1.8
$\langle W_r \rangle$ (kCal/mol)	27.8 \pm 4.4	26.6 \pm 1.1	24.0 \pm 1.2
$(\langle W_f \rangle + \langle W_r \rangle)/2$ (kCal/mol)	69 \pm 11	37.5 \pm 1.0	25.5 \pm 1.1
ΔG_{Crooms} (kCal/mol)	NA (too sparse)	37.8 \pm 2.6	NA (overlap)
$\Delta G_{JE(f)}$ (kCal/mol)	NA (too sparse)	20.4	12.2
$\Delta G_{JE(r)}$ (kCal/mol)	NA (too sparse)	59.3	48.4
$\langle \Delta G_{JE} \rangle$ (kCal/mol)	NA (too sparse)	39.7	30.3
$\langle W_{stretchhandle}(f) \rangle$ (kCal/mol)	-3.5 \pm 0.4	-1.8 \pm 0.04	-1.2 \pm 0.05
$\langle W_{stretchPK}(f) \rangle$ (kCal/mol)	27.0 \pm 3.0	21.7 \pm 0.5	19.1 \pm 1.0
$\langle W_{stretch}(f) \rangle$ (kCal/mol)	23.5 \pm 3.0	19.9 \pm 0.5	17.9 \pm 2.1
$\langle -W_{stretchhandle}(r) \rangle$ (kCal/mol)	-1.5 \pm 0.3	-1.34 \pm 0.05	-1.20 \pm 0.06
$\langle -W_{stretchPK}(r) \rangle$ (kCal/mol)	15.7 \pm 2.0	15.6 \pm 0.4	15.9 \pm 1.8
$\langle -W_{stretch}(r) \rangle$ (kCal/mol)	14.2 \pm 2.1	14.3 \pm 0.4	14.7 \pm 1.8
$\langle W_{stretch} \rangle$ (kCal/mol)	21.4 \pm 2.8	17.7 \pm 1.0	16.5 \pm 1.3
ΔG_{PK} using $\langle W \rangle$ (kCal/mol)	48 \pm 11	19.5 \pm 1.5	9.0 \pm 1.8
ΔG_{PK} using CFT (kCal/mol)	NA (too sparse)	20.2 \pm 2.8	NA (overlap)
ΔG_{PK} using JE (kCal/mol)	NA (too sparse)	22.0	13.8

TABLE F.3: Calculated ΔG using subdistributions, simple area approach for the PK 6/11 double Gaussian distribution subdivision. Calculations are based on the distributions in Figure 7.4 (c). Variable names explained in text in section 7.1.1. No uncertainty is given for ΔG_{JE} since its calculation requires bootstrapping or a similar statistical technique (see [46]). No ΔG_{Crooks} was found for the 3G PK 11/6 distributions since they overlapped so much that the method was meaningless.

	PK 6/11 2G	
	Subdistr A \pm SEM	Subdistr B \pm SEM
n_f	38	11
n_r	38	8
$\langle W_f \rangle$ (kCal/mol)	36.9 ± 1.4	19.7 ± 2.3
$\langle W_r \rangle$ (kCal/mol)	30.8 ± 1.8	25.0 ± 4.7
$(\langle W_f \rangle + \langle W_r \rangle)/2$ (kCal/mol)	33.9 ± 1.1	22.3 ± 2.6
ΔG_{Crooks} (kCal/mol)	31.2 ± 1.7	NA (overlap)
$\Delta G_{JE(f)}$ (kCal/mol)	15.3	NA (too sparse)
$\Delta G_{JE(r)}$ (kCal/mol)	47.6	NA (too sparse)
$\langle \Delta G_{JE} \rangle$ (kCal/mol)	31.4	NA (too sparse)
$\langle W_{stretchhandle}(f) \rangle$ (kCal/mol)	-1.93 ± 0.09	-1.54 ± 0.09
$\langle W_{stretchPK}(f) \rangle$ (kCal/mol)	18.0 ± 0.4	14.5 ± 1.6
$\langle W_{stretch}(f) \rangle$ (kCal/mol)	16.1 ± 0.4	12.9 ± 1.6
$\langle -W_{stretchhandle}(r) \rangle$ (kCal/mol)	-1.79 ± 0.09	-0.4 ± 0.3
$\langle -W_{stretchPK}(r) \rangle$ (kCal/mol)	17.0 ± 0.5	3.3 ± 2.4
$\langle -W_{stretch}(r) \rangle$ (kCal/mol)	15.2 ± 0.5	2.9 ± 2.4
$\langle W_{stretch} \rangle$ (kCal/mol)	15.7 ± 0.8	8.7 ± 1.7
ΔG_{PK} using $\langle W \rangle$ (kCal/mol)	18.2 ± 1.3	13.6 ± 2.3
ΔG_{PK} using CFT (kCal/mol)	15.5 ± 1.9	NA (overlap)
ΔG_{PK} using JE (kCal/mol)	22.7	NA (too sparse)