



## PhD Thesis

Jesper Tholstrup Hansen

# Physical and Functional Properties of messenger RNA Pseudoknots

Ribosome stalling *in vivo*, a putative ribosomal helicase, and single  
molecule mechanical unfolding using Optical Tweezers

### Academic advisors

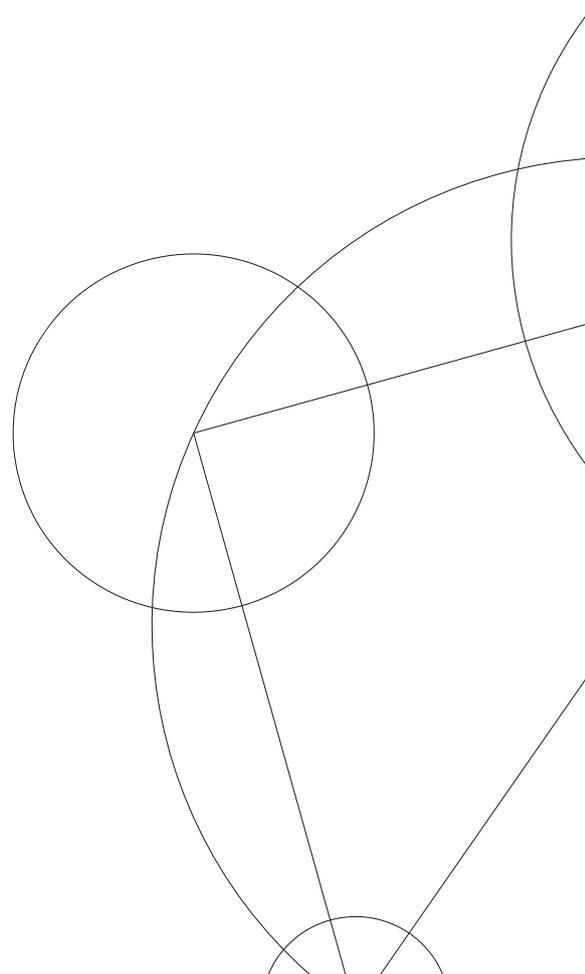
Michael Askvad Sørensen, Department of Biology

Lene Broeng Oddershede, the Niels Bohr Institute

The PhD School of Science, Faculty of Science, University of

Copenhagen, Denmark

Submitted: 07/02/12





## Acknowledgements

I started at the University of Copenhagen as a young, and fairly juvenile, student of biochemistry almost a decade ago. When I look back, this period represents not only a slightly discomfoting large fraction of my life, but also one of the best and most challenging periods. During the last three years, quite a few people have contributed with ideas and practical assistance while I have been working on my thesis.

My supervisor from the Section for Biomolecular Sciences at the Department of Biology, Michael Askvad Sørensen, has been an ever present source of inspiration and have provided invaluable advice and input. Michaels willingness to share his knowledge of the intricate details of mRNA, ribosomes, the process of translation was been crucial for my work. Michael also taught me that nothing starts a good fight like a Northern blot. Michael, it has been fun and you have been an excellent mentor.

Michael's laboratory technician, Marit, has been extremely helpful and has helped continuously through all three years - thanks! All the other students in Michael's lab, Mette, Kim, Maria, and Michael, you have all contributed to a friendly and scientific environment.

My supervisor from the Optical Tweezers group at the Niels Bohr Institute, Lene Broeng Oddershede has also shared essential knowledge and knowhow during my time as a Ph.D. student, most often in a format understandable for non-physicists. Lene was not nearly as scary as I originally envisioned biophysicists when I first heard of our future collaboration. My time in the Optical Tweezers group have been very inspiring and every member of the group helped to create a very welcoming atmosphere.

I worked together with Rebecca Bolt Ettliger during her Master thesis work at the Niels Bohr Institute and we had some interesting discussions regarding the calibration procedure of the optical tweezer we used, how to interpret our results and how to interpret results presented in literature. Congratulation with your recent thesis defense. I hope that you will stay in Denmark and that you invent a renewable source of pure energy - we need it urgently! I also had the pleasure of sharing office and whiteboard with my fellow Ph.D. student Fabian Czerwinski at the Niels Bohr Institute and he introduced me to the magics of iGor Pro and helped me with highly skilled advise during my first visits to the world of single molecule biophysics.

I also had a great time during my stay at Michael O'Connors laboratory at the University of Missouri-Kansas City where I created mutations in the putative ribosomal helicase. I had a very nice time in Kansas City, and Michael and Deepali provided all the support I needed and gave me a delightful and educational time abroad. Not even the occasional tornado warning could ruin the good mood in the lab.

I would also like to thank my family for their understanding during stressful periods of my education and ever present support. Last but not least, my girlfriend Desiree who have been absolutely essential for the completion of this thesis. After giving birth to two beautiful twin girls in April, she has given me the time needed to complete this thesis - I could not have completed this without you! TL!

## Abstract

Programmed -1 ribosomal frameshift is a translational recoding event used by many viruses to create a well defined stoichiometric relationship between translational products of its genome. This recoding event occurs at specific points during the process of translating its genome into its encoded proteins. It is believed that all features required for this frameshift process is location on the translated RNA. One of the requirements for programmed -1 ribosomal frameshift is a downstream RNA structure, typically a pseudoknot, and it is possible that the helicase activity of the ribosome is responsible for “untying” these structures following the frameshift event. Single molecule experiments have demonstrated that, at least for similar structures, a correlation exists between the mechanical strength of these pseudoknots and their ability to induce -1 frameshift during translation.

This thesis have focused on describing functional and physical aspects of completely artificial RNA pseudoknots and have demonstrated that structures with sufficiently high stability is able to stall translating ribosomes on the messenger RNA. This phenomenon had previously been suggested based on single molecule experiments, but is shown here for the first time.

Single molecule mechanical unfolding of RNA structures using force spectroscopy demonstrated that most of the artificial pseudoknots were able for fold as expected and our results were not in disagreement with the correlation between mechanical strength and ability to induce -1 frameshift. Additionally, our sinle molecule experiments showed that the folding dynamics of RNA is complex. A simulation tool was constructed to aid in the quantification of possible sources of error and to help classify the observed unfolding events.

Attempts were made to disrupt the function of the putative ribosomal helicase by creating point mutations in ribosomal proteins S3, S4 and S5 in *Escherichia coli*. Although no effect was observed on -1 frameshift, the putative helicase is still an interesting target, which deserves future attention.

## Resumé

Programmeret -1 ribosomalt læserammeskift er en translationel omkodning som benyttes af mange vira til at skabe et veldefineret støkiometrisk forhold mellem forskellige translationelle produkter. Denne begivenhed indtræffer på bestemte tidspunkter i processen med at omsætte viraens genom til proteiner. Man forventer at alle nødvendige elementer for at inducere -1 rammeskift findes på det RNA der translateres. Et af kravene til programmeret -1 ribosomalt læserammeskift er en nedstrøms RNA struktur, typisk en RNA pseudoknude, og det er muligt at ribosomets helikaseaktivitet er ansvarlig for efterfølgende at opløse disse strukturer. Enkelt-molekyle eksperimenter har vist at der, i hvert fald for beslægtede strukturer, findes en korrelation mellem den mekaniske styrke af disse pseudoknuder og deres evne til at inducere -1 læserammeskift.

Denne afhandling har fokuseret på at beskrive funktionelle og fysiske aspekter af helt kunstige RNA pseudoknuder og viser, at strukturer med tilstrækkelig høj stabilitet er i stand til at bremse translaterende ribosomer på messenger RNA'et. Dette var tidligere foreslået som en mulighed baseret på enkelt-molekyle eksperimenter, men har ikke tidligere været vist.

Enkelt molekyle mekanisk udfoldning af RNA strukturer ved hjælp af en optisk pincet viste, at de kunstige pseudoknuder var i stand til folde som forventet. Der blev ikke fundet uoverensstemmelser med den sammenhæng der er observeret mellem mekanisk styrke og evne til at fremkalde -1 læserammeskift. Ydermere viste vores enkelt-molekyle forsøg at refoldingsdynamikken for RNA strenge er kompleks. Et simuleringsværktøj blev desuden konstrueret til at hjælpe i kvantificering af mulige fejlkilder og til at hjælpe med at klassificere de observerede udfoldingsbegivenheder.

Det blev også forsøgt at hæmme den foreslåede ribosomale helikase ved at skabe punktmutationer i de tre ribosomale proteiner S3, S4 og S5 i *Escherichia coli*. Selvom der ikke blev observeret en effekt på programmeret -1 ribosomalt læserammeskift, er den formodede helikase stadig et interessant mål som fortjener fremtidig opmærksomhed.

## Thesis Objectives and Outline

RNA pseudoknots are used by many viruses to deliberately change the reading frame of translating ribosomes through a process called programmed -1 ribosomal frameshift. Many aspects of this translational recoding process are not fully understood despite more than two decades of research since the description of the frameshift signal from the Rous Sarcoma Virus was described. Given the wide range of viruses which utilize this translational recoding, including HIV and SARS-CoV, it is of considerable interest to understand the recoding mechanism in greater detail.

The first objective was to create a set of completely artificial RNA pseudoknots and measure their ability to induce -1 frameshift *in vivo*, and subsequently use single molecule force spectroscopy to measure their mechanical stability. A previous Ph.D. student, Thomas Møller Hansen, observed a correlation between the mechanical stability of RNA pseudoknot and their ability to induce -1 frameshift. We wished to construct artificial pseudoknots with a wide range of thermodynamical stabilities and investigate if the correlation between mechanical strength and ability to induce -1 frameshift also exists for artificial pseudoknots. Also, we wished to investigate how translation is affected by very stable pseudoknots.

The second objective was to investigate the role of the putative ribosomal helicase in relation to programmed -1 ribosomal frameshift. All models for programmed -1 ribosomal frameshift presented thus far, fails to explain how translation proceeds following -1 frameshift, i.e. how is the RNA structure, responsible for inducing the frameshift, is dissolved allowing continued translation.

### Part I: Programmed -1 Ribosomal Frameshift

Introduces the concept of ribosome stalling during translation of messenger RNA containing thermodynamically stable pseudoknots. The main content is an appended peer-reviewed article:

Tholstrup, J., Oddershede, L. B., and Sørensen, M. (2012). mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Research*, 40(1):303-313.

Co-author statements are given in the Acknowledgements in this article.

### Part II: The Putative Ribosomal Helicase

Work performed both at the University of Copenhagen and at the University of Missouri-Kansas City in the laboratory of Michael O'Connor. The putative ribosomal helicase, located in the mRNA entrance tunnel of the 30S ribosomal subunit, was altered and the effect on pseudoknot induced -1 frameshift was determined *in vivo*.

### Part III: Single Molecule Force Spectroscopy

In an attempt to identify interesting physical features of our pseudoknots, and to verify that they fold into the expected structures, we subjected our pseudoknots to mechanical

unfolding using a dual trap optical tweezers.

The single molecule experiments were conducted in the Optical Tweezers Group at the Niels Bohr Institute using the commercially available dual trap optical tweezers JPK NanoTracker™. In addition to a general introduction to the field of single molecule force spectroscopy, a calibration protocol for the JPK NanoTracker™ is presented along with results of mechanical unfolding of a select subset of the pseudoknots investigated in Part I.

#### **Part IV: Optical Tweezers - Simulations**

Given my background in the field of molecular biology I had only limited knowledge about single molecule force spectroscopy and what potential pitfalls one should be aware of when conducting experiments using optical tweezers. To gain quantitative information about the consequences of e.g. bead-bead misalignment in the  $z$ -direction and trap-stiffness, a simulation tool was constructed *in silico*. This tool is based on a theoretical description of RNA/DNA as a flexible polymer and allows one to simulate the effects of pulling geometry, trap-stiffness and polymer parameters.

## Contents

<b>1</b>	<b>General Introduction</b>	<b>1</b>
1.1	Translation of Messenger RNA . . . . .	1
1.2	Programmed -1 Ribosomal Frameshift . . . . .	2
1.3	Frameshift Efficiency and Pseudoknot Stability . . . . .	5
1.4	Possible Mechanism . . . . .	6
<b>I</b>	<b>Ribosomal Stalling During Translation <i>in vivo</i></b>	<b>9</b>
<b>2</b>	<b>Introductory Remarks</b>	<b>9</b>
	mRNA pseudoknot structures can act as Ribosomal Roadblocks	10
	Supplementary Information	21
<b>II</b>	<b>The Putative Ribosomal Helicase</b>	<b>31</b>
<b>3</b>	<b>Introduction and Objectives</b>	<b>31</b>
3.1	Target mutations . . . . .	32
<b>4</b>	<b>Materials and Methods</b>	<b>36</b>
4.1	Bacterial strains and plasmids . . . . .	36
4.2	DNA oligonucleotides . . . . .	38
4.3	Equipment . . . . .	39
4.4	Extraction of genomic DNA . . . . .	39
4.5	Electroporation . . . . .	39
4.6	P1vir lysate . . . . .	39
4.7	P1 transduction . . . . .	40
4.8	PCR . . . . .	40
4.9	Sucrose Gradient Sedimentation . . . . .	40
4.10	Frameshift assay . . . . .	41
4.11	Construction of <i>rpsC</i> mutations . . . . .	41
4.11.1	Insertion of FRT-cat/kan-FRT upstream/downstream of <i>rpsC</i> . . . . .	41
4.11.2	S3(R131A) . . . . .	42
4.11.3	S3(R132A) . . . . .	42
4.11.4	S3(R164A) . . . . .	42
4.12	Construction of <i>rpsD</i> mutations . . . . .	42
4.12.1	S4(R44A) . . . . .	42
4.12.2	S4(K45A) . . . . .	42
4.12.3	S4(R47A/X) . . . . .	43
4.13	Construction of <i>rpsE</i> mutations . . . . .	43
4.13.1	S5(R20A) . . . . .	43

---

<b>5</b>	<b>Results</b>	<b>43</b>
5.1	Constructed strains . . . . .	43
5.2	Sucrose Gradient Sedimentation . . . . .	44
5.3	Frameshift assay . . . . .	45
<b>6</b>	<b>Discussion - The Putative Ribosomal Helicase</b>	<b>48</b>
<b>III</b>	<b>Single Molecule Force Spectroscopy</b>	<b>51</b>
<b>7</b>	<b>Introduction to Force Spectroscopy</b>	<b>51</b>
7.1	Theory of Optical Trapping . . . . .	51
7.1.1	The Ray Optics Regime . . . . .	52
7.1.2	The Rayleigh Regime . . . . .	53
7.2	The JPK NanoTracker™ . . . . .	54
7.3	Calibration . . . . .	55
7.3.1	Calibration by Power Spectrum Analysis . . . . .	56
7.3.2	Stokes Calibration . . . . .	58
7.3.3	Calibration of the JPK NanoTracker™ . . . . .	60
7.4	Kinetics . . . . .	62
7.5	Nonequilibrium work and free energy . . . . .	63
7.5.1	Jarzynski's equality . . . . .	63
7.5.2	Crooks Fluctuation Theorem . . . . .	64
<b>8</b>	<b>Materials and Methods</b>	<b>65</b>
8.1	Buffers . . . . .	65
8.2	DNA oligonucleotides . . . . .	65
8.3	DNA Template for <i>in vitro</i> transcription . . . . .	65
8.4	Preparation of pseudoknot containing RNA . . . . .	65
8.5	Single stranded DNA handles . . . . .	65
8.6	RNA/ssDNA hybridisation . . . . .	66
8.7	Sample chamber . . . . .	66
8.8	Pulling experiments . . . . .	66
8.9	Data Analysis . . . . .	67
8.10	Estimation of expected rip-lengths . . . . .	67
<b>9</b>	<b>Results</b>	<b>68</b>
9.1	Validation of Conversion Factor $\beta$ . . . . .	68
9.2	Instrumental Noise . . . . .	70
9.3	Negative Control . . . . .	72
9.4	Pulling Statistics . . . . .	72
9.5	Pseudoknot 10/6 . . . . .	73
9.5.1	Possible structures . . . . .	73
9.5.2	Mechanical unfolding . . . . .	73

9.5.3	Kinetics of 10hp, 10/6U, and 10/6 . . . . .	79
9.5.4	Thermodynamics of 10hp, 10/6U and 10/6 . . . . .	80
9.5.5	Unfolding of 10hp, 10/6U, and 10/6 - Summary . . . . .	84
9.6	Pseudoknot 11/6 and 6/11 . . . . .	84
9.7	Pseudoknot 16/6 . . . . .	86
9.8	Pseudoknot 22/6a . . . . .	86
9.9	Pseudoknot 22/6b . . . . .	89
9.9.1	Mechanical Unfolding . . . . .	89
9.9.2	Kinetics of 22hp/b, 22/6bU, and 22/6b . . . . .	95
9.9.3	Thermodynamics of 22hp/b, 22/6bU, and 22/6b . . . . .	96
9.9.4	Unfolding of 22hp/b, 22/6bU, and 22/6b - Summary . . . . .	99
9.10	Pseudoknot 22/6c . . . . .	100
9.10.1	Mechanical Unfolding . . . . .	100
9.10.2	Kinetics of hairpin 22hp/c . . . . .	105
9.10.3	Thermodynamics of hairpin 22hp/c . . . . .	106
9.10.4	Unfolding of Hairpin 22hp/c - Summary . . . . .	107
9.11	Summary - Single Molecule Experiments . . . . .	108
9.11.1	Unfolding data . . . . .	108
9.11.2	Structural kinetics . . . . .	108
9.11.3	Thermodynamics . . . . .	109
<b>10</b>	<b>Discussion - Single Molecule Experiments</b>	<b>111</b>
10.1	Unfolding Length and Unfolding Force . . . . .	111
10.2	Refolding Length and Refolding Force . . . . .	113
10.3	Unfolding Kinetics . . . . .	114
10.4	Thermodynamics . . . . .	115
10.5	Mechanical Stability and Frameshifting Efficiency . . . . .	117
10.6	Breaking Tethers . . . . .	117
<b>IV</b>	<b>Optical Tweezers - Simulations</b>	<b>119</b>
<b>11</b>	<b>Introduction - Nucleic Acid as Biopolymers</b>	<b>119</b>
11.1	Stretching of Biopolymers . . . . .	119
11.2	Polymer Parameters . . . . .	120
<b>12</b>	<b>Implementation</b>	<b>121</b>
12.1	Step 2: Calculate trap-trap distance for zero extension of the tether . . . . .	122
12.2	Step 3: Calculate surface distance and pulling angle . . . . .	122
12.3	Step 4. Increase trap-trap distance - maintain bead-bead distance . . . . .	123
12.3.1	WHILE EWLC length of tether > bead-bead yz-distance . . . . .	124
12.3.2	WHILE EWLC length of tether < bead-bead yz-distance . . . . .	125
12.4	Summary of Implementation . . . . .	125

---

<b>13 Results</b>	<b>126</b>
13.1 Force-extension Curve of a DNA Tether . . . . .	126
13.2 Force-extension Curve for a Compound Tether . . . . .	126
13.3 Practical Implications of $z$ -level Misalignment . . . . .	128
13.4 Estimation of Unfolding Length . . . . .	129
13.4.1 Effect of trap-stiffness on unfolding length . . . . .	130
13.4.2 Impact of $z$ -level misalignment . . . . .	133
13.4.3 Impact of parameter variation . . . . .	133
<b>14 Discussion - Optical Tweezers Simulations</b>	<b>134</b>
<b>V Final Conclusions and Perspectives</b>	<b>137</b>
<b>References</b>	<b>141</b>
<b>A Effective Spring Constant and Damping Function</b>	<b>151</b>
A.1 Effective Spring Constant $\kappa_e$ . . . . .	151
A.2 Damping Function $Q(F, \kappa)$ . . . . .	152
<b>B Data Analysis - iGor Pro</b>	<b>155</b>
B.1 Construction of Force-Extension Curves . . . . .	155
<b>C Viscosity of buffer R</b>	<b>157</b>
<b>D Crooks Fluctuation Theorem - Implementation</b>	<b>159</b>
D.1 Integrating Force-extension Curves . . . . .	160
D.2 Calculating the Correction Term . . . . .	160

## List of Abbreviations

Abbreviation	Defined	Abbreviation of...
-1 PRF	Page 2	Programmed -1 Ribosomal Frameshift
bp	Page 5	Base pair
BWYV	Page 5	Beet Western Yellow Virus
CFT	Page 64	Crooks Fluctuation Theorem
CDF	Page 74	Cumulative Distribution Function
<i>E. coli</i>	Page 1	<i>Escherichia coli</i>
FRT	Page 36	FLP recognition target
HIV	Page 3	Human Immunodeficiency virus
IBV	Page 3	Infectious Bronchitis virus
IPTG	Page 41	Isopropyl $\beta$ -D-Thiogalactopyranoside
mRNA	Page 1	Messenger RNA
nt	Page 4	Nucleotide
OD	Page 39	Optical Density
ORF	Page 3	Open Reading Frame
PDF	Page 64	Probability Density Function
pN	Page 51	pico Newton
PSD	Page 55	Power Spectral Density
QPD	Page 55	Quadrant Photo Diode
RRL	Page 5	Rabbit Reticulate Lysate
rRNA	Page 1	Ribosomal RNA
RSV	Page 3	Rous Sarcoma Virus
<i>S. cerevisiae</i>	Page 48	<i>Saccharomyces cerevisiae</i>
SARS-CoV	Page 3	Severe Acute Respiratory Syndrome Coronavirus
SD		Standard Deviation
SEM		Standard Error of the Mean
SRV	Page 5	Simian retrovirus-1
ssDNA	Page 62	Single Stranded DNA
<i>T. thermophilus</i>	Page 31	<i>Thermus thermophilus</i>
tRNA	Page 1	Transfer RNA
WT	Page 5	Wild-type

# 1 General Introduction

The work presented in this thesis was performed in a collaboration between the Section for Biomolecular Sciences at the Department of Biology, University of Copenhagen and the Optical Tweezers Group at the Niels Bohr Institute, University of Copenhagen. As such, it is intended for researchers with mixed backgrounds such as biology, biochemistry or biophysics. To make it readable for all, regardless of educational background, the thesis starts with a general introduction to the fundamentals of translation, reading frames, and programmed -1 ribosomal frameshift. Though limited to the biological aspects of programmed -1 ribosomal frameshift, this introduction will provide the motivation for all the work of this thesis. A separate introduction for the single molecule biophysics of Part III, and to some extent Part IV, is provided later in the thesis.

## 1.1 Translation of Messenger RNA

The process of translating a sequence of bases on a messenger RNA (mRNA) into an appropriate sequence of amino acids is crucial to all three kingdoms of life. In *Escherichia coli* (*E. coli*) this task is performed by the 70S ribosome, a large and complex machine consisting of 3 different ribosomal RNA species (rRNA) and more than 50 different ribosomal proteins distributed on two subunits of different size, the small subunit (30S) and the large subunit (50S) [Gabashvili et al., 2000]. While the exact function(s) of each component is not known, it has been demonstrated that mutations in either ribosomal proteins or rRNA can have a significant impact on the function of the ribosome. For example, mutations in either ribosomal proteins or rRNA can result in loss of translational accuracy, increased translational accuracy, antibiotic resistance, or ribosomal assembly defects [Agarwal et al., 2011, Kirthi et al., 2006, Björkman et al., 1999, Andersson et al., 1986].

The process of translational elongation can be boiled down to three steps; proper codon-anticodon recognition, peptide bond formation, and mRNA movement through the ribosome (we shall ignore translational initiation and termination in this introduction). The mRNA sequence is decoded in triples of three bases (a codon) matched by an appropriate anticodon on the transfer RNAs (tRNAs), which constitute the link between mRNA sequence and amino acid sequence. After some initial debate as to the number of tRNA binding sites in the ribosome, elegant experiments conducted by Rheinberger and Noller established that the ribosome contains three distinct tRNA binding sites, a finding supported by more recent crystal structures of 70S ribosomes containing all three tRNAs [Rheinberger et al., 1981, Yusupov et al., 2001]. The three tRNA binding sites, called the A-, P-, and E-sites and each tRNA will move through all three sites at different steps of the translational cycle. Later, Moazed and Noller showed that, if translation was divided into smaller steps, the tRNAs occupy hybrid sites where one domain of the tRNA occupies one site on the large subunit and another site on the small subunit [Moazed and Noller, 1989].

In the following description we will consider a situation where translation is well under way and a new amino acid is about to be added to the growing peptide chain. The ribosomal the P-site is occupied by a tRNA linked to the growing chain of amino acids

(the peptidyl-tRNA) and the addition of the next amino acid begins with the arrival of the ternary complex, Aminoacyl-tRNA·EF-Tu·GTP, to the ribosome. The aminoacyl-tRNA (a tRNA with its amino acid) is delivered at the A/T-site<sup>1</sup> of the ribosome where correct codon-anticodon recognition is verified. EF-Tu catalysed GTP hydrolysis moves the tRNA from the A/T-site into the A/A-site where the growing peptide chain is transferred from the peptidyl-tRNA in the P/P-site to the aminoacyl-tRNA in the A/A-site (peptidyl transfer). Presumably though altered affinities in the large subunit and possibly through a partial reorientation of the 30S subunit, the deacetylated-tRNA (empty tRNA) in the P/P-site moves to the P/E-site and the peptidyl-tRNA in the A/A-site moves to the A/P-site [Moazed and Noller, 1989, Frank and Agrawal, 2000]. Binding of an EF-G·GTP complex and subsequent GTP hydrolysis brings about the translocation step during which several things occur: the two tRNAs in the P/E-site and in the A/P-site are moved to the E/E-site and the P/P-site, respectively, pulling the mRNA through the 30S subunit, the 30S subunit rotates about 6 degrees relative to the 50S subunit, and the head of the 30S subunit moves relative to the shoulder which changes the configuration of the mRNA entrance tunnel. Release of EF-G causes the 30S subunit to return to its original orientation relative to the 50S subunit, reconstructing the original configuration of the mRNA entrance tunnel. The ribosome is now ready for another cycle of translation [Yusupova et al., 2001, Frank and Agrawal, 2000].

The complexity involved in proper tRNA selection is profound, and the exact mechanism is probably not fully understood. It appears, however, that aminoacyl-tRNA verification takes place both before and after GTP-hydrolysis by EF-Tu [Blanchard et al., 2004, Ogle et al., 2002]. Incredibly, the ribosome is able to perform this cycle with an average rate 12 cycles per second with an error rate of around  $10^{-4}$  [Kramer and Farabaugh, 2007, Sørensen et al., 1989]

### 1.2 Programmed -1 Ribosomal Frameshift

As mentioned above, translation involves the decoding of mRNA in codons of three bases on the mRNA, consequently the mRNA contains three reading frames (in each direction). As the primary structure of proteins is crucial for their function, and thus overall cellular fitness, the ribosome must maintain the initial reading frame throughout the process of translation. There are, however, an increasing number of known cases where certain factors are able to deliberately alter the translational reading frame and induce programmed translational frameshift. The RECODE database currently contains approximately 1,500 genes known to utilize programmed frameshift [Bekaert et al., 2010]. A translational frameshift refers to an event where forward or backward ribosomal movement relative to the mRNA encompass one or two bases, and not the usual three bases. Such an event will change the reading frame and subsequent translation will decode a different set of mRNA codons. This is illustrated for -1 and -2 frameshift in Figure 1.

Though several types of programmed frameshift are known to occur, the focus here will be limited to programmed -1 ribosomal frameshift (-1 PRF), a type of frameshift

---

<sup>1</sup>The first letter denotes the site on the small subunit, the second letter denotes the site on the large subunit

Protein	E	L	V	I	S	L	I	V	E	S	#	in-frame		
mRNA	5'-	GAG	UAA	GUC	AUC	UCG	CUA	AUA	GUU	GAA	UCA	UAA	-3'	
<hr/>														
Protein	E	V	S	H	L	A	N	S	#			-1 frameshift		
mRNA	5'-	GAG	GUA	AGU	CAU	CUC	GCU	AAU	AGU	UGA	AUC	AUA	A	-3'
<hr/>														
Protein	E	S	K	S	S	R	#					-2 frameshift		
mRNA	5'-	GAG	AGU	AAG	UCA	UCU	CGC	UAA	UAG	UUG	AAU	CAU	AA	-3'

Figure 1: During translation the ribosome decodes the mRNA in codons of three bases which define the amino acid sequence of the resulting protein. Translation terminates when the ribosome encounters a stop codon (red box). In the illustration here, a -1 or -2 frameshift occurs after the GAG codon, resulting in an alteration of the protein sequence and premature termination. In the case of -1 frameshift (red wedge) the ribosome moves forward by only two bases after the GAG codon which results in a “reuse” of the 3' G of the GAG codon (magenta). In the case of -2 frameshift (blue wedge) the ribosome moves forward by only one base resulting in the “reuse” of AG in the following codon (magenta).

often seen in translation of viral genomes. Frameshift signals which purposely change the reading frame, have been found in several positive stranded RNA and double stranded RNA viruses, including retroviruses such as the Alpharetrovirus Rous Sarcoma Virus (RSV, the first characterized viral frameshift signal) [Jacks and Varmus, 1985], the Deltaretrovirus Bovine Leukemia Virus [Rice et al., 1985], the Lentivirus Human Immunodeficiency Virus (HIV) [Jacks et al., 1988b], and positive strand viruses such as the Infectious Bronchitis Virus (IBV) [Brierley et al., 1987], the coronavirus responsible for Severe Acute Respiratory Syndrome (SARS-CoV) [Thiel et al., 2003, Baranov et al., 2005], and the Flavivirus West Nile Virus [Melian et al., 2010]. Typically, -1 PRF occurs near the 3'-end of one open reading frame (ORF) allowing the ribosome to continue translation of another overlapping ORF. In the Mouse Mammary Tumor Virus, -1 PRF is induced at *two* different locations, first in the overlap between the *gag-pro* ORFs and in the overlap between *pol-pro* ORFs [Jacks et al., 1987].

The frameshift frequency<sup>2</sup>, i.e. the fraction of ribosomes which changes reading frame, is not 100% and only a fraction of the translating ribosomes continue into the downstream ORF. This creates two different translational products with the same N-terminus but with different C-termini as is illustrated for the SARS-CoV genome in Figure 2.

It is believed that the frameshift frequency serves to create an appropriate ratio between viral components which allow proper viral particle assembly and packaging of genetic material. In support of this, it has been demonstrated that the -1 frameshift frequency, which vary between viruses, is critical to viral proliferation [Dinman and Wickner, 1992, Du-

<sup>2</sup>Frameshift frequency and frameshift efficiency will be used interchangeably in this thesis and denote the fraction of ribosomes which undergo -1 frameshift during translation.

lude et al., 2006, Melian et al., 2010].

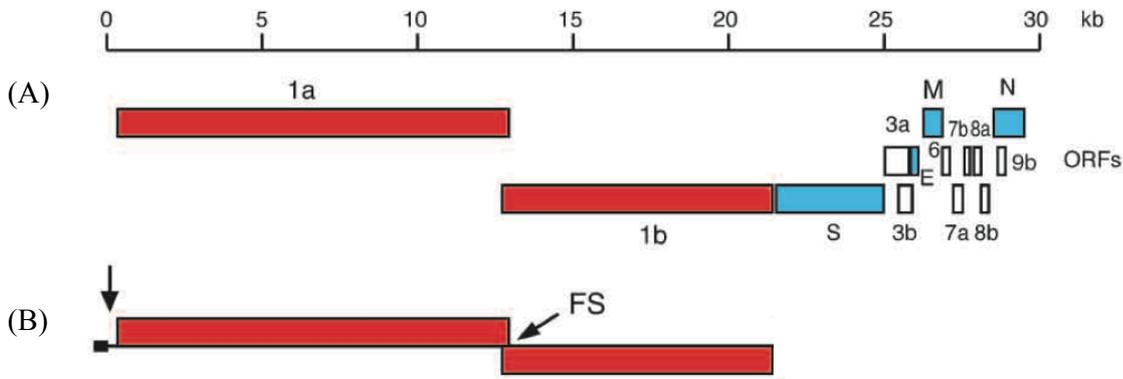


Figure 2: The SARS-CoV genome. (A): Open reading frames in the genome of SARS-CoV including the two overlapping reading frames 1a and 1b. Reading frame 1b is located in the -1 reading frame relative to 1a.(B): Translation with -1 frameshift (FS) results in a fusion of protein 1a and protein 1b while translation without frameshift terminates after the 1a reading frame. The frameshift frequency determines the ratio of 1a to 1a-1b protein. Figure is modified from [Thiel et al., 2003]

A number of well known viruses are among those utilizing -1 PRF: Flaviviruses annually infect more than 100 million people, SARS-CoV infected almost 10,000 people and killed just under 1,000 people during the 2003 outbreak, and HIV/AIDS is a well known killer [Thiel et al., 2003, Melian et al., 2010]. As -1 PRF constitutes a potential therapeutic target it is of great interest to understand the underlying mechanism.

It has been shown that relevant levels of -1 PRF requires three elements; a homopolymeric slippery sequence on the mRNA where the frameshift occurs, a downstream RNA structure, and a spacer region between the slippery sequence and the RNA structure [Brierley et al., 1991, Brierley et al., 1992, Naphthine et al., 1999]. It is believed that the spacer region of typically 6-9 nucleotides (nt) serves to place the ribosome on the slippery sequence when the downstream structure is encountered and that the structures serve a barrier for the forward motion of the ribosome. The slippery sequence for -1 PRF complies with the consensus sequence X XXY YYZ, where X denote identical bases and Y denote identical bases, Z denotes any base, and spaces indicated incoming reading frame (0 frame: X XXY YYZ, -1 frame: XXX YYY Z) [Rice et al., 1985, Brierley et al., 1992, Dinman and Wickner, 1992]. It has been shown that slippery sequences of this nature, where Z is different from X and Y, promotes -1 frameshifting exclusively whereas other sequences can induce e.g. +1 frameshift [Brierley et al., 1992, ten Dam et al., 1994, Belcourt and Farabaugh, 1990]. The downstream structure is often an mRNA pseudoknot but hairpin structures are used by some viruses. Initially there was some debate as to whether or not hairpin structures could induce frameshift, but it has been demonstrated that both hairpins and pseudoknots are able to induce -1 both frameshift *in vivo* and *in vitro* [Yu et al., 2011, Brierley et al.,

1997].

Although it cannot be excluded as a possibility, it is believed that the mRNA contains all features required to induce frameshift and that no additional cellular factors, such as pseudoknot binding proteins, are required. Naturally, it is difficult to prove that no additional factors are involved, but ten Dam et al. showed that the frameshift efficiency of the Simian Retrovirus-1 (SRV) frameshift signal remained constant even when a short pseudoknot-forming RNA species was added as competitor [ten Dam et al., 1994].

The difference between a hairpin and an H-type pseudoknot is illustrated in Figure 3 and a more general description of pseudoknots is provided by C. W. Pleij [Pleij, 1990].

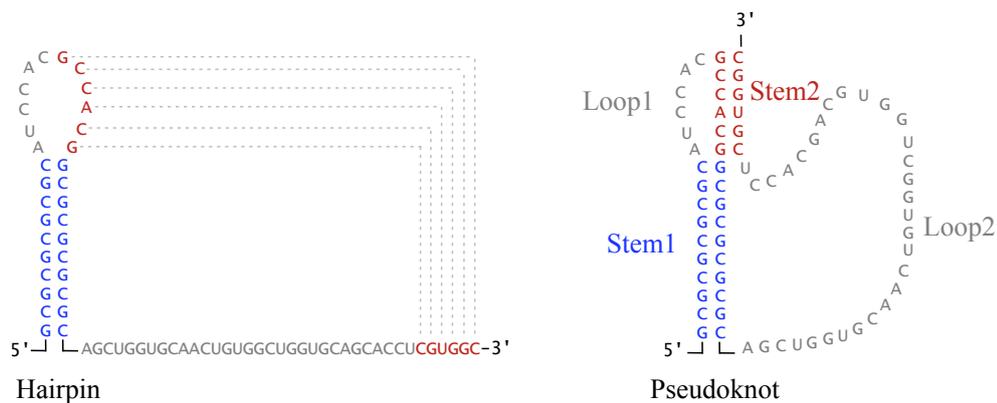


Figure 3: Illustration of two RNA structures, a hairpin (left) and a pseudoknot (right). The pseudoknot is formed by base pairing between the loop sequence of a hairpin and a downstream sequence. The pseudoknot, a classical H-type pseudoknot in this illustration, contains two stems (blue and red) and two loops.

### 1.3 Frameshift Efficiency and Pseudoknot Stability

Some of the earliest investigations of the structural requirements for -1 PRF was performed using the IBV frameshift signal subjected to thorough mutational analysis and used for *in vitro* translation using rabbit reticulate lysate (RRL) and *in vivo* translation in *E. coli* [Brierley et al., 1991, Brierley et al., 1992, Napthine et al., 1999, Brierley et al., 1997]. These experiments indicated, that the simple thermodynamic stability of the frameshift inducing pseudoknot was inadequate to explain their role in frameshift. For example, substituting a G-C base pair (bp) with a C-G bp was permitted at certain positions but not at other positions in the structure, and an increase in -1 frameshift efficiency by a factor of six was observed when increasing the length of stem1 from 10 bp to 11 bp [Brierley et al., 1991, Napthine et al., 1999]. Similar effects were observed for a variant of the SRV pseudoknot where C-G for G-C substitutions changed frameshift efficiency and point mutations (A28C, A26C) in loop2 decreased frameshift efficiency by approximately 70% compared to the wild-type (WT) [Olsthoorn et al., 2010]. In agreement with these mutational studies, the crystal structure of the Beet Western Yellow Virus (BWYV) pseudoknot and the NMR structure of the SRV pseudoknot (not the WT SRV pseudoknot but an NMR-optimized

version, referred to as the “NMR” SRV in literature, the same used by Olsthoorn et al. above) revealed extensive tertiary interactions between the stems and the loop-nucleotides. For the SRV pseudoknot, the number of hydrogen bonds involved in tertiary interactions exceeded the number of hydrogen bonds involved in Watson-Crick base pairing [Michiels et al., 2001, Su et al., 1999].

The development of single molecule techniques have added to the understanding of -1 PRF by investigating mechanical stability, thermodynamics, and kinetics of a wide range of pseudoknots and hairpin structures [Hansen et al., 2007, Green et al., 2008, Chen et al., 2007, Chen et al., 2009]. Hansen et al. proposed, based on mechanical unfolding of two similar IBV-inspired pseudoknots, that a correlation existed between a pseudoknots ability to induce -1 frameshift and its mechanical strength [Hansen et al., 2007]. The correlation was later confirmed by Chen et al. who investigated the effect of potential triplex interactions on the overall mechanical stability of the  $\Delta$ U177 human telomerase pseudoknot [Chen et al., 2009]. This correlation between mechanical stability and frameshift efficiency makes intuitive sense in light role attributed to pseudoknots as translational barriers which somehow forces the ribosome to slip back into the -1 reading frame.

### 1.4 Possible Mechanism

Although the exact mechanism underlying -1 PRF is unclear, a number of possible mechanisms have been proposed. Originally, Jacks et al. proposed, based on protein sequences, that both the A/A-site and P/P-site tRNA slipped into the -1 reading frame simultaneously during translation of the RSV frameshift signal. This gave rise to the “simultaneous slippage” model where the consensus slippery sequence ensures that non-cognate codon-anticodon interaction is isolated to the wobble position [Jacks et al., 1988a]. According to this model, the P/P-site and A/A-site tRNAs slips into the -1 reading frame prior to peptidyl transfer, which places the time of frameshift immediately following complete tRNA accommodation into the A/A-site. This model was later modified by Weiss et al. who placed the frameshift event after peptidyl-transfer and possibly during the translocation step [Weiss et al., 1989].

Jacks et al. found no effect on frameshift efficiency *in vitro* of mutations in the mRNA immediately upstream from the slippery sequence, a finding which was challenged by findings of Léger et al. who showed that mutations in the nucleotides upstream from the slippery sequence from several different viruses, including HIV and SARS-CoV, resulted in changes of frameshift efficiency *in vivo* [Jacks et al., 1988a, Léger et al., 2007]. It is possible that this discrepancy can be explained by a limited resolution in the experiments of Jacks et al. where changes in frameshift efficiency were divided into only three groups: no change,  $\approx 5$  fold, and  $>10$  fold [Jacks et al., 1988a]. Léger et al. also isolated mutations in the 16S rRNA which affected frameshift efficiency - mutations which could be implicated in tRNA binding at the E-site [Léger et al., 2007]. Based on their observations Léger proposed an updated version of the simultaneous slippage model, called “The three tRNA -1 PRF model”, which include a three tRNAs as a model for -1 PRF. This model is illustrated in Figure 4.

In the model by Léger et al. the tRNAs in the hybrid P/E and A/P-sites (called P/E\*

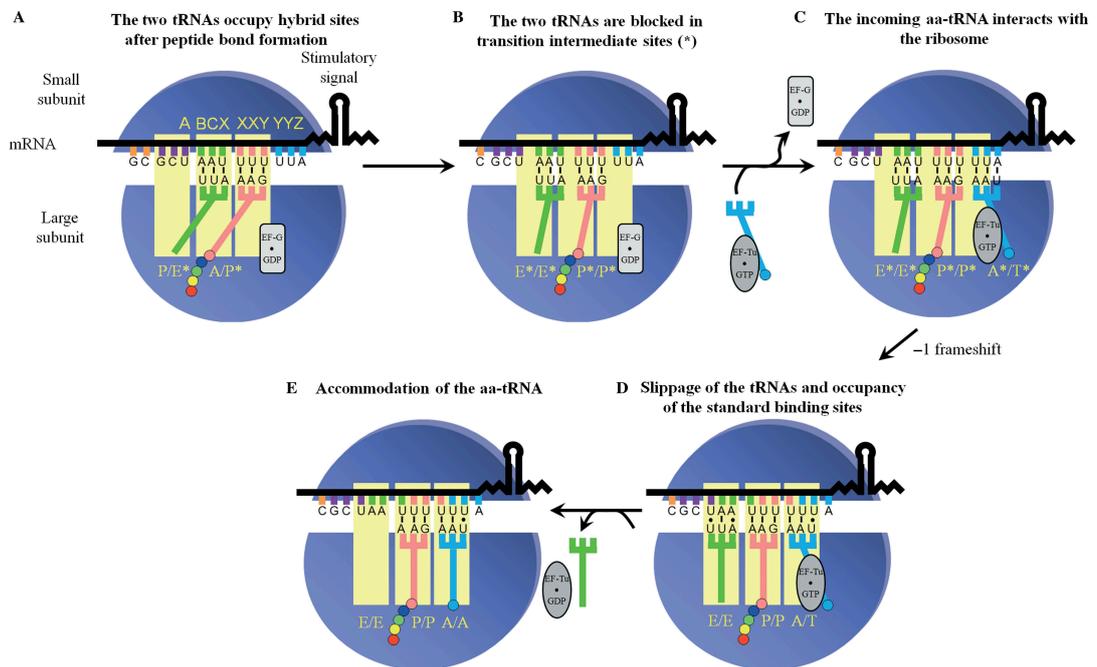


Figure 4: A model for -1 PRF adopted from [Léger et al., 2007]. The RNA structure at the mRNA entrance tunnel prevents proper translocation of the ribosome and accommodation of the next tRNA (A-C). Tension builds up and successive tRNA slippage into the -1 reading frame relieves this tension (D) after which translation continues in the -1 reading frame (E).

and A/P\* in Figure 4A) after peptidyl transfer, are prevented from moving to the E/E and P/P-sites by the mechanical hindrance caused by the RNA structure in the mRNA entrance tunnel. Consequently, they occupy states between the P/E and E/E and A/P and P/P states, called E\*/E\* and P\*/P\* (figure 4B). Because the P\*/P\*-tRNA partially occupies the 30S A-site, the next Aminoacyl-tRNA·EF-Tu·GTP complex places the tRNA at a site called A\*/T\* which differs from the A/T-site it would normally occupy (Figure 4C). The -1 frameshift occurs during accommodation of the A\*/T\*-tRNA into the A/T-site via successive tRNA slippage of all three tRNAs (Figure 4D). After dissolving the RNA structure, translation continues as usual in the -1 reading frame [Léger et al., 2007]. According to this model, the main driving force behind -1 PRF is the tRNAs affinity for their proper binding sites.

The “three tRNA -1 PRF model” proposed by Léger et al. agrees with many experimental observations, e.g. tRNA tension during translocation of a ribosome-pseudoknot complex [Namy et al., 2006], effects of E-site mRNA mutations and ribosomal mutations [Léger et al., 2007], effects of mutations in 16S rRNA which increase the error rate (*ram* mutations) [Léger et al., 2004], that the mRNA entrance tunnel is too narrow to accommodate the pseudoknot [Yusupova et al., 2001, Takyar et al., 2005], and that heelprinting have indicated that ribosomes *may* pause when they encounter the RNA structure [Tu et al.,

1992, Kontos et al., 2001]. This pausing has however not been observed *in vivo* during translation of hairpin structures [Sørensen et al., 1989].

It is, however, still possible that -1 frameshift occurs through different pathways depending upon e.g. slippery sequence, RNA structure and organism.

## Part I

# Ribosomal Stalling During Translation *in vivo*

## 2 Introductory Remarks

The possible correlation between mechanical strength and frameshift efficiency observed by Thomas Møller Hansen during his Ph.D. work [Hansen et al., 2007] prompted us to create a set of completely artificial pseudoknots with varying size and predicted stability. These pseudoknots were designed to challenge our understanding of pseudoknot induced -1 frameshift and included pseudoknots with increasing length of stem1 and pseudoknots with a constant length of stem1 but with destabilizing C-G to A-U substitutions.

All the pseudoknots were created with inspiration from the IBV pseudoknot, i.e. the length of stem2 was held constant at 6 bp and the length of loop2 was 31-32 nt. The length of loop1 was, however, increased from 1-2 nt to 6 nt, to allow increased lengths of stem2 if desired.

The structures were constructed *in silico* under the following restrictions:

- The sequence would fold into a H-type pseudoknot as determined by pknotsRG (v. 1.3) with a loop1 length of 6 nt, a loop2 length of 31-32 nt, a stem2 length of 6 bp, and the desired length of stem1 [Reeder et al., 2007]
- No stop codons would be present in-frame or in the -1 reading frame as they would interfere with the frameshift assay
- Codon usage should be appropriate for translation in *E. coli*

The objective was to investigate the frameshift efficiency of these pseudoknots *in vivo* using a *E. coli* frameshift assay. Subsequently, some or all of these pseudoknots would be subjected to mechanical unfolding using Optical Tweezers to determine their mechanical strength in an attempt to further describe the correlation between frameshift efficiency and mechanical strength.

# mRNA pseudoknot structures can act as ribosomal roadblocks

Jesper Tholstrup<sup>1</sup>, Lene B. Oddershede<sup>2</sup> and Michael A. Sørensen<sup>1,\*</sup><sup>1</sup>Department of Biology, Ole Maaløes vej 5, University of Copenhagen, DK-2200 Copenhagen and <sup>2</sup>Niels Bohr Institute, Blegdamsvej 17, University of Copenhagen, DK-2100 Copenhagen, Denmark

Received March 24, 2011; Revised August 5, 2011; Accepted August 7, 2011

## ABSTRACT

Several viruses utilize programmed ribosomal frameshifting mediated by mRNA pseudoknots in combination with a slippery sequence to produce a well defined stoichiometric ratio of the upstream encoded to the downstream-encoded protein. A correlation between the mechanical strength of mRNA pseudoknots and frameshifting efficiency has previously been found; however, the physical mechanism behind frameshifting still remains to be fully understood. In this study, we utilized synthetic sequences predicted to form mRNA pseudoknot-like structures. Surprisingly, the structures predicted to be strongest lead only to limited frameshifting. Two-dimensional gel electrophoresis of pulse labelled proteins revealed that a significant fraction of the ribosomes were frameshifted but unable to pass the pseudoknot-like structures. Hence, pseudoknots can act as ribosomal roadblocks, prohibiting a significant fraction of the frameshifted ribosomes from reaching the downstream stop codon. The stronger the pseudoknot the larger the frameshifting efficiency and the larger its roadblocking effect. The maximal amount of full-length frameshifted product is produced from a structure where those two effects are balanced. Taking ribosomal roadblocking into account is a prerequisite for formulating correct frameshifting hypotheses.

## INTRODUCTION

The reading frame of the vast majority of mRNAs is determined by the start codon after which the downstream cistron is translated in the same frame. Maintenance of the reading frame occurs without further signals to the ribosome. However, examples of genes containing information for programmed frameshifts can be found in most

organisms, or in some of their IS sequences, transposable elements, retroelement-derived sequences or viruses. The sequence-information needed for programmed ribosomal frameshift varies and both +1 and -1 frameshifts can be induced (1–3).

Here, we focus on the frameshifting signal found in several viruses (1), including infectious bronchitis virus (IBV) and SARS-CoV. The signal leads to programmed ribosomal -1 frameshift, whereby multiple proteins are produced from a single polycistronic messenger RNA (mRNA) (4,5). The frameshift efficiency, i.e. the fraction of ribosomes, which change reading frame, is important to ensure a correct stoichiometric relationship between the different products of translation. It has been shown that altered frameshift efficiency has detrimental effects on the proliferation of HIV-I and the yeast L-A viruses (6,7). In order to induce -1 frameshift, these viruses rely on three physical features on the mRNA: a heptanucleotide sequence, a spacer and a downstream structure (8). The heptanucleotide sequence, called the slippery sequence, is where the -1 frameshift occurs and typically has the following sequence: X XXY YYZ, where X, Y and Z denote nucleotide species and spaces indicate initial reading frame. The spacer is a stretch of 6–9 nt positioning the ribosome correctly at the slippery site when encountering the downstream structure. The downstream structure is most often found to be a pseudoknot. The pseudoknot structure probably functions as a physical barrier deforming upon approach of the translating ribosome (9), thereby assisting the frameshifting process; however, geometry and surface charge of the structure may also play a role for the frameshifting (10).

In bacteria and yeast, programmed frameshift signals can have rather different elements, as, e.g. the upstream Shine–Dalgarno binding element in the autoregulatory RF2 gene frameshift site first described in *Escherichia coli* (11) or the different pattern of the +1 frameshift stimulating heptanucleotide sequences present in *Saccharomyces* Ty elements (2). However, many frameshift signals deviate little from those described for the virus-derived system used here and many signals are of

\*To whom correspondence should be addressed. Tel: +45 3532 3711; Fax +45 3532 2128; Email: mas@bio.ku.dk

such general character that ribosomes from different kingdoms of life will respond to them by shifting frame (12). This happens not always with the same efficiency as in the original organism (12,13) and there are even examples found where a frameshift element can direct the ribosomes into  $-2$  or  $+1$  frameshift depending on the test organism (14). Here, we challenged *E. coli* ribosomes by constructing artificial frameshifting signals containing pseudoknot-like structures with strong stems. Using a refined frameshift assay, involving two-dimensional (2D) gel electrophoresis of pulse labelled proteins, we show that a significant amount of frameshifted ribosomes permanently stall within the strongest pseudoknots which therefore efficiently act as roadblocks.

The small ribosomal subunits have been shown to be sensitive towards mRNA secondary structure in the process of translation initiation and mRNA structures can exclude initiation both in eukaryotes during the scanning process (15) and in prokaryotes for binding between the mRNA and the 3'-end of 16S RNA (16). The fully assembled and translating 70S or 80S ribosomes seem to be more robust. It is, however, broadly accepted that mRNA secondary structures can function as obstacles to translating ribosomes (17,18) although examples exist of large secondary structures in mRNA that are translated without any ribosomal delay (19). Nevertheless, there is compelling evidence from *in vitro* experiments showing that ribosomes may pause upstream to such structures, most pronounced if the structures form pseudoknots (20–22). Possibly the lack of rotational freedom in the helix of stem 1, due to the pairing in stem 2, makes pseudoknot structures harder to 'unzip' by the ribosome than simple stem-loop structures (23). This may explain why pseudoknots can pause ribosomes. Examples from nature show the existence of diverse peptide sequences, often present in regulatory circuits, which will stall ribosomes (24), but to our knowledge, a permanent halt of ribosomes caused by mRNA structures has not been shown previously.

Recent single molecule investigations suggest that the mechanical strength of pseudoknots correlate with the ability of the pseudoknot to stimulate frameshift (25–27), at least in a certain interval. However, the calculated Gibbs free energy does not always correlate with frameshift efficiency. Not only the strength of the stems, but also the interaction between the loop and the stems might be of importance for the ability to induce frameshift and for the overall mechanical strength and brittleness of the structure. If the pseudoknot becomes too strong the ribosome, frameshifted or not, might not be able to open it and continue translation, whereby the pseudoknot acts as a roadblock. Often in literature (25–31) frameshifting assays were performed on constructs exhibiting the common feature that the stop codon for the normal reading frame was located at the entrance of the pseudoknot (or inside the pseudoknot) and the stop codon for the successful  $-1$  frameshift was located downstream of the pseudoknot. In most frameshifting assays, the amount of frameshifting is determined by quantifying the amount of full-length frameshifted

versus non-frameshifted products. However, for this to be a correct measure, the frameshifted ribosome must continue translation through the pseudoknot and beyond to the  $-1$  frameshifted stop codon. If the  $-1$  frameshifted ribosome permanently stalls inside the pseudoknot, it would falsely be interpreted as if the ribosome did not frameshift. Therefore, there is a serious pitfall in the classical methods which renders the amount of frameshifted ribosomes to be non-correctly determined, i.e. be underestimated, potentially leading to false hypotheses regarding the physical mechanism of frameshifting.

The observation that strong pseudoknot-like structures can stop translation lead to the hypothesis that the largest amount of frameshifted product will be produced if the pseudoknot is mechanically strong but without a significant roadblocking effect. Most likely, this is exactly the balance exhibited by naturally occurring viral pseudoknots.

## MATERIALS AND METHODS

### Bacterial growth

*Escherichia coli* strain MAS90 [*E. coli* K-12, *recA1*  $\Delta$ (*pro-lac*) *thi* *ara F'*: *lacI<sup>q1</sup>* *lacZ*::*Tn5 proAB<sup>+</sup>*]. Liquid cultures were grown in minimal MOPS media (32) using glycerol as carbon source. Cultures were incubated with shaking at 37°C for at least 10 generations in the log phase prior to being used in frameshift assays.

### Plasmid construction

Pseudoknots were designed using custom-made software, which ensued that the codon usage was appropriate for expression in *E. coli* and that the sequences were likely to fold into the correct structure as determined by pKnotsRG (33). Hence, the resulting sequences are artificial pseudoknot-like structures and there is always a risk that the structure does not fold as anticipated. The selected sequences were synthesized by GeneScript and were subsequently inserted into plasmid OFX302 [containing slippery sequence, spacer and pseudoknot (25)] between HindIII and ApaI restriction sites.

### Frameshift assay

The *in vivo* frameshift assays were performed as described previously (25). Briefly, 1 ml of an exponentially growing culture was induced with Isopropyl  $\beta$ -D-Thiogalactopyranoside (IPTG) to a final concentration of 1 mM at an optical density of 0.4–0.7 measured at 436 nm (OD<sub>436</sub>). After induction for 15 min, the culture was pulse-labelled with  $\sim 10 \mu\text{Ci}$  L-[<sup>35</sup>S]-methionine for 20 s and chased with 100 mg L-methionine for 2 min before being transferred to 25  $\mu\text{l}$  of chloramphenicol (100  $\mu\text{g/ml}$ ) on ice. The cells were harvested by centrifugation and proteins were boiled in SDS buffer and separated by 9% SDS-PAGE. The gel was dried and placed on a phosphor imager screen (Molecular Dynamics) and left to expose for 1–3 days. Relative amount of protein of the

relevant polypeptides was quantified using ImageQuant software and the frameshift efficiency ( $e$ ) was determined as follows:

$$e = \frac{V_{FS}/n_{met,FS}}{V_{FS}/n_{met,FS} + V_{STOP}/n_{met,STOP}}$$

where  $V_{FS}$  is the relative radioactivity in the frameshift product,  $n_{met,FS}$  is the number of methionines in the frameshift product,  $V_{STOP}$  is the relative radioactivity in the in-frame stop product and  $n_{met,STOP}$  is the number of methionines in the in-frame stop product.

### Two-dimensional SDS-PAGE

Two-dimensional SDS-gels were performed as described (34) with a few modifications (35) using samples from the frameshift assay described above. The frameshift efficiency was determined as described for the frameshift assay above, although polygonal shapes were used to encircle the polypeptides of interest and quantify the relative amount of radioactivity in them.

Polypeptides originating from stalled ribosomes were found as radioactive polypeptides with appropriate isoelectric point and molecular weight appearing on gels when the translated transcript contained a pseudoknot. These polypeptides were absent when a transcript without a pseudoknot was translated. The weakest stalled protein spots were difficult to distinguish from spots originating from endogenous gene expression on these gels (compare to the 0 construct in Supplementary Figure S5) and their determination is connected with some uncertainty. The statistical analysis used to compare the stalling efficiency between pseudoknot 22/6a and 22/6b was an unpaired one-tailed Student's  $t$ -test with a significance level of 0.05.

### Northern blots

Total RNA was extracted from 1.5 ml culture samples by the 'Hot-phenol' extraction method and separated according to size by electrophoresis on 1.2% agarose, 6% formaldehyde gels in recirculating 1xMOPS buffer. Capillary blots were performed onto Hybond-N<sup>+</sup> (Perkin Elmer) membranes, and the RNA was crosslinked to the membrane by 0.12 J/cm<sup>2</sup> UV light in a Stratilinker 1800. Riboprobes covering mRNA sequences as described in Figure 4 were made by T7 RNA polymerase transcripts from the pMAS39 'downstream' template (19) or from templates made by PCR where one primer included 'hanging out' T7 promoter sequences (*gene10* and *lacZ* 5' probes). The riboprobes were synthesized in the presence of 32-P-UTP and the final specific activity was about 40 Ci/mmol of nucleotide. Hybridization and stripping of membranes were performed following standard protocols (Amersham, Hybond-N+ booklet, 2006). The membranes were wrapped in Saran wrap and placed on a phosphor imager screen (Molecular Dynamics) and left to expose over night. Signals were visualized using ImageQuant software.

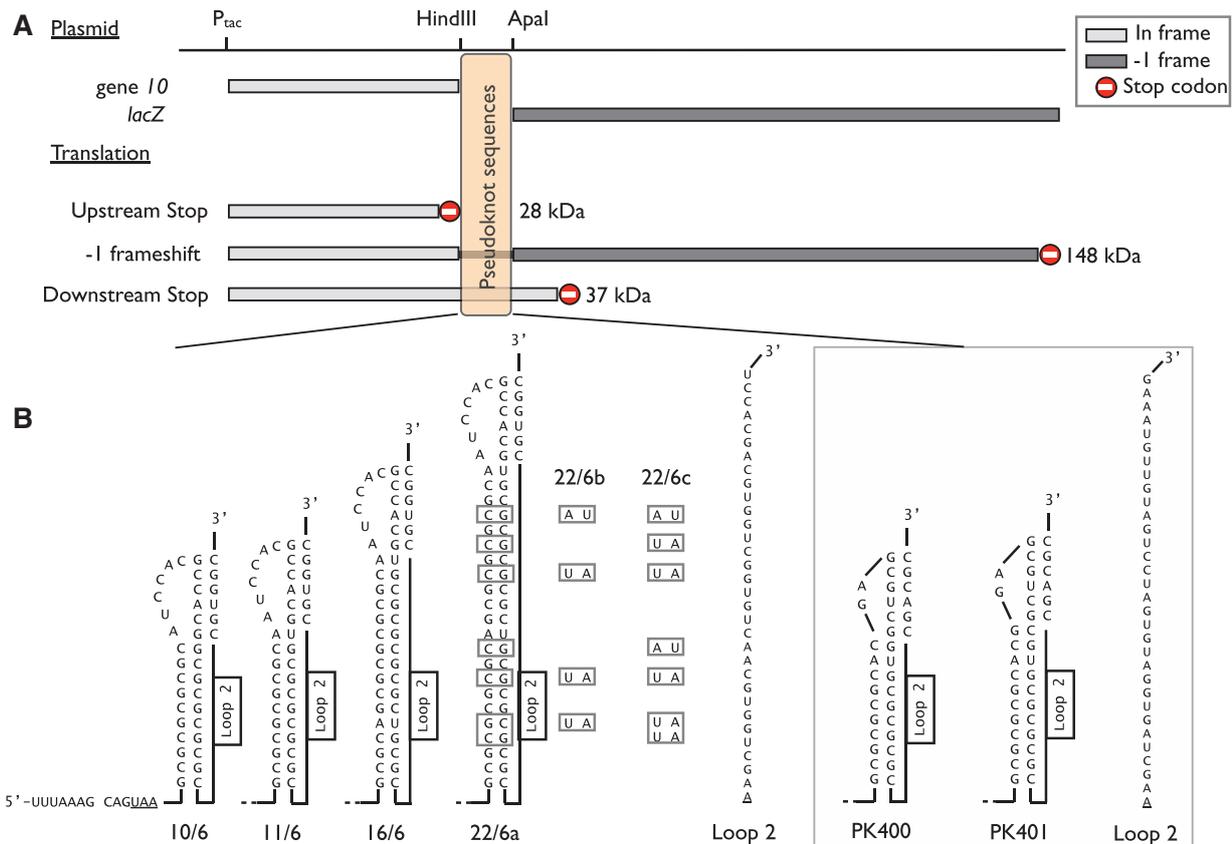
## RESULTS

### mRNA pseudoknot constructs to separate programmed stop from ribosome stalling

We created a series of plasmids containing different pseudoknots and where the in-frame stop codon was placed either immediately upstream ('Upstream stop') or ~150 nt downstream ('Downstream stop') from the pseudoknot (Figure 1A). The 'Upstream stop' constructs had an in-frame stop codon in the spacer between the slippery sequence and the pseudoknot. This caused non-frameshifted ribosomes to produce a 28 kDa polypeptide (*gene10* from phage T7) while ribosomes undergoing a  $-1$  frameshift continued through the pseudoknot and into *lacZ* producing a 148 kDa fusion protein of the T7 *gene10* and *lacZ* sequences. In the 'Downstream Stop' constructs we replaced the UAA stop codon immediately upstream from the pseudoknot with a lysine codon (AAA). This change caused non-frameshifting ribosomes to continue through the pseudoknot and terminate at a downstream UGA codon producing a 37 kDa polypeptide. The pseudoknot constructs based on the plasmid OFX302 (25) are detailed in Figure 1B. We systematically increased the length of stem 1 and in pseudoknot 22/6a through 22/6c, we exchanged GC with UA base pairs, thus, gradually decreasing the stability of stem 1.

Often, the number of ribosomes which undergo  $-1$  frameshift has been determined from constructs such as our 'Upstream stop' constructs, by separating radioactively labelled proteins by SDS-PAGE and quantifying the relative amount of protein in each of the two polypeptides (28 versus 148 kDa). Given the limited resolution of SDS-PAGE, it is, however, impossible to clearly differentiate between polypeptides produced by ribosomes that terminate at the in-frame UAA stop codon and ribosomes that undergo  $-1$  frameshift but stall within the pseudoknot. In order to overcome this problem, we invoked 2D SDS-PAGE (34) whereby polypeptides were separated not only by molecular weight but also by their isoelectric point (pI).

While polypeptides originating from ribosomes stalled in the pseudoknot varied only slightly in molecular weight, they varied significantly in their pI. Based on the 'Downstream Stop' construct, we calculated a theoretical 2D SDS-PAGE assay of a growing polypeptide as consecutive codons are translated (shown in Figure 2A). At around 28 kDa, the trace splits into two, the triangles denote the non-frameshifted product and the circles denote the  $-1$  frameshifted product. Red symbols denote codons inside the pseudoknot. Experimental data originating from the 'Downstream Stop' construct is shown in Figure 2B, the theoretically expected features are indeed present, e.g. both the non-frameshifted (DS-stop) and the  $-1$  frameshifted (FS) products are visible. The heat shock proteins GroEL and DnaK serve as landmarks on the gel. Interestingly, a series of polypeptides originating from ribosomes stalled inside the pseudoknot appeared (inside dashed red line). For comparison, a standard SDS-PAGE of the same sample is shown in Figure 2C, here, the second level of information



**Figure 1.** Frameshift assay and pseudoknot structures. (A) All plasmid constructs contain an IPTG inducible promoter in front of T7 gene10 (light grey), a complete frameshift signal, and lacZ (dark grey). The frame shift stimulating pseudoknot-like structure is inserted downstream of gene10. Immediately, downstream from the pseudoknot lacZ is inserted in the -1 reading frame relative to gene10. In the 'Upstream Stop' construct the non-frameshifting ribosomes will translate gene10 and terminate at a UAA stop codon in the spacer sequence and produce a 28 kDa polypeptide. Ribosomes undergoing -1 frameshift at the slippery sequence translate lacZ thus producing ~148 kDa polypeptide. In the 'Downstream Stop' construct the UAA stop codon is replaced by an AAA lysine codon thus resulting in ~37 kDa polypeptide being produced by non-frameshifting ribosomes which terminate at an UGA stop codon downstream from the pseudoknot. (B) Sequence and structure of the inserted pseudoknots, the slippery sequence and the spacer. In pseudoknot, 10/6, 22/6a, 22/6b and 22/6c the first base in loop 2 has been removed in order to maintain the downstream reading frame (underlined). The boxed insert in panel B shows the structure and sequence of previously described constructs (25).

(isoelectric point) is lost and the relative blurry bands are difficult to interpret.

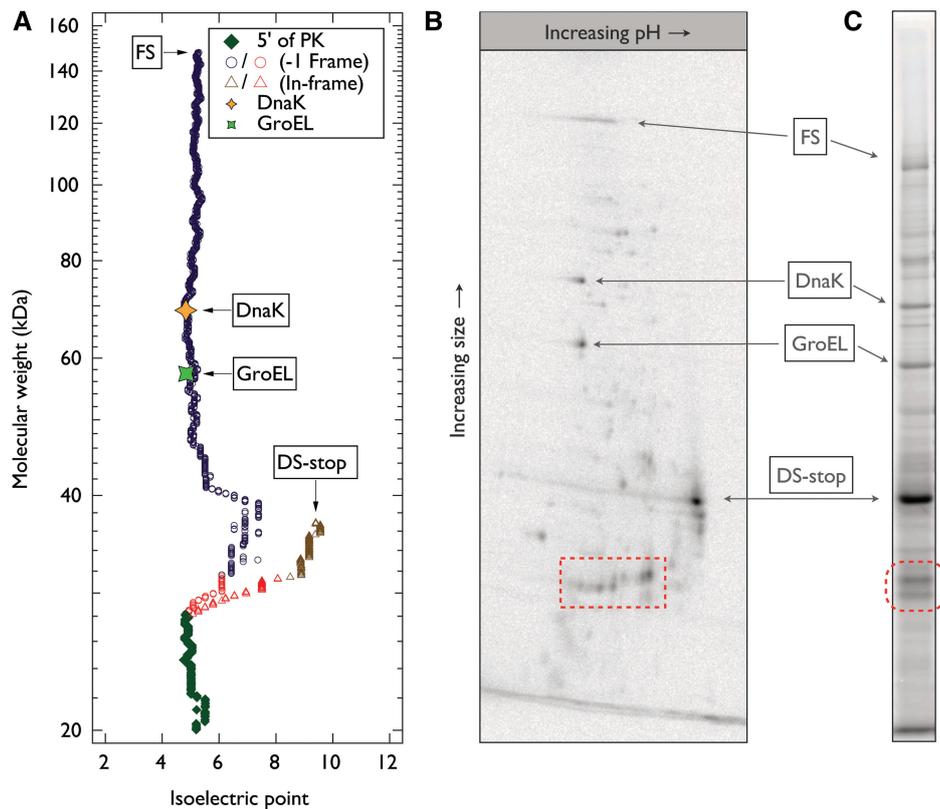
#### Quantification of ribosome stalling and correlation with stem strength

The results shown in Figure 2 revealed that a 1D SDS-PAGE assay could not firmly identify polypeptides originating from a -1 frameshifted ribosome stalled in the pseudoknot from the non-frameshifted product in a 'Downstream Stop' construct. In order to quantify the amount of -1 frameshifted ribosomes stalled inside the pseudoknot, we performed a 2D SDS-PAGE separation of the radioactively labelled proteins originating from the 'Upstream Stop' construct (Supplementary Figures S4 and S5), which is the type of construct most commonly used throughout literature. The advantage of a 2D-gel analysis is that all the unfinished protein chains with different lengths concentrate in a common spot when they have the same pI. This made it possible to identify randomly stalled translation products inside the pseudoknot sequence and we quantified the amount of radioactivity

in all identified additional spots. This produced a conservative estimate of the amount of stalled translations.

The result of quantifying the fraction of *in vivo* -1 frameshifted ribosomes, both those which made it all the way to the lacZ stop codon (gene10/lacZ fusion) and those which stalled inside the pseudoknot, is shown in Figure 3A. The hatched bars denote the -1 frameshift efficiency taking into account only the end product of -1 frameshift (148 kDa gene10/lacZ fusion). This frameshift efficiency was calculated as (intensity of FS product)/(intensity of non-FS product + intensity of FS product). The filled bars denote the -1 frameshift efficiency when both the end product (148 kDa gene10/lacZ fusion) and the products originating from stalled ribosomes are taken into account. This frameshift efficiency was calculated as (intensity of FS product + intensity of stalled product)/(intensity of non-FS product + intensity of FS product + intensity of stalled product).

In addition to the six artificial pseudoknot-like structures, we also analysed two earlier investigated pseudoknots PK400 and PK401 (25), with over-all



**Figure 2.** Stalled ribosomes. The pseudoknot used was 22/6a ‘Downstream stop’. Polypeptides produced by ribosomes stalled inside the pseudoknot are marked by red symbols. (A) Theoretically calculated size and isoelectric point of the growing polypeptide as consecutive codons are translated. Each symbol signifies the addition of a new amino acid and upstream from the pseudoknot this is indicated by green diamonds, inside the pseudoknot (in-frame) by red triangles, after the pseudoknot (in-frame) by brown triangles, inside the pseudoknot (–1 reading frame) by red circles, and after the pseudoknot (–1 reading frame) by purple circles. The expected size and isoelectric point of the completed peptides for both non-frameshifting ribosomes (DS-stop) and for –1 frameshifted ribosomes (FS) are indicated. The sizes and isoelectric points of DnaK (GenBank CAQ30531.1) and GroEL (GenBank CAQ34492.1) are indicated to provide landmarks. (B) Image of phosphor screen with L-[35S]-methionine-labelled proteins from a strain expressing 22/6a ‘Downstream Stop’ separated by 2D SDS-PAGE. (C) Same as B but only separated according to molecular weight by 1D SDS-PAGE.

structures more similar to naturally occurring pseudoknots (Figure 1 insert), inspired from structures in the infectious bronchitis virus (22,28,30). The pseudoknot structures in this type of virus are selected for their effects on vertebrate ribosomes, but the stem1 length variations were found to yield approximately the same relative stimulatory effect in *E. coli* (25) and suggest that stem1 strength is equally important for stimulating bacterial ribosomes to frameshift.

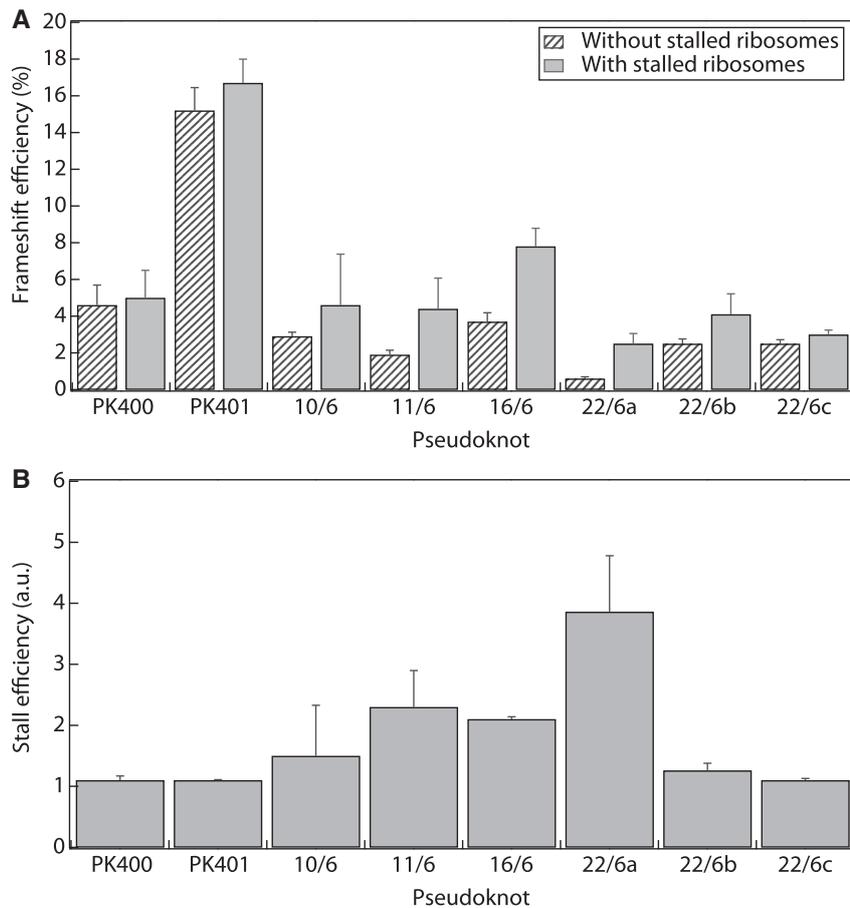
All pseudoknots investigated stalled some fraction of the frameshifted ribosomes, however, significantly more ribosomes stalled in the artificial pseudoknots than in those resembling naturally occurring pseudoknots (PK400 and PK401).

To quantify the amount of ribosomes stalling within a pseudoknot *in vivo* we calculated the ratio of (stalled + non-stalled frameshifted ribosomes) to (non-stalled frameshifted ribosomes), the result is shown in Figure 3B. For the IBV inspired pseudoknots, this ratio was close to 1 signifying that essentially no ribosomes stalled. However, the ratio was significantly larger than 1 for the more artificial pseudoknots which acted as

roadblocks for a large amount of frameshifted ribosomes. The length of stem 1 did not significantly influence on the amount of frameshifted or stalled ribosomes. Interestingly, within pseudoknots with the same overall structure (22/6a–c) 22/6a stalls a significantly higher fraction of frameshifted ribosomes than 22/6b (verified by Student’s *t*-test,  $n = 4$ ,  $\alpha = 0.05$ ,  $P = 0.012$ ), which again stalls more than 22/6c. Hence, the ability to stall a ribosome correlated with the strength of the pseudoknot base pairs, the stronger the base pairs the more frameshifted ribosomes were stalled.

### Messenger RNA pseudoknots are not a barrier to the RNA polymerase

Earlier studies have shown that the insertion of sequences able to form mRNA secondary structures into a gene may cause the RNA polymerase to stall or invoke a target for endonucleolytic attacks (19). Therefore, in our analysis of mRNA pseudoknot-stalled ribosomes, it was important to verify that there was no significant population of mRNAs that ended within the pseudoknot structure. If such truncated transcripts were abundant, it would be difficult



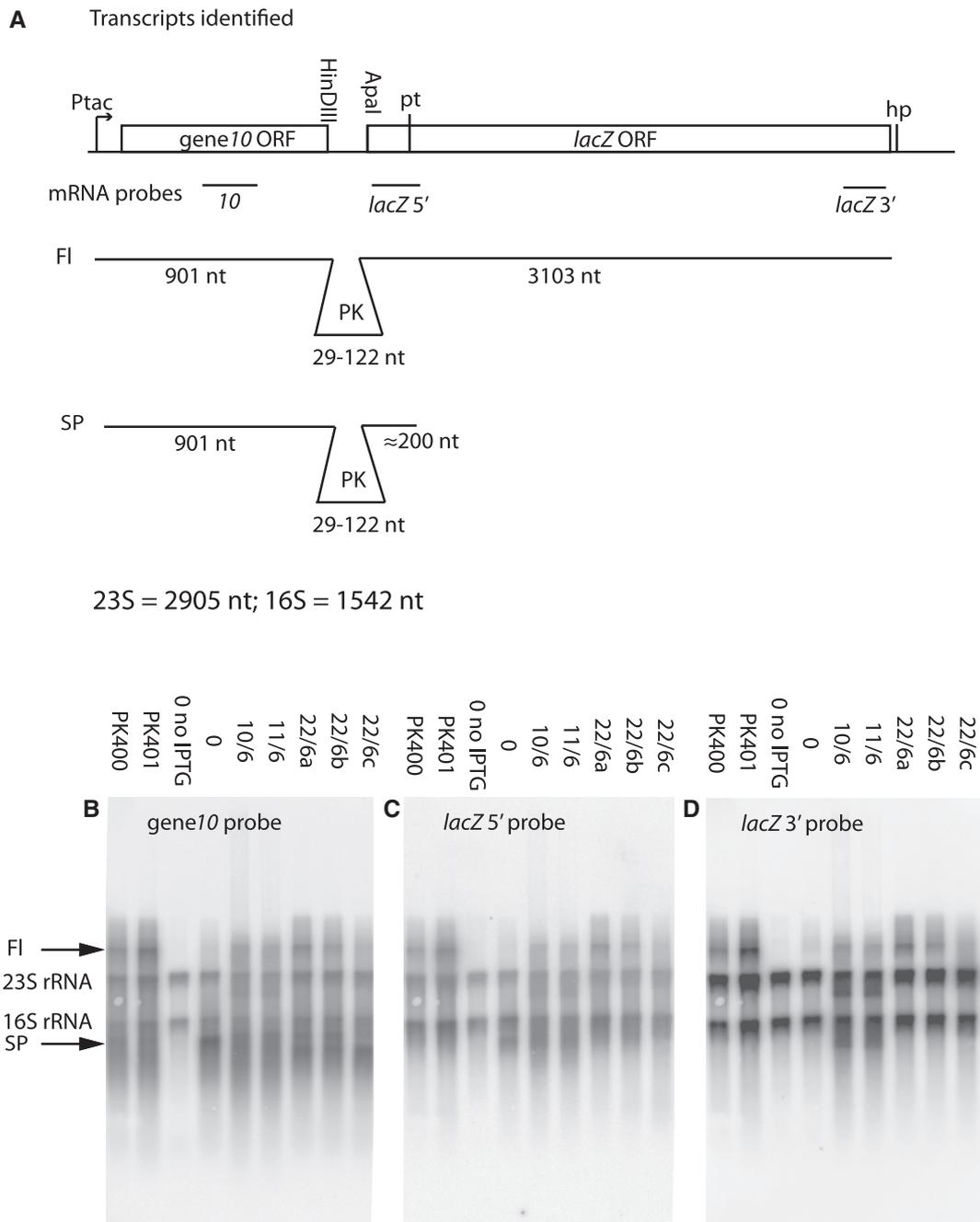
**Figure 3.** Frameshift and stall efficiency. (A) *In vivo* frameshift efficiency of different 'Upstream stop' constructs containing pseudoknots without taking peptides from stalled ribosomes into account (hatched bars) or when the peptides originating from stalled ribosomes are taken into account (filled bars). (B) Stall efficiency (i.e. ratio of all frameshifted ribosomes to non-stalled frameshifted ribosomes). Values are mean  $\pm$  SEM,  $n = 2-4$ .

to distinguish between protein products from ribosomes stalled within the pseudoknot and protein products originating from ribosomes ending translation at 'non-stop' mRNAs having their 3'-ends within the pseudoknot sequence. In the latter case translation would be terminated by tmRNA trans-translation thus rendering the protein products unstable due to the tmRNA-encoded tag (36). In the following subsections 'Identification of transcripts from the T7gene10-PK-lacZ gene fusions', 'Messenger RNA stability' and 'Coupling between translation and transcription is required for full-length transcripts', we will show that the observed proteins did indeed originate from stalled ribosomes and that they were not caused by other effects.

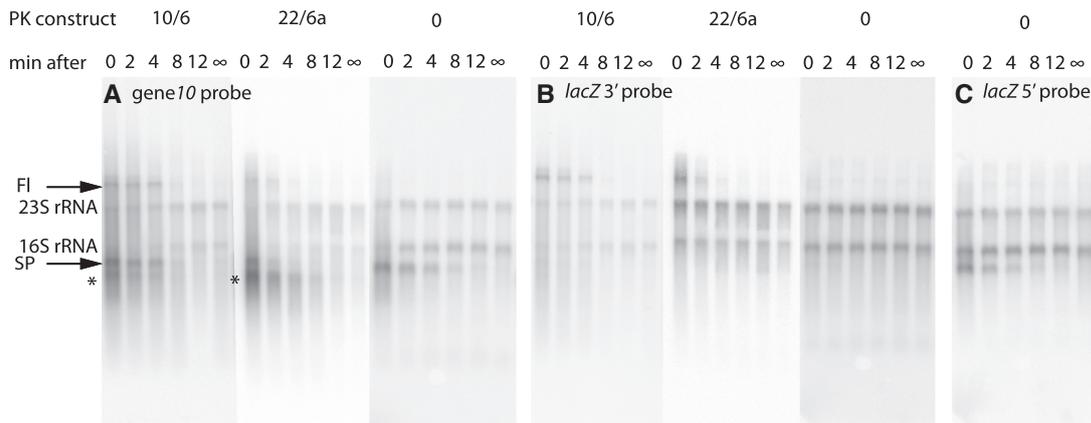
*Identification of transcripts from the T7gene10-PK-lacZ gene fusions.* To identify the major class of transcripts from our pseudoknot containing constructs, we made a northern blot with RNA from all strains used to measure frameshift frequencies, which are those containing the upstream stop. We used three different probes hybridizing either upstream of the pseudoknot, immediately downstream of the pseudoknot or in the very end of the *lacZ* reading frame (Figure 4A).

As seen in Figure 4B-D, there was an unspecific hybridization from all three probes to the 23S and 16S ribosomal RNAs. In *E. coli*, ribosomal RNA constitutes between 80% and 90% of total RNA depending on growth conditions and some cross-hybridization to these species is often seen in northern blots. Here, the uninduced culture in Figure 4B-D, lane '0 no IPTG', made it possible to estimate the unspecific probing to rRNA and the two bands were used as size markers on the blots.

Following induction with IPTG, all strains showed increased hybridization above the 23S RNA band compared to the uninduced control with all three probes. The so-called 0 construct was described in reference (25), and contains a slippery sequence and the UAA stop codon but no pseudoknot-like structure. In all strains, except the one with the 0 construct, there were a distinct band (F1) representing the expected full-length transcript. The full-length transcript reached from transcription start to the stem-loop structure downstream of the 3'-end of the *lacZ* open reading frame ('hp' in Figure 4A). This mRNA stem-loop structure has been shown to stabilize the *lacZ* transcript by reducing 3'-end exonucleolytic attacks (37). The core plasmid contained no distinct transcription termination signal after the



**Figure 4.** Transcripts from pseudoknot containing constructs. Northern blot showing transcripts from ‘Upstream stop’ constructs containing pseudoknot structures. Panel (A) Map of genes, transcripts and probes (not drawn to scale). PK: sequences of pseudoknot structures inserted between the HindIII–Apal restriction sites; FI, full-length transcript from +1 to the mRNA stabilizing hair-pin (hp) in the end of the *lacZ* open reading frame; SP, premature transcription termination fragment originating from transcription from +1 to the premature transcription termination site (pt) where RNA-polymerase terminates in cases where translating ribosomes are uncoupled from the transcribing polymerase; 23S and 16S rRNA: ribosomal RNA from the large (50S) and small (30S) ribosomal subunit, respectively. Panel (B–D) northern blots. The strains were induced by IPTG for 15 min before harvest for RNA extraction. Each lane contains 1  $\mu$ g of RNA from a strain expressing the gene construct indicated above the lane. No IPTG: no induction of  $P_{tac}$  transcription. The blot was probed with three different Ribo-probes: (B) T7 *gene10* probe covering nucleotides +476 to +676 of the induced transcript; (C) *lacZ* 5' probe covering 8–247 nt after the Apal site (approximately nucleotides +1000 to +1300 of the transcript); (D) *lacZ* 3' probe covering nucleotides +2769 to +3010 after the Apal site (probe covering upstream to the last 50 nt of the open reading frame of *lacZ*).



**Figure 5.** mRNA stability. Northern blots showing chemical stability of IPTG inducible transcripts from three PK containing plasmids all having the upstream stop UAA codon. The three strains were induced by IPTG for 15 min, then IPTG was removed by filtration at 37°C and aliquots were harvested for RNA extraction. Each lane contains 1 µg of RNA from the strain indicated above each blot (containing six lanes). The time of harvest after IPTG was removed is indicated above each lane; 0: sample harvested before filtration; ∞: sample harvested either 16 min after (10/6) or before induction (0 and 22/6a). \* indicates an RNA band discussed in text. Other symbols: see caption of Figure 4.

*lacZ* gene, and accordingly we found transcripts that exceeded far beyond the full-length FI band (Figure 4B–D).

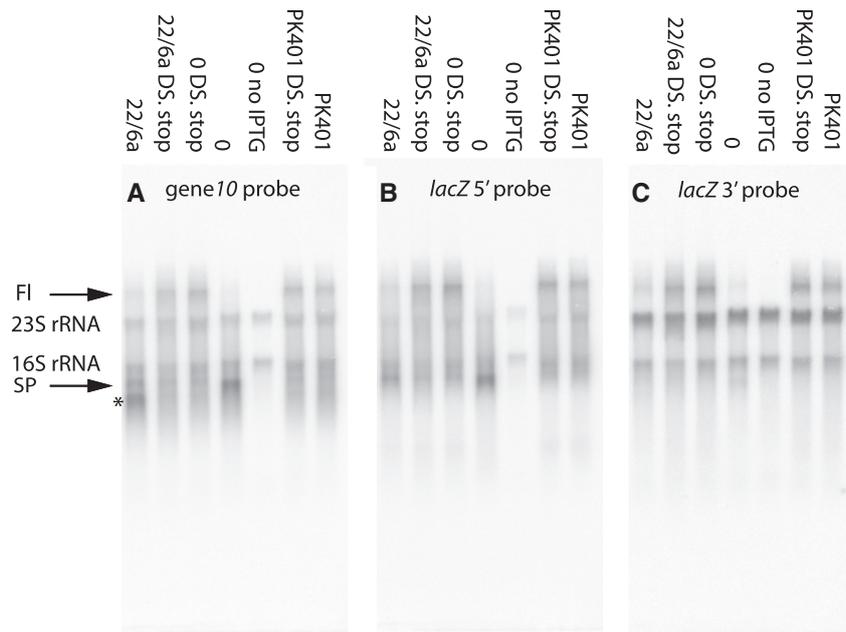
In the beginning of *lacZ*, ~200 nt into the open reading frame, there is a site, called ‘pt’ in Figure 4A, where the RNA polymerase is caused to terminate if there is inefficient translation initiation of the *lacZ* gene (38). In the 0 construct there is no pseudoknot to stimulate frameshift at the slippery site. Therefore, virtually no ribosomes were expected to follow the RNA polymerase from gene 10 into the *lacZ* part of our gene fusion. As expected, Figure 4B and C, lane ‘0’ shows a prominent band (‘SP’ for stop polymerase) corresponding in size and probe-ability to this premature termination product. Also, corresponding low amounts of high molecular weight transcripts are detected for this construct. All the other constructs shown in Figure 4 contained frameshift stimulating pseudoknots and an inspection of the northern blot showed that the ‘SP’ bands probed with both *gene10* and *lacZ5'* sequences were present in sizes which correspond to the sizes of the pseudoknots inserted.

**Messenger RNA stability.** The wild type *lacZ* mRNA half-life is close to the average mRNA half-life in *E. coli* (120 s) and transcription takes close to 80 s due to the length of the *lacZ* gene (three times longer than the average gene). Therefore, a northern blot of wild type *lacZ* mRNA under steady state transcription will always include a lot of unfinished native transcripts, as well as mRNAs under degradation. Here, our *gene10-lacZ* fusion was even longer and transcription should take ~120 s. Accordingly, all induced strains included in Figure 4 show a distinct smear of mRNA fragments recognized by all three probes. In order to examine the half-life of our artificial transcripts, we made experiments where transcription from the  $P_{tac}$  promoter was stopped due to removal of the inducer (Figure 5). Two minutes after IPTG removal, any remaining smear should originate from mRNA degradation because most of the RNA polymerase should have reached the end of the gene fusion at this time.

From the experiment, shown in Figure 5, it is evident that both the ‘FI’ and the ‘SP’ mRNA fragments had a half-life close to the average 2 min *E. coli* mRNA half-life. In addition, both the pseudoknot containing constructs (10/6 and 22/6a) revealed the existence of a short mRNA fragment that was recognized only by the *gene10* probe but not the *lacZ5'* and 3' probes (indicated by ‘asterisks’ in Figure 5). This fragment includes the transcription start in the 5'-end and the pseudoknot in the 3'-end. We suggest that the pseudoknot acts as an exonuclease barrier like the natural stem-loop structure in the 3'-end of the wild type *lacZ* transcript (37) and thereby induces a degradation intermediate of a distinct length with increased half life compared to unstructured mRNA sequences like those from construct 0. Alternatively, but not mutually exclusive, a pseudoknot acts like a rho-independent termination signal to the RNA polymerase. However, the sequences were not followed by a row of uridine residues, which would be necessary to make a stem-loop structure into a functional transcription terminator.

**Coupling between translation and transcription is required for full-length transcripts.** The final test of our model for the transcription pattern in our artificial gene fusion was to establish translational coupling beyond the slippery sequence and into the polar termination site (SP) in *lacZ*. By changing the upstream stop codon between the slippery site and the pseudoknot region into a sense codon ribosomes should, frameshifted or not, follow the RNA polymerase into the beginning of the *lacZ* sequence.

The 22/6a and the 0 constructs are the two constructs with the lowest frequency of frameshifting. Therefore, they have the least ribosome traffic into the *lacZ* sequence. Alteration of the UAA stop codon into a lysine AAA codon in the spacer between the slippery sequence and the pseudoknot changed the pattern of transcripts immensely. These two downstream stop variants (‘DS. stop’ in Figure 6), which did not contain a stop codon upstream from the structure, expressed significantly more full-length (‘FI’) transcript and only insignificant



**Figure 6.** Transcriptional polarity in the *lacZ* gene. Northern blot comparing transcripts from ‘Upstream stop’ and ‘Downstream stop’ (DS. stop) constructs. The strains were induced by IPTG for 15 min before harvest for RNA extraction. Each lane contains 1  $\mu$ g of RNA from the strain indicated above the lane. No IPTG: no induction of  $P_{lac}$  transcription. Other symbols: see caption of Figures 4 and 5. In PK401 DS. stop, the nucleotide T encoding the most 3' U in Loop 2 (Fig. 1) was mutated into an A to avoid an internal stop codon.

amounts of premature transcription stop fragment (‘SP’) compared to their sister constructs containing the UAA stop codon upstream from the structure (Figure 6). Our control construct, PK401, which stimulated 14% frameshift, showed no premature transcription stop fragment (‘SP’) and therefore no change in transcription pattern was observed as a consequence of removing the upstream UAA stop codon (Figure 6) thus confirming that the major effect causing the ‘SP’ fragment is polarity in the *lacZ* gene and not transcription termination caused by the pseudoknot sequences.

Also, the very short band marked by ‘asterisks’ that appeared from the 22/6a construct was not present in the ‘DS. stop’ variant (Figure 6). This exclude this mRNA fragment to be causal for the appearance of stalled protein products, because 22/6a (‘DS. stop’) is the construct that caused the highest frequency of stalling (compare Figure 2 and Supplementary Figure S5).

Our conclusion is that the stable proteins observed from within the pseudoknot structures (Figure 2, Supplementary Figure S1, S2, S4 and S5) were products from stalled ribosomes. The stalling of the ribosomes was directly caused by the tertiary structure and not by some secondary effect, as, e.g. stop codon-less mRNA fragments ending within the structure sequences.

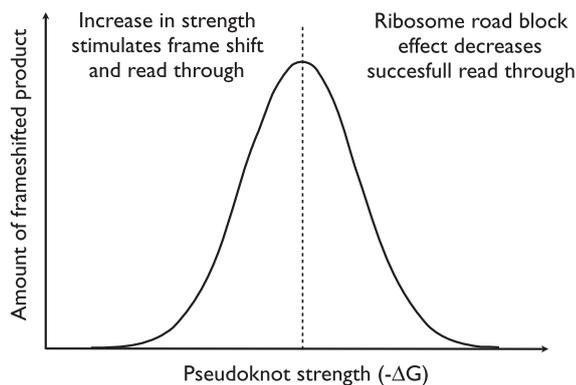
## DISCUSSION

The structures analysed in this study are artificial and were designed to fold into pseudoknot-like structures with a gradually increasing mechanical strength. The mechanical strength was adjusted by changing the base pairs of the two stems, which seems to be a reasonable way of crudely

varying the mechanical strength, as the energy involved in base pairing is higher than the energies involved in, e.g. the electrostatic interaction of the loop with the stems. It is, however, likely that the loop–stem interaction, surface charges or other players than just mechanical strength influence frameshift stimulating effect of mRNA structures. As there is a consensus in recent literature that pseudoknot mechanical strength correlates with frameshifting efficiency (23–25), it was intriguing that the amount of frameshifted product was reduced by the stronger pseudoknot 22/6a compared to the weaker 22/6b or c (Figure 3A). This proved to be caused by stalling of a significant amount of frameshifted ribosomes by the strong pseudoknots (Figure 3B). Future studies will show whether significant stalling can also be caused by naturally occurring pseudoknots.

Quantitative northern blot analysis was used to examine whether the observed translation products ending within the pseudoknot structure arose from fragments of mRNA produced either by low RNA–polymerase processivity or specific endonucleolytic attacks by RNases at the pseudoknot sequences. No evidence was found of a specific population of transcripts that could explain the amounts of protein products attributed to originate from pseudoknot-stalled ribosomes. Also, our protein-stability assay showed that the translational products from the stalled ribosomes were stable for at least 80 min (Supplementary Figure S1), thus indicating that the stalled ribosomes are not rescued by tmRNA and that the stalled proteins do not originate from truncated mRNA.

We also checked whether the protein products from within the pseudoknot structure could arise from very



**Figure 7.** Model of frameshifting efficiency. Increasing the strength of a pseudoknot causes the pseudoknot to induce frameshifting at a higher frequency. However, the stronger the pseudoknot the larger the likelihood that it will act as a roadblock for the ribosome, reducing the amount of frameshifted product produced. The optimal frameshifting efficiency is achieved by balancing the two contributions.

slow rather than permanently stalled ribosomes. A pulse chase experiment (Supplementary Figure S2) revealed that within 16 min there was no sign of a redistribution of label between the stalled spots and the stop codon-terminated downstream stop product, thus proving the possibility of very slow ribosomes to be unlikely.

It is possible that the newly discovered ribosome rescue factor, ArfA (39) could be active at pseudoknot-stalled ribosomes and that nascent proteins would be more stable than if saved by tmRNA. However, as can be seen in Supplementary Figure S3, the growth of strains expressing pseudoknot 22/6a was severely affected by induction and showed a decrease in growth rate correlating to the amount of stall product observed. Because ribosomes are limiting in growing cells (40), the sequestration of ribosomes by engagement in induced overexpression of a gene from a plasmid will often cause a strain to grow slower than the uninduced counterpart. The enhanced reduction in growth rate upon induction of 22/6a compared to the 0 construct (Supplementary Figure S3) could indicate that stalled ribosomes were not rescued at a sufficiently high rate and we suggest that either the ribosomal rescue systems were titrated by the large amount of mRNA induced from the plasmid alleles, or alternatively, that no rescue is possible for pseudoknot-stalled ribosomes.

Our results are in agreement with the observation that the amount of protein produced from an mRNA can be reduced when a pseudoknot is located upstream (29). Also, they provide a possible explanation for the reduction in frameshift efficiency observed by, e.g. Napthine *et al.* (30) when increasing the thermodynamic stability of stem 1 above a certain threshold. This apparent reduction in frameshift efficiency (observed by 1D SDS-PAGE) could be caused by the fact that a significant fraction of the 'frameshifted' ribosomes permanently stalled within the pseudoknot.

We propose that pseudoknot induced frameshifting efficiency can be viewed as a balance between two

competing effects (as visualized in Figure 7), the mechanically stronger the pseudoknot, the larger the frameshifting efficiency (25–27), however, the stronger the pseudoknot the larger the likelihood of stalling the frameshifted ribosome, thus preventing the translation of full-length frameshift product. Possibly, evolution optimized viral pseudoknots to balance these two effects. Hence, in measurements of frameshifting efficiency it is important to take into account the roadblocking effect of mRNA pseudoknots.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank M. Warrer for excellent technical support. J.T., L.B.O. and M.A.S. designed research, analysed data and wrote paper. J.T. and M.A.S. performed research.

## FUNDING

Funding for open access charge: University of Copenhagen Excellence Program.

*Conflict of interest statement.* None declared.

## REFERENCES

- Brierley, I., Gilbert, R.J.C. and Pennell, S. (2010) Pseudoknot-Dependent Programmed-1 Ribosomal Frameshifting: Structures, Mechanisms and Models. In Atkins, J.F. and Gesteland, R.F. (eds), *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, Vol. 24. Springer, New York/Dordrecht/Heidelberg/London, pp. 149–174.
- Farabaugh, P.J. (2010) Programmed Frameshifting in Budding Yeast. In Atkins, J.F. and Gesteland, R.F. (eds), *Recoding: Expansion of Decoding Rules enriches Gene Expression*, Vol. 24. Springer, New York/Dordrecht/Heidelberg/London, pp. 221–247.
- Fayet, O. and Prère, M.-F. (2010) Programmed Ribosomal-1 Frameshifting as a Tradition: The Bacterial Transposable Elements of the IS3 Family. In Atkins, J.F. and Gesteland, R.F. (eds), *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, Vol. 24. Springer, New York/Dordrecht/Heidelberg/London, pp. 259–280.
- Brierley, I. and Dos Ramos, F.J. (2006) Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.*, **119**, 29–42.
- Firth, A.E. and Atkins, J.F. (2009) A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. *Viol. J.*, **6**, 14.
- Shehu-Xhilaga, M., Crowe, S.M. and Mak, J. (2001) Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol.*, **75**, 1834–1841.
- Dinman, J.D. and Wickner, R.B. (1992) Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J. Virol.*, **66**, 3669–3676.
- Giedroc, D.P. and Cornish, P.V. (2009) Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.*, **139**, 193–208.
- Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J. and Brierley, I. (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, **441**, 244–247.

10. Pallan, P.S., Marshall, W.S., Harp, J., Jewett, F.C., Wawrzak, Z., Brown, B.A., Rich, A. and Egli, M. (2005) Crystal Structure of a Luteoviral RNA Pseudoknot and Model for a Minimal Ribosomal Frameshifting Motif. *Biochemistry*, **44**, 11315–11322.
11. Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F. and Gesteland, R.F. (1988) Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *EMBO J.*, **7**, 1503–1507.
12. Weiss, R.B., Dunn, D.M., Shuh, M., Atkins, J.F. and Gesteland, R.F. (1989) *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol.*, **1**, 159–169.
13. Garcia, A., van Duin, J. and Pleij, C.W.A. (1993) Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. *Nucleic Acids Res.*, **21**, 401–406.
14. Ivanov, I.P., Gesteland, R.F., Matsufuji, S. and Atkins, J.F. (1998) Programmed frameshifting in the synthesis of mammalian antizyme is +1 in mammals, predominantly +1 in fission yeast, but -2 in budding yeast. *RNA*, **4**, 1230–1238.
15. Kozak, M. (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol. Cell. Biol.*, **9**, 5134–5142.
16. Hall, M.N., Gabay, J., Debarbouille, M. and Schwartz, M. (1982) A role for mRNA secondary structure in the control of translation initiation. *Nature*, **295**, 616–618.
17. Von Heijne, G., Nilsson, L. and Blomberg, C. (1977) Translation and messenger RNA secondary structure. *J. Theor. Biol.*, **68**, 321–329.
18. Qu, X., Wen, J.-D., Lancaster, L., Noller, H.F., Bustamante, C. and Tinoco, I. (2011) The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature*, **475**, 118–121.
19. Sørensen, M.A., Kurland, C.G. and Pedersen, S. (1989) Codon usage determines the translation rate in *Escherichia coli*. *J. Mol. Biol.*, **207**, 365–377.
20. Tu, C., Tzeng, T.H. and Bruenn, J.A. (1992) Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc. Natl Acad. Sci. USA*, **89**, 8636–8640.
21. Somogyi, P., Jenner, A.J., Brierley, I. and Inglis, S.C. (1993) Ribosomal pausing during translation of an RNA pseudoknot. *Mol. Cell. Biol.*, **13**, 6931–6940.
22. Kontos, H., Napthine, S. and Brierley, I. (2001) Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol. Cell. Biol.*, **21**, 8657–8670.
23. Plant, E.P. and Dinman, J.D. (2005) Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic Acids Res.*, **33**, 1825–1833.
24. Ito, K., Chiba, S. and Pogliano, K. (2010) Divergent stalling sequences sense and control cellular physiology. *Biochem. Biophys. Res. Commun.*, **393**, 1–5.
25. Hansen, T.M., Reihani, S.N., Oddershede, L.B. and Sørensen, M.A. (2007) Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proc. Natl Acad. Sci. USA*, **104**, 5830–5835.
26. Chen, G., Chang, K.Y., Chou, M.Y., Bustamante, C. and Tinoco, I. Jr (2009) Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc. Natl Acad. Sci. USA*, **106**, 12706–12711.
27. Green, L., Kim, C.-H., Bustamante, C. and Tinoco, I. Jr (2008) Characterization of the Mechanical Unfolding of RNA Pseudoknots. *J. Mol. Biol.*, **375**, 511–528.
28. Brierley, I., Digard, P. and Inglis, S.C. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell*, **57**, 537–547.
29. Plant, E.P., Rakauskaitė, R., Taylor, D.R. and Dinman, J.D. (2010) Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *J. Virol.*, **84**, 4330–4340.
30. Napthine, S., Liphardt, J., Bloys, A., Routledge, S. and Brierley, I. (1999) The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *J. Mol. Biol.*, **288**, 305–320.
31. Liphardt, J., Napthine, S., Kontos, H. and Brierley, I. (1999) Evidence for an RNA pseudoknot loop-helix interaction essential for efficient -1 ribosomal frameshifting. *J. Mol. Biol.*, **288**, 321–335.
32. Neidhardt, F.C., Bloch, P.L. and Smith, D.F. (1974) Culture medium for enterobacteria. *J. Bacteriol.*, **119**, 736–747.
33. Reeder, J., Steffen, P. and Giegerich, R. (2007) pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic Acids Res.*, **35**, W320–W324.
34. O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, **250**, 4007–4021.
35. Sørensen, M.A. and Pedersen, S. (1997) Determination of the Peptide Elongation rate In Vivo. In Martin, R. (ed.), *Methods in Molecular Biology: Protein Synthesis: Methods and protocols*, Vol. 77. The Humana Press, Inc., Totowa, New Jersey, NJ, USA, pp. 129–142.
36. Farrell, C.M., Grossman, A.D. and Sauer, R.T. (2005) Cytoplasmic degradation of ssrA-tagged proteins. *Mol. Microbiol.*, **57**, 1750–1761.
37. Cannistraro, V.J., Subbarao, M.N. and Kennell, D. (1986) Specific endonucleolytic cleavage sites for decay of *Escherichia coli* mRNA. *J. Mol. Biol.*, **192**, 257–274.
38. Stanssens, P., Remaut, E. and Fiers, W. (1986) Inefficient translation initiation causes premature transcription termination in the lacZ gene. *Cell*, **44**, 711–718.
39. Chadani, Y., Ono, K., Ozawa, S., Takahashi, Y., Takai, K., Nanamiya, H., Tozawa, Y., Kutsukake, K. and Abo, T. (2010) Ribosome rescue by *Escherichia coli* ArfA (YhdL) in the absence of trans-translation system. *Mol. Microbiol.*, **78**, 796–808.
40. Vind, J., Sørensen, M.A., Rasmussen, M.D. and Pedersen, S. (1993) Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes: expression from reporter genes does not always reflect functional mRNA levels. *J. Mol. Biol.*, **231**, 678–688.

## Supplementary information to Tholstrup *et al.*

### Supporting text

#### Ribosome chase assay.

In order to examine if the stalled protein products were caused by paused ribosomes, we performed a pulse chase experiment. Very slow translation would cause radioactivity to disappear, as a function of time, from the stalled products and eventually to appear in the stop codon terminated protein product. The results from such a pulse chase experiment (Fig. S2) showed that the amount of <sup>35</sup>S-Met incorporated during the pulse (20 s) appeared in the final protein spots after a short translation time between the first (10 sec. chase) and the second sampling point (60 sec. chase) and stayed stable for at least 900 sec.

#### Two-dimensional SDS-PAGE – additional gels

The results in figure 3 are based on two-dimensional SDS-PAGE of proteins from constructs that contain a UAA stop codon immediately upstream from the pseudoknot structure. Peptides visible only in IPTG induced cultures showing the expected molecular weight and pI were used in the quantification frameshift in addition to the stop product and full-length frameshift product. Figure S4 below is an example of such a gel pair (induced and un-induced) for pseudoknot 22/6a. In panel B the in-frame stop product (Stop) is clearly visible as are the frameshift product (FS), GroEL and DnaK. The region containing peptides originating from ribosomes that are stalled within the pseudoknot is indicated with a red dashed rectangle.

In addition, Figure S5 shows examples of two dimensional SDS-PAGE representing all the different constructs that were analyzed quantitatively in Figure 3. The 0 construct (Fig. S5) has no pseudoknot sequence inserted and was used as the background control. 21

## Supplementary Materials and Methods

**Ribosome chase assay.** 5 ml of an exponentially growing culture in MOPS media was induced with 1 mM IPTG ( $t = 0$ ). At  $t = 2$  min  $80 \mu\text{Ci L-[4,5-}^3\text{H(N)]lysine}$  was added. At  $t = 10$  min  $610 \mu\text{g L-lysine}$  was added as a lysine chase. At  $t = 15$  min  $50 \mu\text{Ci L-[}^{35}\text{S]-methionine}$  was added. At  $t = 15$  min 20 sec  $500 \mu\text{g L-methionine}$  was added as a methionine chase. At the indicated time points aliquots of the culture were transferred to tubes on ice containing chloramphenicol to a final concentration of 2.5 mg/ml. Cells were harvested and proteins were separated by two-dimensional SDS-PAGE as described in Materials and Methods in the main text. The gels were dried and left to expose an X-ray film (Kodak). Relevant proteins samples were extracted from the gel using a hollow needle and the isotope ratio was determined in a liquid scintillation counter (PerkinElmer, Winspectral 1414).

**Protein stability assay.** An exponentially growing culture in MOPS media was induced with 1 mM IPTG (final) for 15 min at  $\text{OD}_{436} \sim 0.6$ . Two 5 ml aliquots of the induced culture were extracted (culture A and B hereafter). Culture A was labeled with  $20 \mu\text{Ci L-[}^{35}\text{S]-methionine}$  for 30 sec after which  $500 \mu\text{g L-methionine}$  was added as a methionine chase. At the indicated time points aliquots were extracted to tubes at  $0^\circ\text{C}$  containing chloramphenicol to a final concentration of 2.5 mg/ml. Culture B was labeled with  $200 \mu\text{Ci L-[4,5-}^3\text{H(N)]-lysine}$  for 30 min and transferred to tubes at  $0^\circ\text{C}$  containing chloramphenicol to a final concentration of 2.5 mg/ml. Equal amounts of culture B was added to the aliquots of culture A. The mixed aliquots were harvested and proteins were separated on two-dimensional SDS-PAGE as described in Materials and Methods in the main text. The isotope ratio of relevant proteins was investigated as described for the ribosome chase assay. The isotope ratio in the individual protein samples was divided by the isotope ratio of TCA precipitated total protein for each time-point.

## Supplementary figure legends

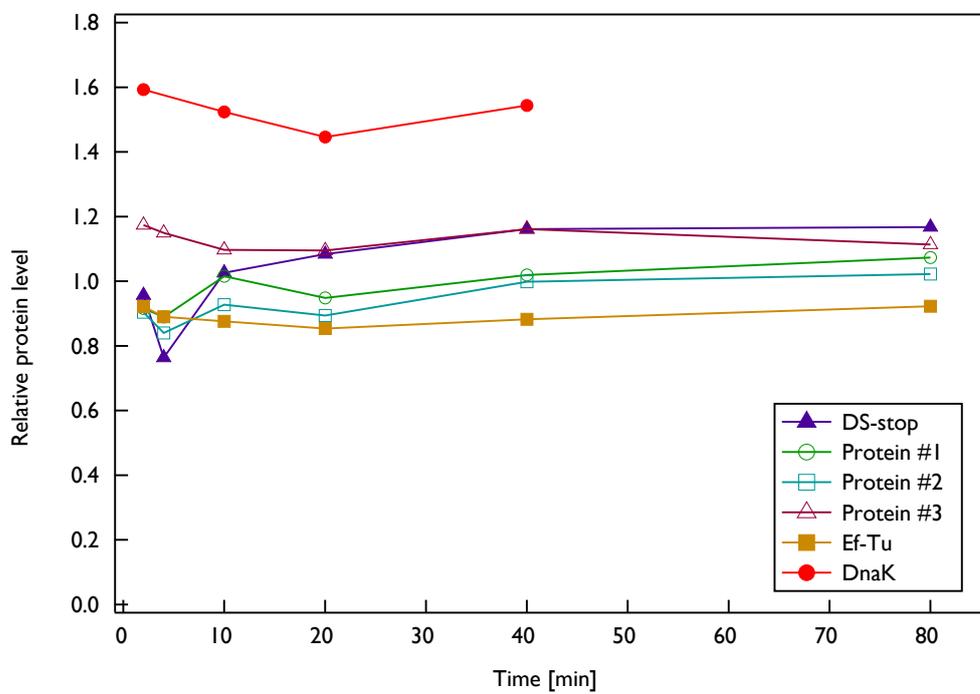
**Figure S1: Result from protein stability assay on pseudoknot 22/6a with downstream stop.** Relative expression levels of proteins determined as  $^3\text{H}/^{35}\text{S}$  in specific protein spots divided by the  $^3\text{H}/^{35}\text{S}$  ratio in total protein (TCA precipitable material) as a function of time in the methionine chase. The specific proteins are the same as those in Figure S2 with the addition of elongation factor Tu (Ef-Tu).

**Figure S2: Result from ribosome chase assay on pseudoknot 22/6a with downstream stop.** **A)** Two-dimensional SDS-PAGE of radioactively labeled proteins harvested at time point 980 sec. The proteins selected for analysis are indicated. For each time point a two-dimensional SDS-PAGE was made and the identical proteins were extracted from the gels and the ratio of  $^{35}\text{S}$ -methionine to  $^3\text{H}$ -lysine was determined using a liquid scintillation counter. The selected proteins were DnaK (control), in-frame termination (“Downstream stop”) and three polypeptides produced by stalled ribosomes (“Protein #1”, “Protein #2” and “Protein #3”). **B)** The isotope ratio for the selected proteins indicated in (A) as a function of time in the methionine chase.

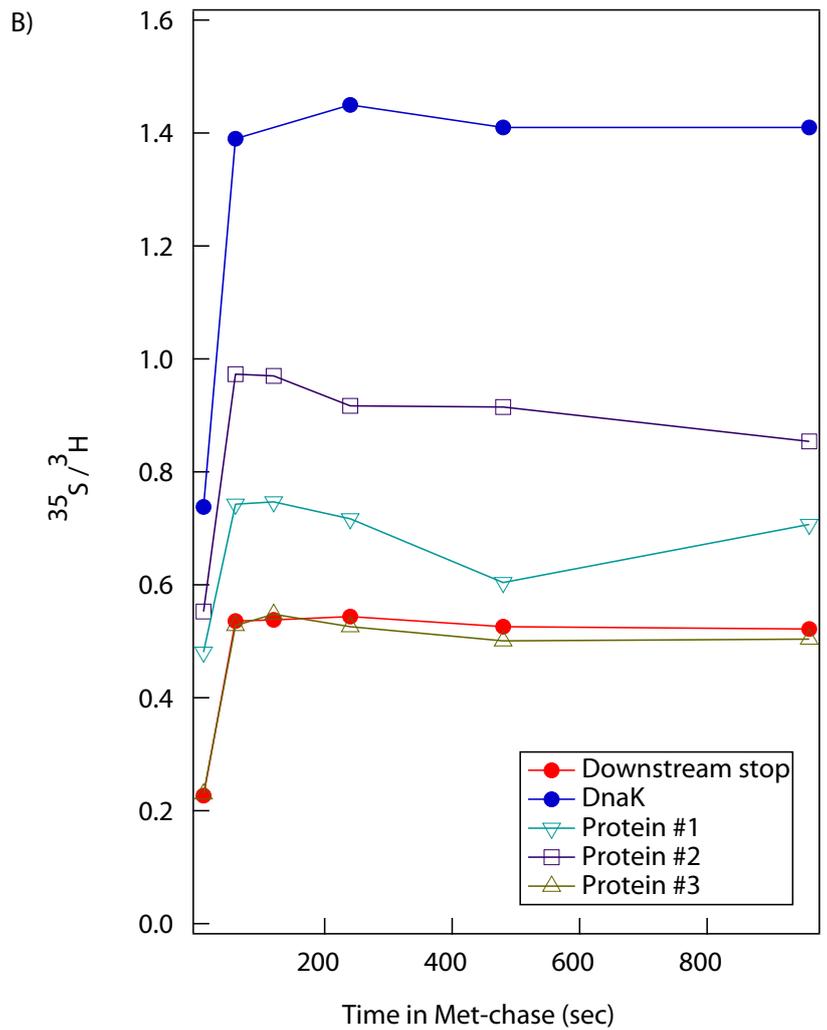
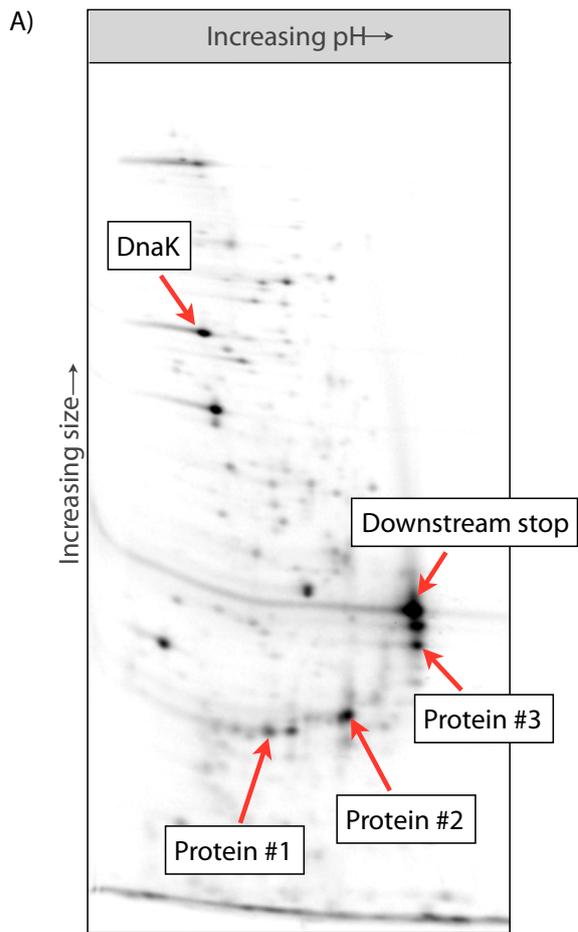
**Figure S3: Growth curves for strains containing pseudoknots after induction.** Exponentially growing cultures were diluted to  $0.001 \text{ OD}_{436}$  and allowed to grow to  $\text{OD}_{436} \approx 0.06$  where IPTG was added (time point zero) to induce expression. See insert for construct description.

**Figure S4: Two-dimensional SDS-PAGE used to estimate true frameshift efficiency.** **A:** Theoretical estimation of the molecular weight and pI of peptides produced as translation progresses. GroEL and DnaK is shown as landmarks to ease orientation on experimental gels in panel B and C. **B:** Image of exposed PhosphorImage screen for pseudoknot 22/6a containing a UAA-stop codon immediately upstream from the pseudoknot. GroEL and DnaK is shown to help orientation with panel A, in-frame stop product (Stop) and frameshift product is indicated (FS). Red rectangle indicate region of interest in relation to stalled ribosomes. **C:** Image of PhosphorImage screen for an uninduced culture containing pseudoknot 22/6a. **D:** SDS-PAGE of an uninduced (-) or IPTG induced (+) strain carrying PK 22/6a. Red dashed rectangle denote region of interest in relation to stalled ribosomes.

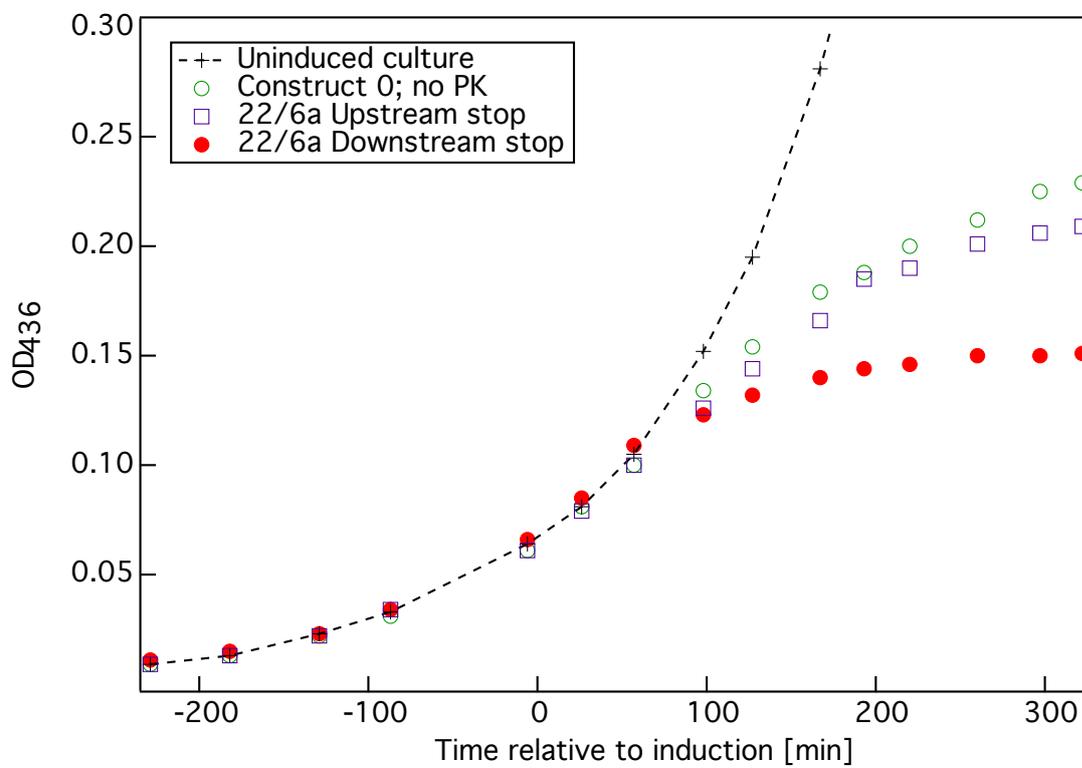
**Figure S5: Two-dimensional SDS-PAGE representing all strains used to estimate frameshift efficiency.** All strains contained a UAA-stop codon immediately upstream from the pseudoknot. Images are of exposed PhosphorImage screens for the pseudoknots indicated above each top panel. Each of the lower panels show an enlargement (some are slightly distorted to fit in the panel) of the area indicated by a red dashed rectangle in the top panel right above and denotes the region of interest in relation to stalled ribosomes. The 0 construct ('0 No PK'), which only contain the slippery sequence and the UAA stop codon is indicated by a red rectangle.



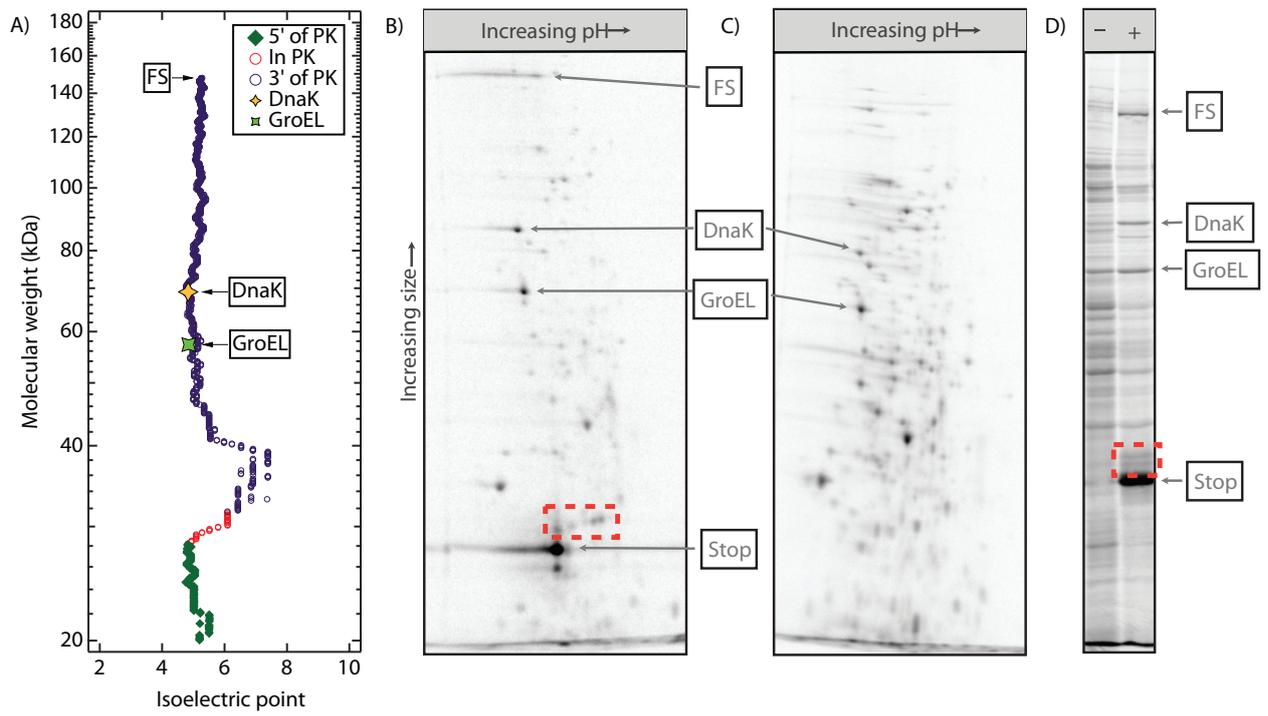
Supplementary figure S1; Tholstrup *et al.*



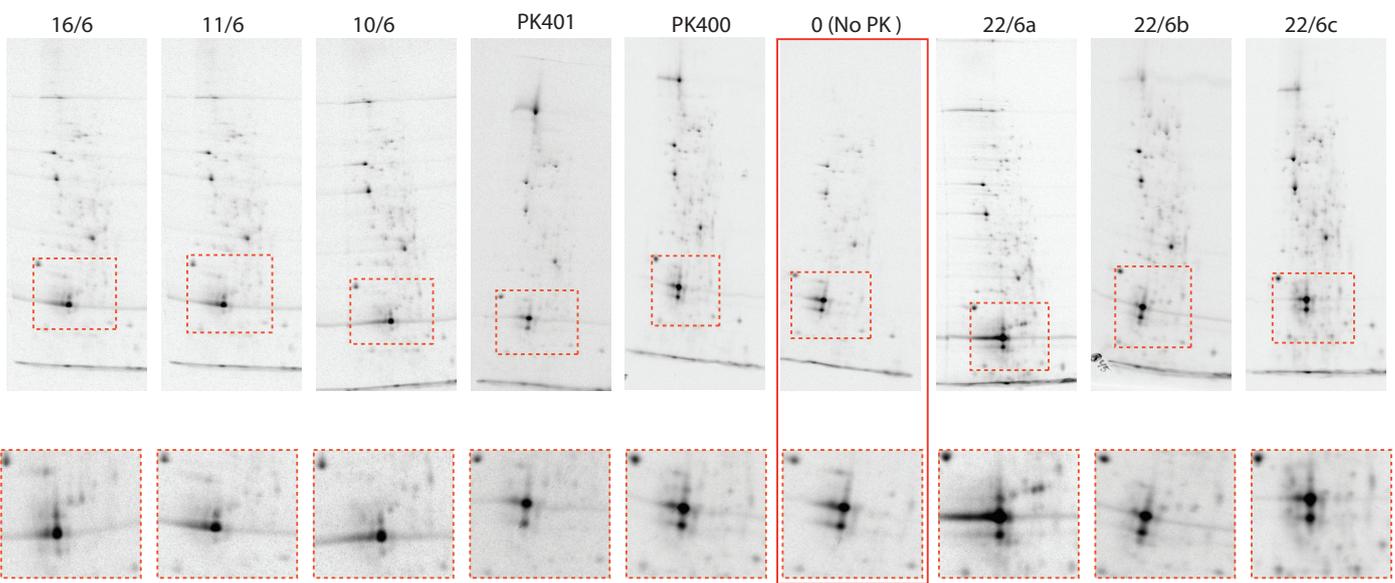
Supplementary figure S2; Tholstrup *et al.*



Supplementary figure S3; Tholstrup *et al.*



Supplementary figure S4; Tholstrup *et al.*



Supplementary figure S5; Tholstrup *et al.*



## Part II

# The Putative Ribosomal Helicase

## 3 Introduction and Objectives

The process of translation involves decoding the sequence of nucleotides on a messenger RNA into a sequence of amino acids. This requires that the mRNA is presented to the decoding center of the ribosome in a single stranded conformation, allowing proper codon tRNA-anticodon recognition by the ribosome. The propensity of polynucleotides to form double stranded structures is thus a barrier which must be overcome during the process of translation. Experiments have indicated that approximately 40% of nucleotides on a random sequence of RNA are base paired in helical structures, indicating that some component of the translational apparatus must be able to overcome this barrier and force the RNA into a single stranded conformation prior to decoding by the ribosome [Doty et al., 1959].

Although the existence of helicase proteins, which use the energy of phosphate hydrolysis to drive the separation of duplex oligonucleotides, are known (reviewed in [Caruthers and McKay, 2002]), this introduction will consider a more recent suggestion. In 2001 Yusupova et al. proposed, based on the crystal structure of the 70S ribosome from *Thermus thermophilus* (*T. thermophilus*), that the ribosome itself was able to separate duplex structures on the template mRNA [Yusupova et al., 2001]. The suggestion was based on the observation that the mRNA entrance tunnel is too narrow to accommodate an RNA helix and that the entrance tunnel, which is defined by three ribosomal proteins (S3, S4, and S5), is lined with positively charged amino acid residues. These three proteins are located on two different domains of the 30S subunit. Ribosomal protein S3 is part of the “head” of the 30S subunit while both ribosomal protein S4 and S5 are part of the “body” - these two domains have been shown to move relative to each other during EF-G-GTP accommodation and subsequent GTP hydrolysis [Yusupova et al., 2001, Frank and Agrawal, 2000]. Yusupova et al. suggested that while the positive amino acid residues bind to the negative phosphate backbone of the RNA, the relative movement of the proteins would force the strands of any double stranded RNA apart, thus providing helicase activity to the ribosome itself [Yusupova et al., 2001]. Using a rather crude *in vitro* strand-displacement assay, Takyar et al. later showed that reconstituted *E. coli* ribosomes were able to exercise helicase activity in the absence of other known helicases [Takyar et al., 2005]. Furthermore, Takyar et al. showed that the ribosomal helicase activity could be compromised by substitution mutations of positive amino acid residues in S3 (R131A, R132A, and K135A) and S4 (R44A and R47A) but not in S5 (R19A and R28A) [Takyar et al., 2005]. These findings have been supported by computational analysis of various crystal structures of the *T. thermophilus* ribosome, which indicate that R131, R132, K135, and R164 from S3; and R47, R49, and R50 from S4, all maintain a close proximity with the mRNA under dynamic conditions [Kurkcuoglu et al., 2008].

The putative ribosomal helicase is interesting in the context of programmed frameshift, as the ability of pseudoknots to induce frameshift could originate in a structural mismatch

between the pseudoknot and the ribosomal helicase and/or limited rotational freedom during the untying of pseudoknots [Yusupova et al., 2001, Plant and Dinman, 2005]. Using cryo-EM, Namy et al. was able to visualize a rabbit 80S ribosome while stalled upon encountering the minimal IBV pseudoknot [Namy et al., 2006, Brierley et al., 1991]. Their findings indicated that the pseudoknot partially deformed the entrance tunnel in the region of ribosomal protein rpS3 (S3 in prokaryotes) and rpS2 (S5 in prokaryotes). This could indicate that the putative helicase is unable to effectively dissolve mRNA pseudoknots explaining why pseudoknots are frequently employed to induce programmed frameshift. It is also interesting to investigate how the structure is unfolded by the ribosome, as models of -1 PRF presented thus far, simply consider the RNA structure as an entity which dissolves after the frameshift event [Léger et al., 2007, Weiss et al., 1989, Jacks et al., 1988a, Plant et al., 2003].

To investigate the effect of mutations in the putative ribosomal helicase on pseudoknot induced -1 frameshift *in vivo*, a number of chromosomal mutations was constructed in which lysine or arginine residues in the mRNA entrance tunnel were substituted with alanine residues. Bacterial strains carrying these mutations were subsequently used in frameshift assays where the effect on pseudoknot induced -1 frameshift *in vivo* could be evaluated.

### 3.1 Target mutations

Yusupova et al. originally identified R131, R132, K135 and R164 from S3; R47, R49, and R50 from S4; and R15 and R24 from S5 as potential components of the ribosomal helicase of the 70S ribosome from *T. thermophilus* [Yusupova et al., 2001]. Takyar et al. investigated the effect of a triple mutation in S3 (R131A, R132A, and K135A), a double mutation in S4 (R44 and R47), and a double mutation in S5 (R19 and R28) using reconstituted *E. coli* 70S ribosomes [Takyar et al., 2005]. Based on the the 3Å structure of the *T. thermophilus* 30S ribosome (PDB file 1J5E) and the 5.8Å cryo-EM structure of the *E. coli* 30S ribosome with the *tnaC* leader mRNA (PDB file 2WWL), we decided to target the same amino acid residues as Takyar et al. with a few exceptions. In S3, R164 was targeted instead of K135 as R164 appears closer to the mRNA and in S5 only R20 would be targeted. In addition to R44 and R47 in S4, K45 would also be targeted as it also appears to protrude into the mRNA tunnel<sup>3</sup>. Sequence alignments showed that *T. thermophilus* R47, R49, and R50 in S4 corresponds to K45 and R47 in *E. coli* (R50 did not appear to be conserved in *E. coli*). Also, *T. thermophilus* R15 and R24 in S5 corresponds to R20 and R29 in *E. coli*. The sequence alignments between *T. thermophilus* ribosomal proteins and *E. coli* ribosomal proteins are shown in Figure 5.

The location of the target residues relative to the mRNA entrance tunnel are shown in Figure 6 (red) while mutations investigated by Takyar et al. (but not here) are shown in

---

<sup>3</sup>To avoid confusion about residue numbers, the residue numbers of the mutations constructed here are based on counting the first (AUG) methionine as position 1.

cyan. The figure depicts the mRNA entrance tunnel as seen from the solvent side of the 30S ribosomal subunit.

## Ribosomal protein S3

T. thermophilus	MGNKIHPIGFRLGITRDWESRWYAGKKQYRHLLLEDQIRIRLLEKELYSA	50
E. coli	MGQKVHPNGIRLGIVKPWNSTWFANTKEFADNLDSDFKVRQYLTKELAKA	50
	**:*:** *:*:**: *:* *:*:**: . * . * ::* * ** * *	
T. thermophilus	GLARVDIERAADNAVTVHVAKPGVIGRGGIRVLRREELAKLTGKNVA	100
E. coli	SVSRIVIERPAKSIRVTIHTARPGIVIGKKGEDVEKLRKVVADIAGVPAQ	100
	:::*: ***. * . : *:* *:*:**: ** . . ** :* . : * *	
T. thermophilus	LNVQEVQNPNSAPLVAQRVAEQIERRFAVRAIKQAVQVRMESGAKGAK	150
E. coli	INIAEVRKPELDAKLVADSITSQLERRVMFRAMKRAVQNAMRLGAKGIK	150
	::* *:*:**: * ** * * * : : * : * * * . * * * : * * * * * *	
T. thermophilus	VIVSGRIGGAEQARTEWAAQGRVPLHTLRANIDYGFALARTTYGVLGVKA	200
E. coli	VEVSGRLGGAETATEWYREGRVPLHLRADIDYNTSEHTTYGVIGVKV	200
	* * * * * : * * * * * : * * * * * : * * * * * : * * * * * *	
T. thermophilus	YIFLGEVIGGQPKARPELPAEERPRRRRPAVRVKKEE	239
E. coli	WIFKGEILGGMAAVEQPEKPAAPKQQRK---GRK--	233
	::* *:*:**: * * * * * : * * * * * : * * * * * : *	

## Ribosomal protein S4

T. thermophilus	MGRYIGPVCRLCRREGVKLYLK-GERCYSPKCAMERRPYPGQHGGKRA	49
E. coli	MARYLGPKLKLRRREGTDLFLKSGVRAIDTKCKIEQ---APGQHGARKPR	47
	*:*:**: *:*:**: * * * * * * * . . * * * : * * * * * : *	
T. thermophilus	RPSDYAVRLREKQKLRRIYGISERQFRNLFEEASKKKGVTSVFLGLLES	99
E. coli	-LSDYGVQLREKQKVRRIYGVLERQFRNYYKAAARLKGNTGENLLALLEG	96
	** *:*:**: * * * * * : * * * * * : * * * * * : * * * * *	
T. thermophilus	RLDNVVYRLGFAVSRQARQLVRHGHITVNGRRVDLPSYRVPGDEIAVA	149
E. coli	RLDNVVYRMFGATRAEARQLVSHKAIMVNGRVVNIASYQVSPNDVVSIR	146
	* * * * * : * * * * * * * * * * * * * * * * : * * * * * * * * * * *	
T. thermophilus	EKSRNLELIRQNLAMKGRKVPWLSLDVEGMKGFRLRDPREDLALPVN	199
E. coli	EKAKKQSRVKAALAEQREKPTWLEVDAGMMEGTFKRKPERSDLSADIN	196
	* * * * * . . : * * : * : * * : * : * * * * * * * * * * : *	
T. thermophilus	EQLVIEFYSR	209
E. coli	EHLIVELYSK	206
	* * * * * : * * * * *	

## Ribosomal protein S5

T. thermophilus	----MPETDFEEKMILIRRTARMQAGRRFRFGALVVVGDQRVGLGF	45
E. coli	MAHIEKQAGELQEKLIAVNRVSKTVKGGRIFSFTALTVVGDGNGRVGFGY	50
	::*:**: *:*:**: * * * * * * * * * * * * * * * * : * * * * * *	
T. thermophilus	GKAPEVPLAVQKAGYYARRNMVEVPLQNGTIPHEIEVEFGASKIVLKPA	95
E. coli	GKAREVPAAIQKAMEKARRNMINVALNNGTLQHPVKGVHTGSRVFMQPAS	100
	** *	
T. thermophilus	PGTVIAGAVPRAILELAGVTDILTKEGSRNPINIAATMEALRQLRTK	145
E. coli	EGTGIIAGGAMRAVLEVAGVHNLAKAYGSTNPINVVRATIDGLENMNSP	150
	* * * * * . *	
T. thermophilus	ADVERLRKGEAHAQAQG-	162
E. coli	EMVAAKR-GKSVVEILGK	167
	* * * * * : * * * * *	

Figure 5: Sequence alignments of ribosomal proteins S3, S4 and S5 from *T. thermophilus* and *E. coli*. Green boxes indicate positions of target amino acids which were changed to alanines. Accession numbers for *T. thermophilus* proteins are AEG34099.1, BAA75548.1, and AEG34088.1 for S3, S4, and S5 respectively. Accession numbers for *E. coli* proteins are ACI76865.1, ACI76955.1, and ACI76920.1 for S3, S4 and S5 respectively. Alignments were performed with ClustalW2 using the Gonnet weight matrix, a gap open penalty of 25, and a gap extension penalty of 0.1 [Larkin et al., 2007]. Asterisks denote identical residues, colon represents residues with highly similar properties, and period represents residues with less similar properties.

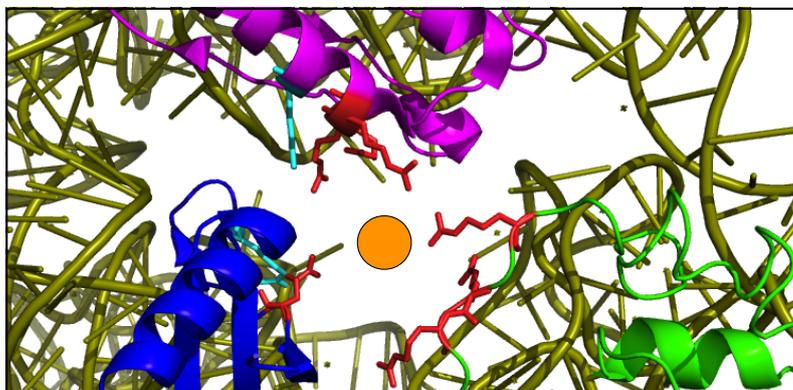


Figure 6: mRNA entrance tunnel. The figure is based on 3Å structure of *T. thermophilus* 30S ribosome (PDB 1J5E). A view from solvent side *into* the mRNA entrance tunnel (orange circle) which is defined by ribosomal protein S3 (magenta), S4 (green) and S5 (blue). 16S ribosomal RNA is shown in pale green. The target amino acid residues are shown in red (R131, R132, and R164 in S3; R44, K45 and R47 in S4; R20 in S5). Residues shown in cyan were investigated by Takyar et al. but were not investigated here. [Takyar et al., 2005]

## 4 Materials and Methods

The mutations were constructed using a combination of site directed mutagenesis and recombineering using the lambda red system as described by Datsenko and Wanner [Datsenko and Wanner, 2000]. This method leaves an 84 bp residual sequence in the genome after removal of the selective marker, a sequence which will be referred to as the FRT-scar (FLP recognition target) in the following.

### 4.1 Bacterial strains and plasmids

Strain	Source	Genotype
MAS78	Michael A. Sørensen	<i>E. coli</i> K-12: <i>recAI</i> $\Delta(lac-pro)$ <i>thi ara</i>
MAS90	Michael A. Sørensen	MAS78 F': <i>lacI</i> <sup>q1</sup> <i>lacZ::Tn5 proAB</i> <sup>+</sup>
MAS327	Michael A. Sørensen	MAS78 + pMAS53
MG1655	UMKC	<i>E. coli</i> K-12: F <sup>-</sup> lambda <sup>-</sup> <i>ilvG</i> <sup>-</sup> <i>rfb-50 rph-1</i>
MC323	UMKC	MG1655: <i>lacZ521</i>
MC323 DKan	Deepali Agarwal (UMKC)	MC323: <i>rpsD</i> -FRT- <i>kan</i> -FRT
MC323 EKan	Deepali Agarwal (UMKC)	MC323: <i>rpsE</i> -FRT- <i>kan</i> -FRT
NF1830	Michael A. Sørensen	MC1000: <i>recA1</i> F': <i>lacI</i> <sup>q1</sup> <i>lacZ::Tn5 proAB</i> <sup>+</sup>

Details about the insertion sites of the FRT-*kan*-FRT cassettes in MC323 DKan and MC323 EKan are given in Figure 7.

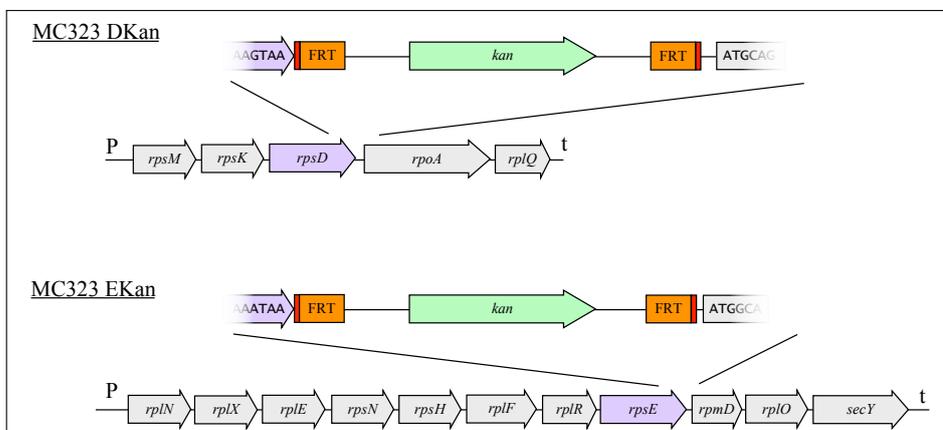


Figure 7: FRT-*kan*-FRT insert in MC323 DKan and EKan. **MC323 DKan:** *rpsD* (magenta) is expressed on a polycistronic messenger RNA [Lindahl and Zengel, 1986]. The FRT-*kan*-FRT cassette is inserted immediately downstream from the *rpsD* stop codon. **MC323 EKan:** *rpsE* (magenta) is expressed on a polycistronic messenger RNA [Lindahl and Zengel, 1986]. The FRT-*kan*-FRT cassette is inserted immediately downstream from the *rpsE* stop codon. The FRT-*kan*-FRT cassette is flanked by priming sites P1 and P2 (red) from Datsenko and Wanner [Datsenko and Wanner, 2000]. “P”: Promotor and “t”: terminator.

---

Plasmid	Source	Features
pCP20	[Cherepanov, 1995]	Thermal induction of FLP synthesis, infers resistance towards ampicillin and chloramphenicol, ts for replication
pKD3	[Datsenko and Wanner, 2000]	FRT- <i>cat</i> -FRT, chloramphenicol and ampicillin resistance
pKD4	[Datsenko and Wanner, 2000]	FRT- <i>kan</i> -FRT, kanamycine and ampicillin resistance
pKD46	[Datsenko and Wanner, 2000]	Expresses the Red system under the <i>araB</i> promotor, infers resistance towards ampicillin, ts for replication
pMAS53	Michael A. Sørensen	<i>recA</i> <sup>+</sup> , infers tolerance towards chloramphenicol, ts for replication

---

## 4.2 DNA oligonucleotides

Name	Sequence (5' → 3')	Use
oJT_rpsC1	TCCTGAAGCGCACCAGCCACATCACTGTGGTTGTGT CCGATCGCTGAGACGTGTAGGCTGGAGCTGCTTCG	F primer for FRT-cat/kan-FRT in pKD3 and pKD4. Insertion of cassette upstream of <i>rpsC</i>
oJT_rpsC2	AGGCGAATACCATTAGGATGTACTTTCTGACCCATT GCTAGTCTCCAGACATATGAATATCCTCCTTAG	R primer for FRT-cat/kan-FRT in pKD3 and pKD4. Insertion of cassette upstream of <i>rpsC</i>
oJT_rpsC3	TGCCGCGTGCAAAAGGTC	F primer for amplification of <i>rpsC</i>
oJT_rpsC4	TGCGCCAGACCGCGGTTAC	R primer for amplification of <i>rpsC</i>
oJT_rpsC5	AAAAACCGGCTGCTCAGCCTAAAAAGCAGCAGCGT AAAGGCCGTAATAAGTGTAGGCTGGAGCTGCTTCG	F primer for FRT-cat/kan-FRT in pKD3 and pKD4. Insertion of cassette downstream of <i>rpsC</i>
oJT_rpsC6	TGCATTTTACGGAATTTTGTACGCTTTGGTTGTAA CATCAGCGACGCTCCCATATGAATATCCTCCTTAG	R primer for FRT-cat/kan-FRT in pKD3 and pKD4. Insertion of cassette downstream of <i>rpsC</i>
oJT_rpsC7	GTCGCGTTATGTTTCGCTCGTGTATGAAGCGTGC	F primer for S3(R131A)
oJT_rpsC8	GCACGCTTCATAGCAGGAGGAACATAACGCGAC	R primer for S3(R131A)
oJT_rpsC9	GTCGCGTTATGTTTCGCTCGTGTATGAAGCGTGC	F primer for S3(R132A)
oJT_rpsC10	GCACGCTTCATAGCAGCAGGAACATAACGCGAC	R primer for S3(R132A)
oJT_rpsC11	GCGGAAATCGCAGCTACCGAATGGTACC	F primer for S3(R164)
oJT_rpsC12	GGTACCATTCCGCTAGCTGCGATTTCGCGC	R primer for S3(R164A)
oJT_rpsD1	CAGCACGGTGGCGCTAAACCGCGTCTGTG	F primer for S4(R44A)
oJT_rpsD2	GACAGACGGGTTTAGCCGACCCGCTGCTG	R primer for S4(R44A)
oJT_rpsD3	CAGCACGGTGGCGCTGACCCGCGTCTGTG	F primer for S4(K45A)
oJT_rpsD4	GACAGACGGTGGCAGCGCACCCGCTGCTG	R primer for S4(K45A)
oJT_rpsD5	GTGCGCGTAAACCGGCTCTGTCTGACTATG	F primer for S4(R47A)
oJT_rpsD6	CATAGTCAGACAGAGCCGGTTACGCGCAC	R primer for S4(R47A)
oJT_rpsD7	CAGCACGGTGGCGNNNAACCGCGTCTGTG	F primer for S4(R44X)
oJT_rpsD8	GACAGACGGGTTTNNNCGACCCGCTGCTG	R primer for S4(R44X)
oJT_rpsD9	GTGCGCGTAAACCGNNNCTGTCTGACTATG	F primer for S4(R47X)
oJT_rpsD10	CATAGTCAGACAGNNNCGGTTTACGCGCAC	R primer for S4(R47X)
oJT_rpsE1	GATCGCGGTAACCGCGTATCTAAAACCG	F primer for S5(R20A)
oJT_rpsE2	CGGTTTTAGATACGGCGTTTACCGCGATC	R primer for S5(R20A)
rpsD-drug5	FTCCAGGCCGGAATCTACT	F primer for <i>rpsD</i> amplification
rpsD-drug3	ACGCTCTAAAGGCTCAAGGG	R primer for <i>rpsD</i> amplification
rpsE-drug5	CTGTGGGTAAAGCTGTGCGT	F primer for <i>rpsE</i> amplification
rpsE-drug3	GTGACCAATACGACGCAGAC	R primer for <i>rpsE</i> amplification
S4mutamp5	CAGGGTAACGCGTTGGGTTG	F primer for <i>rpsD</i> amplification
S4mutamp3	CATCTTTGCCCTGAACTCTC	R primer for <i>rpsD</i> amplification
S4mutseq5	TCAGGTTGCAGCAGAGCGTT	F primer for <i>rpsD</i> sequencing
S4mutseq3	GCCTTCTTTGGTGTGCTACT	R primer for <i>rpsD</i> sequencing
S5mutamp5	GGTGTTCGTTACGCCGACGA	F primer for <i>rpsE</i> amplification
S5mutamp3	CGACGGTACAGAGGCATCTG	R primer for <i>rpsE</i> amplification
S5mutseq5	CGTACCCCGCTCACATTTA	F primer for <i>rpsE</i> sequencing
S5mutseq3	GGCCAGAACCAGTACCACGA	R primer for <i>rpsE</i> sequencing

“F” = Forward, “R” = Reverse. The following oligonucleotides were provided by Deepali Agarwal at the UMKC: rpsD-drug5, rpsD-drug3, rpsE-drug5, rpsE-drug3, S4mutamp5, S4mutamp3, S4mutseq5, S4mutseq3, S5mutamp5, S5mutamp3, S5mutseq5, S5mutseq3. All oligonucleotides were store in water at -20°C. Underline sequence indicate mismatches, “N” denotes any nucleotide and “X” denotes any amino acid.

### 4.3 Equipment

Centrifuges	Allegra 25R Eppendorf tabletop
Electroporation	BioRad Micropulser
Gradient reader	ISCO UA-6 UV/Vis Detector w. BioRad Econo pump
PCR Machine	Applied Biosystems 2720 Thermal Cycler
PCR Purification	Promega Wizard SV PCR+Gel Clean-up system
Photometer	Beckman Coulter, DU 520
Ultracentrifuge	Beckman L8-80M Ultracentrifuge, SW28 roter

### 4.4 Extraction of genomic DNA

0.8 ml of an over night culture was pelleted at 5,000 $xg$  for 4 minutes. Pellet was resuspended in 100  $\mu$ l 150 mM NaCl, 100 mM EDTA, pH 8. A small amount of lysozyme was added (a few crystals) and the sample was incubated at 37°C for 30 minutes. To the clear and viscous lysate 400  $\mu$ l of a mix of phenol:tris/chloroform/isoamyl alcohol (25/24/1) was added and the phases were separated at 18,000 $xg$  for 5 minutes. The supernatant was subjected to another phenol:tris/chloroform/isoamyl alcohol extraction after which 7  $\mu$ l 3M sodium acetate pH 5.2 and 175  $\mu$ l ice cold 100% ethanol was added to the supernatant. The genomic DNA was extracted with an inoculation loop, air dried and solubilized in 100  $\mu$ l 10 mM Tris-HCl pH 8, 2 mM EDTA

### 4.5 Electroporation

Electocompetent cells were prepared as follows: 2.5 ml of an over night culture of MG1655 or MC323 carrying pKD46 was used to inoculate 250 ml LB broth supplemented with 0.2% (w/v) L-arabinose and 100  $\mu$ g/ml ampicillin. At an optical density (OD) of 0.5 (at 600 nm) the cells were pelleted (2,500 $xg$ , 4°C, 5 min) and resuspended in 250 ml ice cold 10% glycerol. The cells were pelleted and resuspended in 185 ml 10% ice cold glycerol. The cells were pelleted again and resuspended 1 ml ice cold 10% glycerol and stored in 50  $\mu$ l aliquots at -80°C.

Electroporation was conducted as follows: Competent cells were thawed on ice and transferred to a cuvette on ice. 1  $\mu$ l purified PCR product was added and the transformation was performed using the preset "Ec1" program. Approximately 600  $\mu$ l SOC media [Hanahan, 1983] was added and the cuvette was placed on ice. The cells were gently resuspended, transferred to an eppendorf tube and incubated at 37°C for two hours. Cells were pelleted at 2,000 $xg$  for 3 min at room temperature and resuspended in 200  $\mu$ l SOC media. 100  $\mu$ l of the resuspended cells were plated on selective LB plates while the remaining 100  $\mu$ l were placed at room temperature over night as back-up.

### 4.6 P1vir lysate

A dilution of a P1vir phage lysate (supplied by Deepali Agarwal at the UMKC) was diluted with LB media to concentrations  $10^{-2}$ ,  $10^{-3}$ ,  $10^{-4}$  and  $10^{-5}$ . An over night culture of the donor strain was diluted 10 fold in LB media and 100  $\mu$ l of this dilution was mixed with 100  $\mu$ l of each P1vir phage dilution in 1 ml LB media. The suspension was incubated at room temperature for 10 minutes whereafter 3 ml melted 0.77% nutrient agar was added. The mix was plated on LB GC plates (LB-agar supplemented with 0.2% glucose and 50 mM CaCl<sub>2</sub>). The plates are incubated (bottom down) in a moist environment at 37°C.

The next day, a plate showing clear and well defined plaques was selected and 3 ml LB media was added to the nutrient agar. The nutrient agar was mashed with a Drigalski spatula and transferred to a centrifuge tube. A small amount of chloroform was added and the tube was incubated at room temperature for one hour with regular inversions. Debris were removed by centrifugation at 3,400 $xg$ , 5°C, 15 minutes and the clear supernatant was transferred to a glass vial. 200  $\mu$ l chloroform was added and the lysate was placed at 5°C. The lysate was left to settle for a few days prior to use.

## 4.7 P1 transduction

To an over night culture of the recipient strain (MAS327), 1 M CaCl<sub>2</sub> was added to 5 mM (final) and aliquots of 650  $\mu$ l was transferred to four Eppendorf tubes. Varying amounts of P1vir lysate was added to each tube, i.e. 5  $\mu$ l, 10  $\mu$ l, 20  $\mu$ l and 40  $\mu$ l, and the suspension was incubated at room temperature for one hour. 100  $\mu$ l of each mix was plated on selective LB plates and incubated at 37°C over night. Transductants were re-purified on selective LB plates followed by PCR analysis to ensure correct location of the resistance cassette. Once the location of the cassette was confirmed a liquid culture was inoculated and the strain was stored in 20% glycerol at -80°C.

The strains were subsequently tested for UV- and ampicillin sensitivity.

## 4.8 PCR

### Normal PCR

PCR was performed either with Phusion High-Fidelity DNA polymerase (New England Biolabs, #M0530) using HF-buffer according to instructions or using Taq DNA polymerase (New England Biolabs, #M0273) with standard Mg-containing buffer according to instructions.

### Two fragment PCR

To make site-directed mutagenesis a two-fragment PCR procedure was employed. First, two PCR fragments were produced with complementary ends containing the mutation. Second, the two products were annealed and extended for ten "PCR" cycles in a mixture lacking primers. Third, after ten cycles the amplification primers (the forward primer from the upstream PCR product and the reverse primer from the downstream product was added) and the product was amplified for 20-25 cycles of PCR.

## 4.9 Sucrose Gradient Sedimentation

Sucrose gradients were prepared as follows: 22.5% sucrose (w/v) was prepared in polysome buffer (25 mM Tris-HCl pH 7.9, 50 mM KCl, 10 mM MgCl<sub>2</sub> and 3 mM DTT). The sucrose solution was transferred to Beckmann 25x89 mm tubes in aliquots of 36 ml. The tubes were stored at -80°C.

Overnight cultures were diluted 100 fold in preheated LB media and grown at 37°C until OD<sub>436</sub> was approximately 0.8. Cells were pelleted at 7,000*xg*, 4°C, 5 min and resuspended in 2 ml wash solution (20% sucrose, 25 mM Tris-HCl pH 8.0, 60 mM KCl and 10 mM MgCl<sub>2</sub>). The suspension was transferred to a 2 ml Eppendorf tube and the cells were pelleted at 14,000*x* for 2 min at room temperature. Cells were resuspended in 150  $\mu$ l lysis solution (wash solution + a few crystals of lysosyme) and lysed through 4-6 freeze-thaw cycles in a ethanol/dry ice bath until the suspension became viscous. The following was added: 450  $\mu$ l TKM buffer (25 mM Tris-HCl pH 8.0, 60 mM KCl and 10 mM MgCl<sub>2</sub>), 90  $\mu$ l 5% Brij 35 in 10 mM Tris-HCl pH 8.0, 30  $\mu$ l 1 mg/ml RNase free DNase in 50 mM Tris-HCl pH 7.2, 150  $\mu$ l 1% deoxycholate in 0.1 M Tris-HCl pH 8.0 (fresh) and 37.5  $\mu$ l 0.1 M MgSO<sub>4</sub>. The suspension was mixed and incubated on ice for approximately 20 min. Debris were pelleted by 14,000*xg* for 10 min at room temperature. The supernatant was kept on ice.

The 22.5% sucrose solutions from -80°C were thawed in a fume hood and the gradient was created during the thawing process. After complete thawing, 30 A<sub>260</sub> units (1 cm light path) was loaded on top of the gradient and the samples were centrifuged at 52,000*xg* for 17.5 hours at 4°C.

After centrifugation the gradients were analyzed by pumping 60% glycerol into the bottom of the gradient (6 ml/min) and measuring the absorbance at 260 nm of the displaced gradient. The ISCO UA-6 UV/Vis Detector prints the absorbance on a physical media (paper) which was scanned to obtain a digital representation of the curve. The data presented here is simply the scanned image converted to grayscale without the background (Figure 9 and 10) .

## 4.10 Frameshift assay

As described in [Tholstrup et al., 2012]. Briefly, expression of a suitable frameshift construct was induced in an exponentially growing culture in minimal MOPS media [Neidhardt et al., 1974] using glycerol as carbon source at  $OD_{436}$  0.4-0.7 using 1 mM Isopropyl  $\beta$ -D-Thiogalactopyranoside (IPTG). After induction for 15 minutes proteins were pulse labelled for 20 seconds using radioactive methionine followed by a 2 min chase with large molar excess of unlabeled methionine. Cells were harvested and proteins were separated using one-dimensional SDS-PAGE. Frameshift efficiency was calculated from the ration of frameshift to non-frameshift product.

## 4.11 Construction of *rpsC* mutations

### 4.11.1 Insertion of FRT-cat/kan-FRT upstream/downstream of *rpsC*

A PCR reaction was set up using pKD4 as template and oJT\_rpsC1 and oJT\_rpsC2 as primers. Another PCR reaction was set up using pKD3 as template and oJT\_rpsC1 and oJT\_rpsC2 as primers. The resulting PCR products (1576 bp and 1113 bp respectively) were purified and electoporated into MC323 and MG1655 and selected on LB with either 25  $\mu$ g/ml kanamycine or 25  $\mu$ g/ml chloramphenicol. Eight transformants from each was re-streaked on selective plates. PCR with oJT\_rpsC3 and oJT\_rpsC4 were used to verify the correct location of the FRT-kan/cat-FRT cassette.

Insertion downstream from *rpsC* was attempted as described for insertion upstream except that oJT\_rpsC5 and oJT\_rpsC6 was used as PCR primers in stead of oJT\_rpsC1 and oJT\_rpsC2. The insertion sites are shown in Figure 8.

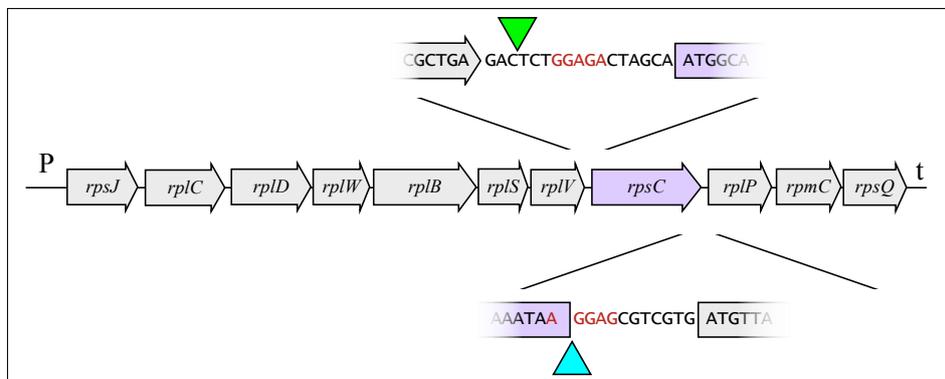


Figure 8: FRT-*kan*-FRT insertion near *rpsC* which is expressed on a polycistronic messenger RNA [Lindahl and Zengel, 1986]. Attempts were made to insert either the FRT-*kan*-FRT or the FRT-*cat*-FRT cassette upstream or downstream from *rpsC* (magenta). The upstream location (green wedge) was located between the stop codon of *rplV* and a potential ribosome binding site for *rpsC* (red). The downstream location (cyan wedge) was immediate after the *rpsC* stop codon and inside a potential ribosome binding site for *rplP* (red). The FRT-*kan*-FRT (from pKD4) and the FRT-*cat*-FRT (from pKD3) cassettes each contain a ribosome binding site and start codon in-frame with *rplP* [Datsenko and Wanner, 2000]. P = promoter, t = terminator.

*This method did not provide transformants with either cassette in the correct location. See main text for details.*

### 4.11.2 S3(R131A)

Two PCR reactions were set up using genomic DNA from MC323 EKan as template and the following primer pairs: oJT\_rpsC7+rpsE-drug3 (504 bp) and oJT\_rpsC3+oJT\_rpsC8 (6379 bp). The two PCR products were purified and used as described for S4(R44A) below.

### 4.11.3 S3(R132A)

Two PCR reactions were set up using genomic DNA from MC323 EKan as template and the following primer pairs: oJT\_rpsC9+rpsE-drug3 (504 bp) and oJT\_rpsC3+oJT\_rpsC10 (6379 bp). The two PCR products were purified and used as described for S4(R44A) below.

### 4.11.4 S3(R164A)

Two PCR reactions were set up using genomic DNA from MC323 EKan as template and the following primer pairs: oJT\_rpsC11+rpsE-drug3 (599 bp) and oJT\_rpsC3+oJT\_rpsC12 (6278 bp). The two PCR products were purified and used as described for S4(R44A) below.

## 4.12 Construction of *rpsD* mutations

### 4.12.1 S4(R44A)

Using genomic DNA from MC323 DKan as template two separate PCR reactions were set up with the following primer pairs: oJT\_rpsD1+rpsD-drug3 (2105 bp) and oJT\_rpsD2+rpsD-drug5 (301 bp). The two products were purified and used (2  $\mu$ l of each) in a 50  $\mu$ l PCR mix (without primers). After 10 cycles (98°C for 30 sec, 61°C for 30 sec, 72°C for 1 min) rpsD-drug5 and rpsD-drug3 was added to 0.2  $\mu$ M each. The mix was subjected to an additional 25 PCR cycles for amplification and the resulting PCR product (2377 bp) was purified.

The purified PCR product was electroplated into MC323 and plated on LB plates supplemented with 25  $\mu$ g/ml kanamycine. Transformants were re-streaked on selective plates and used to inoculate 3 ml LB media with 25  $\mu$ g/ml kanamycine. Cells from 100  $\mu$ l of the over night culture was pelleted and resuspended in 100  $\mu$ l water and incubated at 37°C for 30 min for lysis whereafter 2 $\mu$ l were used for PCR using rpsD-drug5 and rpsD-drug3 as primers. The PCR-fragment (2377 bp) were purified and sequenced.

P1vir lysate was prepared from a strain carrying the S4(R44A) mutation and the FRT-*kan*-FRT cassette. The lysate was used to transduce MAS327 and the cells were plated on LB plates containing 25  $\mu$ g/ml kanamycine at 37°C. Transductants were re-streaked twice on selective plates. A colony was used to inoculate 3 ml LB+25  $\mu$ g/ml kanamycine and PCR (using rpsD-drug5+rpsD-drug3) on lysed over-night culture was used to ensure the correct location of FRT-*kan*-FRT cassette (2377 bp). The cells were also tested for sensitivity towards UV-light and chloramphenicol at 30°C to ensure loss of pMAS53. To remove the FRT-*kan*-FRT cassette, the purified transductants were transformed with pCP20 using a CaCl<sub>2</sub> wash of pelleted over-night cultures, plated on LB+100 $\mu$ g/ml ampicillin and incubated at 30°C. The transformants were re-streaked once at 30°C on selective plates followed by another re-streak on LB plates at 42°C. The purified colonies were tested for ampicillin and kanamycine sensitivity at 30°C. PCR on lysed over-night cultures using S4mutamp5 and S4mutamp3 as primers to validate the loss of FRT-*kan*-FRT-cassette (1331 bp) and to ensure the correct sequence of the FRT-scar region (using S4mutseq5 and S4mutseq3 as sequencing primers). Finally, the strain was made F<sup>+</sup> by cross with NF1830.

### 4.12.2 S4(K45A)

As described for S4(R44A) except that the initial two PCR fragments were created with the following primer pairs: oJT\_rpsD3+rpsD-drug3 (2105 bp) and oJT\_rpsD4+rpsD-drug5 (301 bp).

### 4.12.3 S4(R47A/X)

As described for S4(R44A) except that the initial two PCR fragments were created with the following primer pairs: oJT\_rpsD5+rpsD-drug3 (2098 bp) and oJT\_rpsD6+rpsD-drug5 (309 bp). Random mutations at R47 was also created using oJT\_rpsD9 and oJT\_rpsD10.

## 4.13 Construction of *rpsE* mutations

### 4.13.1 S5(R20A)

As described for S4(R44A) except that the initial two PCR fragments were created with the following primer pairs: oJT\_rpsE1+rpsE-drug3 (2042 bp) and oJT\_rpsE2+rpsE-drug5 (224 bp) using MC323 EKan genomic DNA as template.

## 5 Results

### 5.1 Constructed strains

As opposed to Takyar et al. who used triple and double mutations in their investigation of the ribosomal helicase *in vitro* [Takyar et al., 2005] we made only single mutations. This should both increase the resolution of the assay and reduce the risk of attempting to construct lethal mutations, as our *in vivo* approach requires the host strain to carry the mutations in their genome,

We were unable to obtain any strains carrying mutations in ribosomal protein S3 (*rpsC*) as no transformants with either of the two resistance cassettes, FRT-*kan*-FRT or FRT-*cat*-FRT, either upstream or downstream of the *rpsC* gene was obtained. Some background growth was observed on selective plates following electroporation but none of these survived purification on selective plates. Another attempt was made using the FRT-*kan*-FRT cassette inserted immediately downstream from S5 (*rpsE*) as selective marker for *rpsC* mutants. A few transformants were obtained but sequence analysis showed that they carried the wild-type *rpsC* gene.

All of the desired mutations in ribosomal protein S4 (*rpsD*) were obtained and in addition a S4(R47V) mutant was obtained using primers with random nucleotides at codon 47 (oJT\_rpsD9 and oJT\_rpsD10). The single mutation in ribosomal protein S5 (*rpsE*) was also obtained.

Table 1 lists the mutations constructed here and their strain name with which they will be referred to hereafter. In addition, isogenic strains were created containing the wild-type *rpsD* and *rpsE* gene respectively, with the downstream FRT-scar (JT762 and JT761 respectively).

To obtain a quick assessment of the physiological effect to the mutations, the doubling time in liquid minimal MOPS media [Neidhardt et al., 1974] using glycerol as carbon source at 37°C was determined from two independent colonies of each strain (Table 1).

With the exception of JT756 there was no statistically significant difference between the doubling times of the mutant strains and their isogenic wild-type (for JT756 p-value = 0.02 (Two-tailed, Student's t-test) compared to JT762). Interestingly, although only significant for JT762 (p-value = 0.03, Two-tailed, Student's t-test), the double time of both the isogenic wild-types were longer than that of MAS90 (78.1±3.4 min, not shown).

Strain	Mutation	Genotype	Doubling time
JT726	CGT → GCT	MAS78: <i>rpsD</i> (R44A)-FRTscar	N.D.
JT731	AAA → GCA	MAS78: <i>rpsD</i> (K45A)-FRTscar	N.D.
JT734	CGT → GCT	MAS78: <i>rpsD</i> (R47A)-FRTscar	N.D.
JT738	CGT → GTT	MAS78: <i>rpsD</i> (R47V)-FRTscar	N.D.
JT744	CGC → GCC	MAS78: <i>rpsE</i> (R20A)-FRTscar	N.D.
JT748	<i>none</i>	MAS78: <i>rpsE</i> <sup>+</sup> -FRTscar	N.D.
JT750	<i>none</i>	MAS78: <i>rpsD</i> <sup>+</sup> -FRTscar	N.D.
JT752	CGT → GCT	MAS90: <i>rpsD</i> (R44A)-FRTscar	96.1±3.2 min
JT754	AAA → GCA	MAS90: <i>rpsD</i> (K45A)-FRTscar	93.7±9.8 min
JT756	CGT → GCT	MAS90: <i>rpsD</i> (R47A)-FRTscar	105.3±1.4 min
JT758	CGT → GTT	MAS90: <i>rpsD</i> (R47V)-FRTscar	102.0±8.0 min
JT759	CGC → GCC	MAS90: <i>rpsE</i> (R20A)-FRTscar	87.1±4.0 min
JT761	<i>none</i>	MAS90: <i>rpsE</i> <sup>+</sup> -FRTscar	88.8±3.1 min
JT762	<i>none</i>	MAS90: <i>rpsD</i> <sup>+</sup> -FRTscar	93.7±1.9 min

Table 1: Ribosomal mutations. Each mutation was created as described in section 4 and contained a downstream FRT-scar in addition to the mutation (see Datsenko and Wanner for details about FRT-scar [Datsenko and Wanner, 2000]). Each mutation is listed both on DNA level and protein level. JT761 and JT762 are isogenic wild-types and carry wild-type *rpsE* and *rpsD* respectively along with a downstream FRT-scar. Doubling time in liquid minimal MOPS media [Neidhardt et al., 1974] at 37°C is mean±SD,  $n = 2$ . N.D.: not determined.

This indicates that FRT-scar has an effect on bacterial growth, which for the *rpsD* mutants is supported by the sucrose gradient sedimentation analysis shown below.

## 5.2 Sucrose Gradient Sedimentation

Polysomes were extracted to investigate if the constructed mutations alter the *in vivo* pool of ribosomes. Sucrose gradient sedimentation analysis of all S4 mutants are shown in Figure 9 where a wild-type has been included which does not contain the FRT-scar downstream from *rpsC* (WT, MAS78). The isogenic wild-type with the FRT-scar downstream from *rpsC* (JT750) have also been included. The sucrose gradients show that the *in vivo* ribosomal pool has been altered in the mutants, i.e. relative abundance of free 30S and 50S ribosomal subunits relative to assembled 70S ribosomes have increased. However, the effect most likely originates in the FRT-scar and not in the amino acid substitutions (compare WT and JT750). Although the ribosomal pool have changed, we still observe a significant amount of assembled 70S ribosomes and a polysome population comparable to the wild-type (WT).

Sucrose gradient sedimentation analysis of the S5 mutant is shown in Figure 10, where the wild-type without the FRT-scar has been included (WT) in addition to the isogenic wild-type which carries the FRT-scar downstream from *rpsE*. In contrast to the S4 mutants, it seems reasonable to conclude that neither the presence of the FRT-scar nor the presence

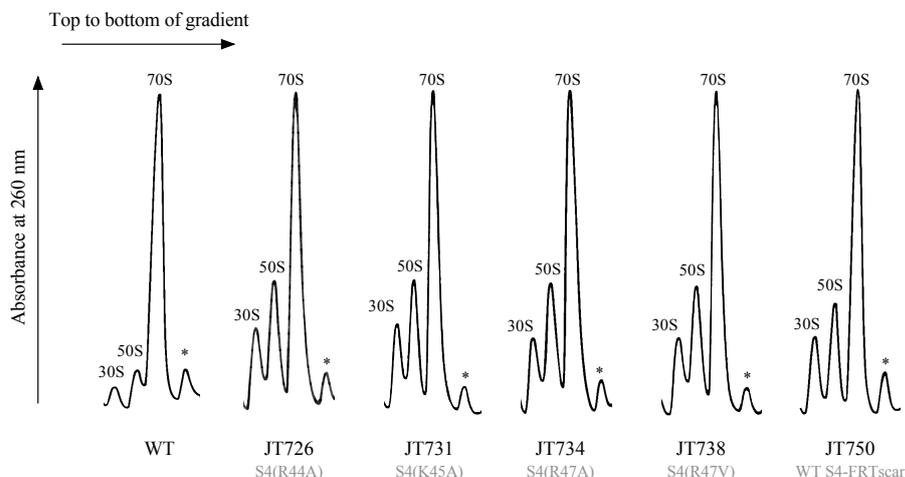


Figure 9: Sucrose gradient sedimentation analysis for each strain including true wild-type with no FRT-scar (MAS78, WT). Gradients were run for two independent colonies of each strain, one of which is shown here. Ribosomal 30S and 50S subunits, 70S ribosomes, and polysome (asterisk) are indicated. The gradients were analyzed from the top to the bottom while absorbance at 260 nm was monitored.

of the R20A substitution significantly affects the *in vivo* pool of ribosomes.

### 5.3 Frameshift assay

Finally, *in vivo* frameshift assays were used to investigate how the mutations in the putative helicase effects pseudoknot induced -1 frameshift. Each strain (MAS90 derivatives) was transformed with a pOFX302 derived plasmid containing the pseudoknot of interest in a suitable context and used in our frameshift assay, as described in materials and methods. In addition to the two pseudoknots 22/6a and 22/6c from this work two constructs from Hansen et al. (TH400 and TH421), were used in these frameshift assays [Tholstrup et al., 2012, Hansen et al., 2007]. The construct TH421, in which the pseudoknot was replaced with 5'-GCCG-3', was used as a negative control. The -1 frameshift efficiencies using these constructs are shown in Figure 11.

There was no difference in the frequency of -1 frameshift between the isogenic wild-types and the mutants for either of the three pseudoknots. In addition, there was no difference in frameshift efficiency between the true WT (MAS90) and these strains [Tholstrup et al., 2012]. This indicates, that if the putative helicase is involved in pseudoknot induced -1 frameshift it is sufficiently robust to remain functional without all of its positive amino acid residues. Interestingly, JT756 seems to have an increased frameshift efficiency in the negative control lacking a pseudoknot, indicating that and R47A mutation in S4 has an impact on reading frame maintenance while a R47V mutation does not.

None of the mutations were found to induce ribosomal stalling when translating either

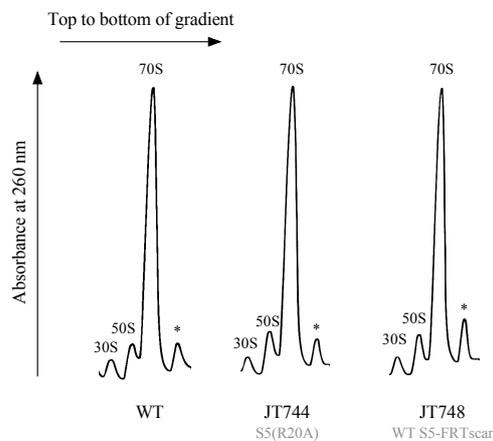


Figure 10: Sucrose gradient sedimentation analysis for each strain including true wild-type with no FRT-scar (MAS78, WT). Gradients were run for two independent colonies of each strain, one of which is shown here. Ribosomal 30S and 50S subunits, 70S ribosomes, and polysome (asterisk) are indicated. The gradients were analyzed from the top to the bottom while absorbance at 260 nm was monitored.

pseudoknot TH400 or pseudoknot 22/6c when 2D SDS-PAGE was used as described in Part I (not shown).

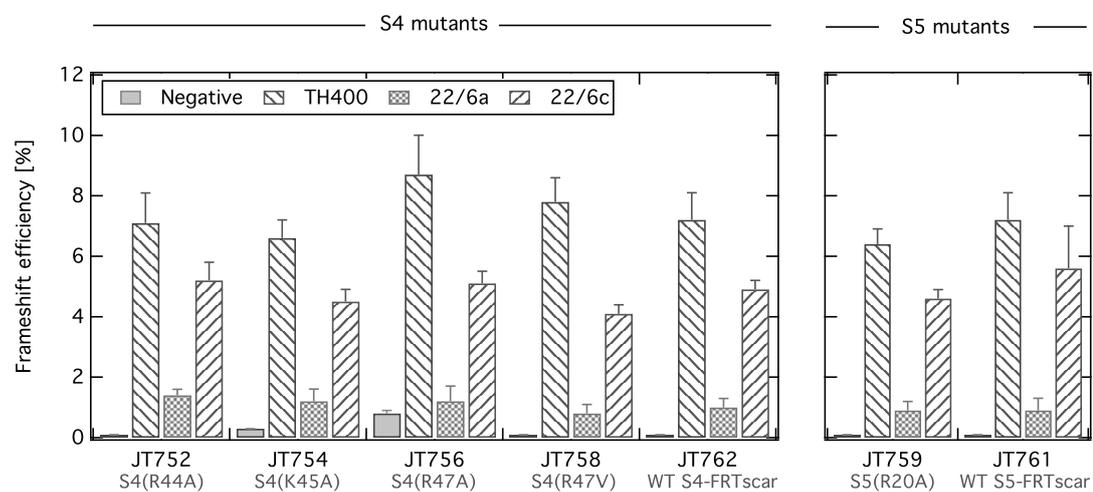


Figure 11: Frameshift efficiency of ribosomal mutations. Left panel show frameshift efficiencies for S4 mutants and right panel show frameshift efficiencies for S5 mutants. Isogenic wild-types were used as reference for both S4 and S5 mutants (JT762 and JT761 respectively). Values are mean $\pm$ SEM,  $n = 2$  for negative control and  $n = 3-5$  for all other.

## 6 Discussion - The Putative Ribosomal Helicase

The experiments conducted here were inspired by an *in vitro* assay conducted by Takyar et al. which indicated that ribosomal proteins S3 and S4 could define a functional ribosomal helicase [Takyar et al., 2005]. Their results although qualitative and not quantitative, are interesting and potentially important for the understanding pseudoknot induced -1 frameshift, as the mechanism by which the pseudoknot is dissolved during the frameshift process is unknown.

Here we were able to obtain mutations in ribosomal proteins S4 and S5. We did not obtain the desired mutations in ribosomal protein S3 which was identified by Takyar et al. as one of the two key proteins in the helicase. Attempts were made to insert an FRT-*cat*-FRT or an FRT-*kan*-FRT cassette both upstream and downstream from *rpsC* without success. Considering the impact of the FRT-scar on the *in vivo* ribosome pool when present downstream from *rpsD*, it is possible that the resistance cassettes near *rpsC* could alter the ribosome pool to a lethal degree. As mentioned in materials and methods, the insertion site downstream from *rpsC* could interfere with the ribosome binding site for *rplP* and that the FRT-*kan*-FRT and FRT-*cat*-FRT cassettes each contained a ribosome binding site and an in-frame start codon relative to *rplP*. It is possible that this could have a negative impact on the function of *rplP* and it could be worth trying to use the FRT-*kan*-FRT cassette from pKD13 instead, which contains neither a ribosome binding site nor a start codon [Datsenko and Wanner, 2000].

Attempts were also made to construct *rpsC* mutations using the FRT-*kan*-FRT cassette downstream from *rpsE* (ME323 E<sub>Kan</sub>), however, these efforts were also futile. It is possible that the target mutations in S3 has a lethal impact on the function of the ribosome either through an inability to translate structured mRNA or due to assembly defects. It has been shown previously, that changing even a single amino acid in one of the ribosomal proteins (S5[G28D]) can have detrimental effects on ribosome assembly and cellular growth [Kirthi et al., 2006].

The frameshift assay showed that none of the mutations in the putative ribosomal helicase investigated here, had an impact on pseudoknot-induced -1 frameshift (Figure 11). Interestingly, the S4(R47A) mutation showed an increased level of frameshift in the negative control where the mRNA lacks the pseudoknot sequence. This is most likely attributed to increased frameshift over the slippery sequence as simple read-through of the UAA stop codon would not result in full length *gene10-lacZ* fusion. It has previously been shown that mutations in S4, and in many other locations in the ribosome, can result in an increase in translational errors (*ram* phenotype) and it is possible that the mutations constructed here results in similar phenotypes, however, this was not addressed in the experiments conducted here [Agarwal et al., 2011, Björkman et al., 1999, Andersson et al., 1986].

In 1994 Dinman and Wickner showed that the *SUP46* allele, originally identified as mutations in ribosomal protein S13 (two silent mutations and two substitutions; R40K and L97W), was able to increase the frequency of -1 frameshift in *Saccharomyces cerevisiae* (*S. cerevisiae*) [Vincent and Liebman, 1992, Dinman and Wickner, 1994]. Ribosomal protein S13 from *S. cerevisiae* is, based on sequence similarity, assumed to be a homolog

of prokaryotic ribosomal protein S4 (one of the putative helicase components) and their findings could therefore constitute the first experimental observations of ribosomal helicase defects *in vivo*. However, a recent 3Å crystal structure of the *S. cerevisiae* (strain S288c) 80S ribosome indicate that ribosomal protein S13<sup>4</sup> is not as central to the mRNA entrance tunnel as seen in *E. coli* and *T. thermophilus* ribosomes [Ben-Shem et al., 2011]. In addition, the two substitution mutations (R40K and L97W) in yeast ribosomal protein S13 is quite distant from the mRNA entrance tunnel, which makes it questionable if the effect observed by Dinman and Wickner is likely to originate explicitly from defects in the ribosomal helicase. In the eukaryotic 80S ribosome, the mRNA entrance tunnel appears to be defined by eukaryotic ribosomal proteins S2 (S5 in prokaryotes), S3 (S3 in prokaryotes), and S30 (unique to eukaryotes) [Ben-Shem et al., 2011].

The strain in which the FRT-*kan*-FRT cassette was inserted downstream from *rpsE* (MC323 EKan), suffered from a minor drawback. There was an in-frame deletion of three base pairs ( $\Delta$ AAA) in *rpmD* (encoding ribosomal protein L30) downstream from *rpsE*, which removes K6 from ribosomal protein L30 (Figure 7, page 36). If additional mutations in S5 are to be investigated it could be worthwhile to construct a new strain in which *rpmD* does not contain this deletion.

Another, perhaps more appealing, approach in the construction of future mutations is to use the integrating plasmid pKO-system developed by Link et al. which does not require the insertion of a selectable marker in the vicinity of the gene of interest and which does not leave a “scar” on the genome [Link et al., 1997]. The latter feature is of particular interest in light of the effect observed for the FRT-scar downstream from *rpsD*. It would also be worth while to attempt to construct the *rpsC* mutants using this system, avoiding any possible effects from the resistance cassette.

Although the mutations constructed here failed to demonstrate a role for the putative helicase in pseudoknot induced -1 frameshift, it was demonstrated that these mutations can be constructed and that their effect on pseudoknot induced -1 frameshift can be investigated *in vivo*. It is worth noticing, that the experiments conducted here only determines whether or not the mutations alter the frequency of pseudoknot induced -1 frameshift, and not whether the putative helicase is functional or not. Given the natural propensity of RNA to form structures, it is however possible, that severe defects in the helicase would render the host strain non-viable thus making the *in vivo* analysis of helicase defects unattainable.

---

<sup>4</sup>Referred to as ribosomal protein S4 in their main text (their figure 1B) and as yeast ribosomal protein S9 in the supplementary (table S4). The confusion originates from the fact that UniProt recommends referring to yeast ribosomal protein S13 as yeast ribosomal protein S9-B (compare P05755.4 with AAB60283.1). Furthermore, eukaryotic ribosomal protein S9 describes the generic ribosomal protein S4 (<http://www.uniprot.org/docs/ribosomp>).



## Part III

# Single Molecule Force Spectroscopy

## 7 Introduction to Force Spectroscopy

In 1619 the German astronomer Johannes Kepler proposed traveling to the Moon by sailing on the “solar breeze”. This idea, although somewhat ahead of its time, illustrates that Kepler had understood something fundamental about the physics of light. Observing that the tail of comets always point away from the sun, Kepler understood that light itself carries a momentum.

The forces from the radiation pressure of light are extremely small, and the force applied by milliwatts of light (a very bright light) is in the range of piconewtons (pN). As a consequence it took almost 400 years from Keplers observations until the radiation pressure of light was used to manipulate small objects [Ashkin, 1970]. Today this phenomenon is referred to as optical trapping and the equipment is referred to as optical tweezers.

Today commercial optical tweezers, like the JPK NanoTracker™, are becoming commercially available making optical tweezers available in an increasing number of research areas. Optical tweezers have been applied to many different biologically relevant areas, including estimating the mechanical strength and kinetic parameters of pseudoknots and hairpins [Hansen et al., 2007, Chen et al., 2007, Green et al., 2008, Li et al., 2006b, Liphardt et al., 2001], identifying kinetic barriers for unfolding of highly complex RNA structures [Onoa et al., 2003] (L-21 derivative of the *Tetrahymena thermophila* ribozyme), extracting structural equilibrium information [Green et al., 2008, Li et al., 2006b, Liphardt et al., 2002, Liphardt et al., 2001], following translation of a single ribosome [Wen et al., 2008], describing hairpin-hairpin interactions [Li et al., 2006a], describing the physical properties of several molecular motors [Adelman et al., 2002, Chemla et al., 2005, Davenport et al., 2000, Smith et al., 2001, Valentine et al., 2006], and resolving individual steps in protein folding/unfolding [Ceconi et al., 2005].

Optical tweezers have added fundamental knowledge about the mechanical properties of mRNA pseudoknots and they are particularly relevant in relation to -1 PRF due to the way in which the structure is unfolded. In optical tweezers, the structure is unfolded through force which most likely resemble the way the ribosome is expected to unfold the structure to a higher degree than e.g. thermal melting.

### 7.1 Theory of Optical Trapping

The theory of optical trapping can be explained in two regimes: the ray optics regime where the diameter of the trapped particle ( $d$ ) is much larger than the wavelength ( $\lambda$ ) of the trapping laser ( $d \gg \lambda$ ) or in the Rayleigh regime where the diameter of the trapped particle is much smaller than the wavelength of the trapping laser ( $d \ll \lambda$ ). As optical trapping is often performed in an intermediate regime where  $d \approx \lambda$ , the theory for both regimes will be introduced. For both regimes it is assumed that the trapping laser has a gaussian intensity profile.

### 7.1.1 The Ray Optics Regime

When the diameter of the trapped particle is much larger than the wavelength of the trapping laser, classical ray optics apply. In this regime the trapped particle can be treated as a lens which causes the light of the trapping laser to be reflected and refracted in accordance with Snell's law. As each photon of the trapping laser contains a momentum, any reflection or refraction will result in the transfer of momentum from the trapping laser to the trapped particle. This introduction will provide only a limited and conceptual description of the forces acting on the trapped particle, however, a thorough description is provided by A. Ashkin [Ashkin, 1992].

Lateral trapping is achieved through an intensity gradient in the profile of the trapping laser. Typically, a gaussian intensity profile is used which creates a high photon density in the center of the beam and a lower photon density at the edges of the beam. This profile will result in a uneven transfer of momentum to the trapped particle as illustrated in Figure 12. The transverse change in momentum of the light is greater for the high intensity area of beam center compared to the low intensity edge. This results in a net force driving the trapped particle towards the center of the beam.

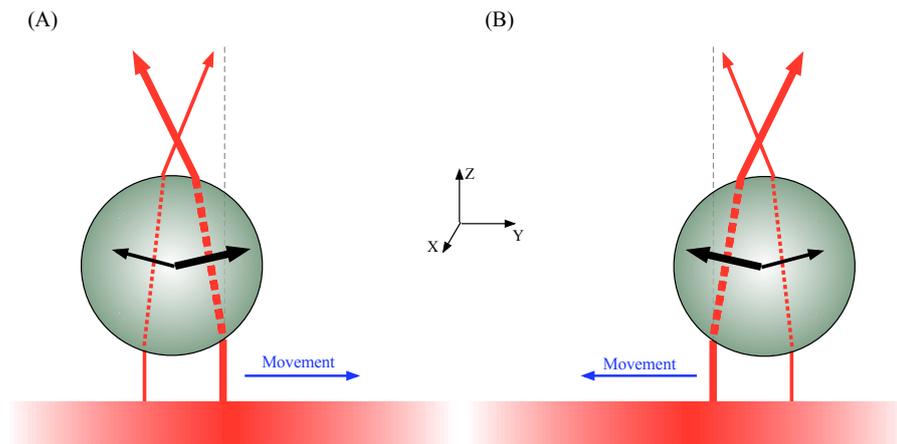


Figure 12: Lateral trapping. If the trapped particle is displaced to either side of the beam center, the gaussian intensity profile will result in a net force driving the bead towards the center of the beam (vertical dashed line). Trapping light (red arrow, thickness illustrate intensity), impulse (black arrow), and bead movement (blue arrows). Figure adopted from [Molloy and Padgett, 2002]

Axial trapping is achieved by using a trapping objective with a high numerical aperture. This results in a tightly focused beam which creates a gradient that drives the bead towards the focus for displacement either above or below the focal point. This is illustrated in Figure 13.

If the particle is too far *above* the focal point (Figure 13B), the restoring force which drives the particle towards the focal point will not be able to overcome the scattering force (see below) and the particle will be pushed away from the focal point. In practice this means that the when trapping beads one must make sure that the bead is *below* the focal

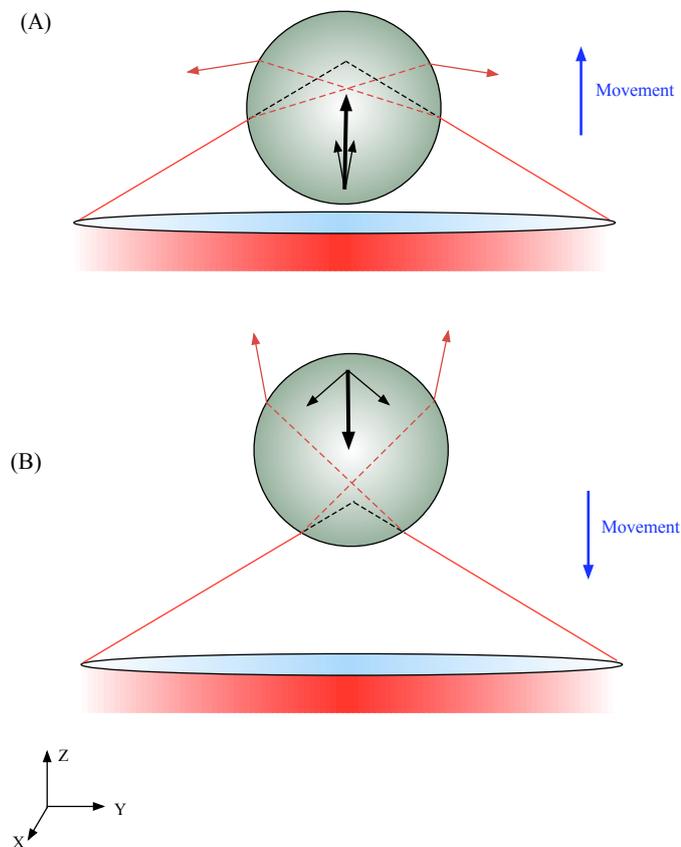


Figure 13: Axial trapping is achieved through the tight focus of the trapping beam. If the bead is placed either below the focal point (A) or above the focal point (B) the focused beam will drive the trapped particle towards the focal point. Trapping objective focus the incoming laser beam (red) and the refraction creates a net force on the particle (black) which drives it towards the focal point. Particle movement is indicated with blue arrows. Focal point is the intersection between dashed black lines. Figure adopted from [Molloy and Padgett, 2002]

point prior to trapping (Figure 13A).

### 7.1.2 The Rayleigh Regime

Due to the particle's small size relative to the wavelength of the trapping laser, it will “feel” the oscillating electromagnetic field, and the trapped particle is considered to be a point dipole which interacts with the electromagnetic trapping laser - explaining why the trapped particle must be dielectric.

Two forces are considered in this regime; the *scattering force* ( $F_s$ ) and the *gradient force* ( $F_g$ ). The scattering force, which describes the momentum transferred to the induced dipole per unit of time, can be described by [Svoboda, 1994]:

$$F_s = \frac{\langle S \rangle}{c} \frac{128\pi^5 r^6 \eta_m^5}{3\lambda^4} \left( \frac{m^2 - 1}{m^2 + 2} \right)^2 \quad (1)$$

where  $\langle S \rangle$  is the time average Poynting vector,  $c$  is the speed of light in vacuum,  $r$  is the bead radius,  $\lambda$  is the wavelength of the trapping laser,  $\eta_m$  is the refractive index of the media, and:

$$m = \frac{\eta_p}{\eta_m} \quad (2)$$

where  $\eta_p$  is the refractive index of the trapped particle. From equation (1) it follows that the scattering force is proportional to the magnitude of the time average Poynting vector, which is the time average of the cross product of the electric field vector and the magnetic field vector for the trapping laser. Consequently,  $F_s$  is proportional to the intensity of the trapping laser and acts in the direction of propagation of the trapping laser (it will push the particle in the direction of laser light propagation).

To explain why the particle is trapped by the laser, the gradient force must be considered

$$F_g = \frac{\eta_m^2 r^3}{2} \left( \frac{m^2 - 1}{m^2 + 2} \right) \nabla E^2 \quad (3)$$

where  $\nabla E$  is the electric field gradient, i.e. the gradient force is proportional to the gradient of the trapping laser. Thus, in the Rayleigh regime, trapping in all directions is explained by equation (3) in the case where  $F_g$  is larger than  $F_s$ .

## 7.2 The JPK NanoTracker™

The focus in the follow sections will be on the JPK NanoTracker™ which is the first and only commercially available optical tweezers used by the Optical Tweezers Group at the Niels Bohr Institute. Being a commercial product, it is somewhat different from the other optical tweezers available to the group which have been constructed by group members. A consequence of the commercial nature of the JPK NanoTracker™ is that some details about its construction are unavailable to us. However, some information can be extracted from the technical sales material as shown in Figure 14 [JPK Instruments, 2011].

From Figure 14 and A. Wozniak et al. a number of features of the JPK NanoTracker™ are apparent [Wozniak et al., 2009, JPK Instruments, 2011]:

- The setup is based on an inverted microscope with the light source on top and laser feed through the bottom.
- Two individually steerable traps, originating from a single 3W 1064 nm laser, are available for manipulation and force measurements. Steering the  $x$ - and  $y$ -direction is accomplished through galvanometric mirrors while steering the  $z$ -direction is achieved through a movable lens.

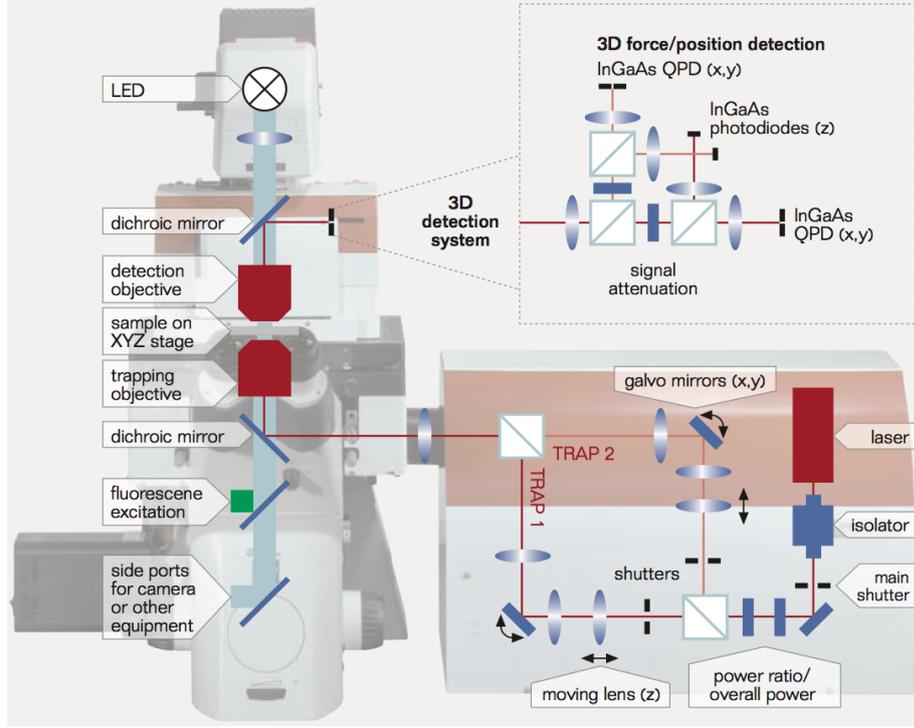


Figure 14: Schematic illustration of the JPK NanoTracker™ [JPK Instruments, 2011].

- The trap stiffness of the two traps can be adjusted independently of each other by adjusting the polarization and intensity of the single input laser beam.
- The detection system, in the back focal plane, allows for position detection in all three dimensions. Position detection in the  $xy$ -plane is accomplished by one quadrant photo diodes (QPD) while detection in the  $z$ -direction is accomplished by another QPD.

### 7.3 Calibration

The objective of the calibration is to obtain two parameters, one ( $\beta$ ) which allows the conversion of measured signal ( $V_{QPD}$ ) into metric distance ( $x$  [m]) and one ( $\kappa$ ) which allows the conversion of metric distance into exerted force ( $F$  [N]):

$$x = \beta V_{QPD} \quad (4)$$

$$F = \kappa x \quad (5)$$

There are several different methods but the main attention in this thesis will concern the power spectral density (PSD) analysis and the Stokes calibration method.

### 7.3.1 Calibration by Power Spectrum Analysis

The most commonly used calibration method relies on power spectral analysis. This method utilizes the fact that the trap will dampen the low frequency Brownian motions of the trapped particle. The motion of a trapped particle (in one dimension) can be described by the Langevin equation [Berg-Sørensen and Flyvbjerg, 2004]

$$m\ddot{x} + \gamma\dot{x}(t) + kx(t) = \sqrt{2k_B T \gamma} \mathcal{F}(t) \quad (6)$$

where  $m$  is the particle mass,  $\gamma$  is the drag coefficient ( $\gamma = 3\pi\eta d$ ),  $x$  is the bead position,  $k$  is the spring constant for the trap, a dot indicates time derivative, and  $\sqrt{2k_B T \gamma} \mathcal{F}(t)$  is a random Gaussian process that represents Brownian motion ( $\langle \mathcal{F}(t) \rangle = 0$ ). In equation (6) the mass and moment of inertia ( $m\ddot{x}$ ) can be omitted as viscous forces are dominant. This is valid only when the Reynolds number is small ( $\mathbf{R} \ll 1$ ) as in the case of micrometer sized polystyrene beads in water. An example of the time series of a trapped particle is shown in Figure 15 (experimental), where the expected Gaussian distribution of visited bead positions is observed.

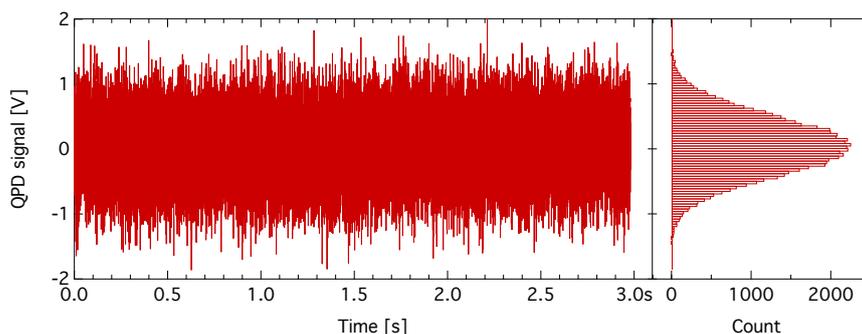


Figure 15: Raw QPD signal of a trapped particle with a diameter of  $2.1 \mu\text{m}$  (left panel). Histogram of visited positions (right panel).

By omitting the moment of inertia from equation (6) it can be expressed as

$$\dot{x}(t) + 2\pi f_c x(t) = \sqrt{2D} \mathcal{F}(t) \quad (7)$$

where the corner frequency  $f_c$  and diffusion constant  $D$  have been introduced.

$$f_c \equiv \frac{\kappa}{2\pi\gamma} \quad (8)$$

$$D \equiv \frac{k_B T}{\gamma} \quad (9)$$

The time ( $x(t)$ ) series can be Fourier transformed to the frequency domain ( $X(f)$ ). The one sided PSD ( $S(f)$ ) is defined as the squared modulus of the Fourier transform and has a Lorentzian shape [Berg-Sørensen and Flyvbjerg, 2004]:

$$S(f) = 2|X(f)|^2 = \frac{k_B T}{\gamma \pi^2 (f^2 + f_c^2)} \quad (10)$$

By insertion of  $D_v$ , the diffusion constant measured in  $V^2/s$ , we obtain:

$$S_v(f) = \frac{D_v}{\pi^2 (f^2 + f_c^2)} \quad (11)$$

By fitting equation (11) to the experimental PSD it is possible to obtain estimates of  $D_v$  and  $f_c$  as shown in Figure 16. From the estimate of the corner frequency it is possible to estimate the spring constant  $\kappa$  (from equation (8)):

$$\kappa = 2\pi\gamma f_c \quad (12)$$

The experimental PSD ( $S_v(f)$ ) has units of  $[V^2/\text{Hz}]$  and the diffusion constant ( $D_v$ ) has units of  $[V^2/s]$ . We can estimate the conversion factor  $\beta$   $[\text{m}/\text{V}]$  by comparing  $D_v$  with  $D$   $[\text{m}^2/\text{s}]$ .

$$D_v \beta^2 = D \Leftrightarrow \beta = \sqrt{\frac{D}{D_v}} = \sqrt{\frac{k_B T}{\gamma D_v}} \quad (13)$$

This completes the calibration and the two desired calibration parameters are given in (12) and (13). This method of calibration requires nothing more than the simple recording of a time series of position data similar to the one shown in Figure 15. Figure 16 shows equation (11) fitted to a PSD in order to obtain the calibration parameters.

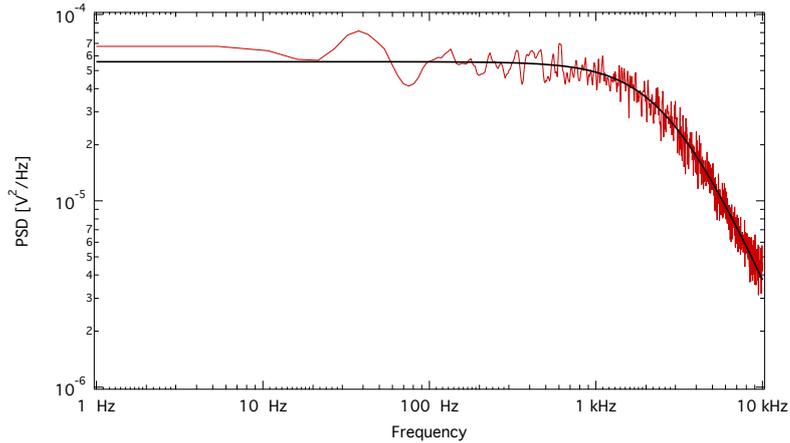


Figure 16: Power spectral density of data shown in Figure 15 (red) and fitted equation (11) (black). Estimated parameters;  $f_c = 2971$  Hz and  $D_v = 3989$   $V^2/s$ , which correspond to a spring constant  $\kappa_x = 0.329$  pN/nm and a conversion factor  $\beta = 7.65$  nm/V (bead diameter  $2.1\mu\text{m}$  @ 298K)

### 7.3.2 Stokes Calibration

Stokes calibration (or drag force calibration) uses Stokes law to find a parameter ( $\xi$ ) which relates the QPD output signal ( $V_{QPD}$ ) to the force exerted on the trapped particle ( $F$ ):

$$\xi V_{QPD} = F \quad (14)$$

To estimate  $\xi$ , a known force is applied to the trapped particle while the QPD signal is recorded. The force is applied to the trapped particle by inducing drag through a laminar flow which can originate either from a pump or from a controlled movement of the sample chamber relative to the trapped particle. The drag force ( $F_d$ ) is determined by Stokes law

$$F_d(v) = 3\pi\eta_m d v = \gamma v \quad (15)$$

where  $\eta_m$  is the dynamic viscosity of the media,  $d$  is the diameter of the trapped particle, and  $v$  the velocity of the flow.

From equation (14) and equation (15) we obtain:

$$\xi V_{QPD} = \gamma v \Leftrightarrow \xi = \frac{\gamma v}{V_{QPD}} \quad (16)$$

The QPD signal caused by a  $2.03 \mu\text{m}$  bead under a forward/reverse oscillating flow of  $150 \mu\text{m/s}$  is shown in Figure 17. In the absence of flow, the QPD signal rests at 0 V while the signal changes to approximately  $\pm 0.15$  V for each direction of flow.

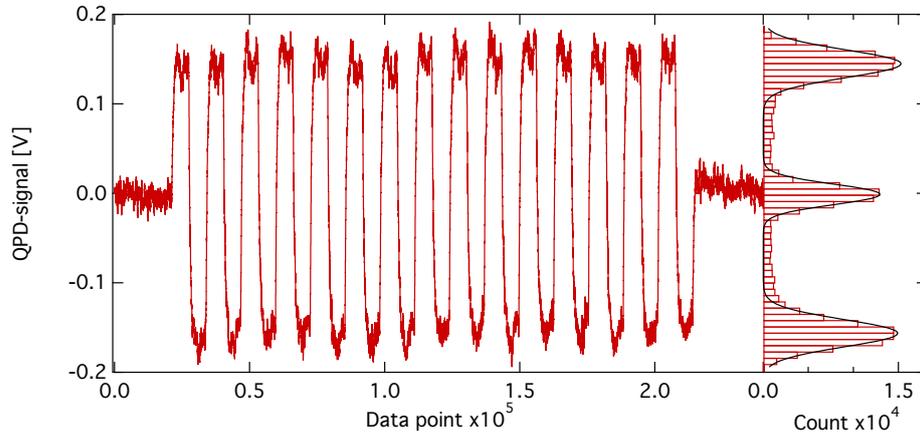


Figure 17: Stokes Calibration. QPD signal from a trapped  $2.03 \mu\text{m}$  bead subjected to a flow of  $150 \mu\text{m/s}$  with oscillating direction (left panel). The signal has been smoothed using a Savitzky-Golay filter with a width of 401 points (sample rate was 10 kHz). Histogram of visited positions (right panel) with fitted triple Gaussian distribution (black). From the triple gaussian the difference in voltage between the two flow directions was estimated to be 0.301 V, yielding a conversion factor  $\xi = 17.0 \text{ pN/V}$ .

### Further considerations

Stokes calibration requires an accurate knowledge of the force applied to the trapped particle and several factors should be addressed if this method of calibration is employed in practice. One of these factors concerns the distance between the trapped particle and the surfaces of the sample chamber. Equation (15) can be corrected for these surface interactions by substituting  $F_d$  for  $F'_d$  [Svoboda, 1994]:

$$F'_d(v, h) = \frac{6\pi\eta r v}{1 - \frac{9}{16} \left(\frac{r}{h}\right) + \frac{1}{8} \left(\frac{r}{h}\right)^3 - \frac{45}{256} \left(\frac{r}{h}\right)^4 - \frac{1}{16} \left(\frac{r}{h}\right)^5} \quad (17)$$

where  $v$  is the velocity of the flow,  $r$  is the particle radius, and  $h$  is the distance between the surface and the center of the trapped particle. By comparing  $F_d(v)$  with  $F'_d(v, h)$  it is possible to calculate how the actual drag force will deviate from the calculated drag force when ignoring the surface interactions. This is shown in Figure 18 where the error in  $F_d(v)$  is plotted as a function of the distance to the surface for a particle with a diameter of 2  $\mu\text{m}$ .

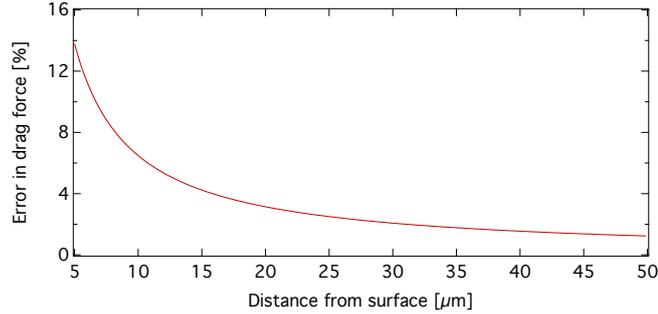


Figure 18: Error in uncorrected drag force. Error of  $F_d(v)$  as a function of distance to surface. Error is defined as  $(F'_d(v, h) - F_d(v))/F_d(v) \cdot 100\%$

From Figure 18 it is evident that if the distance between the trapped particle and the surfaces of the sample chamber is larger than 15  $\mu\text{m}$  the error will be less than 5%. For all experiments conducted here the distance to the nearest surface was more than 20  $\mu\text{m}$  and the surface interaction was neglected.

Another important factor to consider when employing Stokes calibration is the viscosity of the media in which the particle is trapped. Apart from sensitivity to the temperature the viscosity is also sensitive to the concentration of salts in the buffer. One way to estimate how the salts affect the viscosity of water is employ the Othmer rule [Fabuss and Korosi, 1969, Korosi, 1968]. Using this method, the expected change in dynamic viscosity in the buffer used in this study is less than 3% (calculations behind this change is provided in Appendix C (page 157)). Such a small change can be neglected during the calibration.

Stokes calibration also requires accurate knowledge of the bead size and the velocity of the flow. For the experiments conducted here the standard deviation of the 2  $\mu\text{m}$  beads was approximately 0.05  $\mu\text{m}$  according to the manufacturer (batch dependent). The stage

speed was verified by video recoding of beads deposited on the glass surface of the sample chamber. Consequently, these sources of error was considered negligible.

### 7.3.3 Calibration of the JPK NanoTracker™

Calibration by power spectral analysis uses the thermal fluctuations of the trapped particle to derive the spring constant and the conversion factor as described above. This method of calibration is sensitive towards any instrumental noise that is transferred to the trapped particle.

Unfortunately, the JPK NanoTracker™ suffers from such a problem as the galvanometric mirrors add very significant vibrational noise to the trapped particles (personal communication with JPK). As these vibrations are transferred to the trapped particles, they “pollute” the recorded time-signal. To make matters worse, these vibrations are significant from 1kHz to 6kHz - the same region as the corner frequency for any practical trap stiffness. These vibrations are clearly visible in the power spectrum from the JPK NanoTracker™, as shown in Figure 20. Although the software suite associated with the JPK NanoTracker™ contains a calibration routine based on power spectrum calibration, the resulting calibration was notoriously unstable. Consequently, we developed our own calibration scheme based on both Stokes calibration and power spectrum analysis as described below.

#### Calibration Parameter $\xi$

The  $\xi$  calibration parameter was determined by Stokes calibration as described above, however, instead of using only a single drag force to determine the parameter, four different drag forces were used to create a plot of drag force as a function of QPD signal. From linear regression it is possible to estimate  $\xi$  as the slope of this graph, as illustrated in Figure 19.

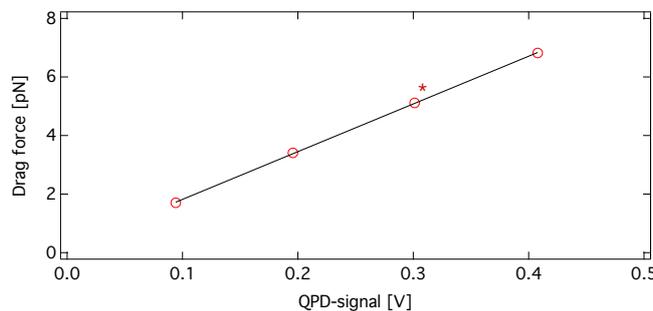


Figure 19: Drag force versus QPD-signal. By linear regression (black solid) to experimentally determined data (red circle)  $\xi$  was determined to be 16.3 pN/V. The point indicated with an asterisks is obtained from the data shown in Figure 17

### Conversion Factor $\beta$

In practice, both the low- and high frequency domain of the PSD are relatively unaffected by the noise introduced by the galvanometric mirrors. Hence, we used these two regimes of the PSD to determine the conversion factor  $\beta$ .  $\beta$  was estimated from an unweighted least-square regression of equation (11) to the PSD in the two frequency domains 50-150 Hz and 20 kHz-55kHz. This is illustrated in Figure 20 where the noise introduced by the galvanometric mirrors is clearly visible as a broad peak in the frequency range 1 kHz - 6 kHz.

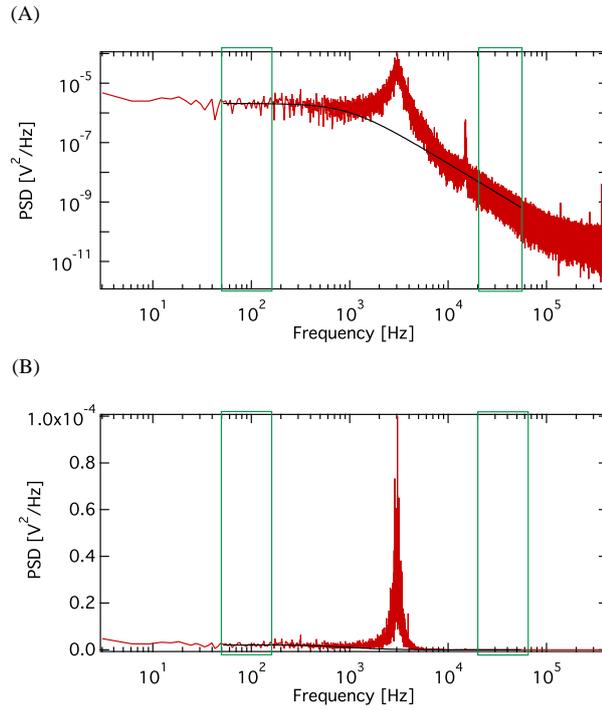


Figure 20: Power spectrum from the JPK NanoTracker™. **A**: Power spectrum with logarithmic ordinate. **B**: Power spectrum with linear ordinate. Equation (11) was fitted to the data in the green boxes (black solid) to obtain calibration parameter  $\beta$ . Sample rate was 800 kHz.

## 7.4 Kinetics

Consider the equilibrium between a folded and an unfolded RNA structure. The rate of unfolding depends upon the free energy difference between the folded structure and the transition state. Consider the same system under the presence of an applied force ( $F$ ). As the force applied favors the species with the longest extension in the direction of the force, the rate of change between the species depend upon the *distance* between the folded structure and the transition state ( $X^\ddagger$ ). Under such conditions the rate constant can be expressed by the Arrhenius like equation [Tinoco, 2004]:

$$k(F) = k_0 e^{FX^\ddagger/k_B T} \quad (18)$$

where  $k_0$  is the zero-force rate constant. Another way of looking at the situation described above is to consider at the lifetime of the folded structure. For a single molecule, transitions occurs stochastically and the change in probability that the structure is *still* folded can be described by:

$$dP = -k(F) P dt \quad (19)$$

Insertion of (18) gives:

$$dP = -k_0 e^{FX^\ddagger/k_B T} P dt \quad (20)$$

In experiments where a harmonic force is applied and the system start out in equilibrium, the force is given by time ( $t$ ) multiplied by loading rate ( $r$ ),  $F = rt$ . By integrating (20) one obtains:

$$\ln(P) = -\frac{k_0 k_B T}{r X^\ddagger} e^{rt X^\ddagger/k_B T} + \frac{k_0 k_B T}{r X^\ddagger} \Leftrightarrow r \ln(P) = -\frac{k_0 k_B T}{X^\ddagger} \left( e^{FX^\ddagger/k_B T} - 1 \right) \quad (21)$$

Thus, it is possible to estimate  $k_0$  and  $X^\ddagger$  from force-extension experiments by plotting  $r \ln(P)$  versus force.

The loading rate is determined by the pulling speed, ( $v$ ), and the effective spring constant ( $\kappa_e$ ) of the setup:

$$r = v \kappa_e Q(F, \kappa) \quad (22)$$

where  $Q(F, \kappa)$  is damping function in the interval (0,1) which corrects for the flexibility of the ssDNA/RNA handles (ssDNA, single stranded DNA). The value of  $Q(F, \kappa)$  depends on the force, the trap stiffness, and the composition of the tether. Due to a convention in the literature, we will consider  $Q(F, \kappa) = 1$  [Tinoco, 2004, Hansen et al., 2007, Green et al., 2008, Chen et al., 2007]. An estimate of  $Q(F, \kappa)$  for our tethers is derived in Appendix A (page 151) and it appears that the loading rate is overestimated by 10-20% for unfolding above 10 pN when using  $Q(F, \kappa) = 1$ .

For single trap optical tweezers  $\kappa_e = \kappa_{trap}$ , while for dual trap optical tweezers  $\kappa_e$  is more complicated (see Appendix A for details):

$$\kappa_e = \frac{\kappa_1 \kappa_2}{\kappa_1 + \kappa_2} \quad (23)$$

We can calculate the loading rate using equation (22) with the effective spring constant described by equation (23).

## 7.5 Nonequilibrium work and free energy

### 7.5.1 Jarzynski's equality

In the single molecule experiments conducted in this study, we add work to a single molecule until some desired event takes place - unfolding of RNA structures in our experiments. During this process, the parameters of the system are changed at a finite rate whereby the average work ( $\langle W \rangle$ ) required to reach this new state on average will be larger than the Helmholtz free energy difference ( $\Delta A$ ) between the two states [Jarzynski, 1997]:

$$\langle W \rangle \geq \Delta A \quad (24)$$

where angled brackets denote an average of ensemble measurements. At constant pressure and volume,  $\Delta A$  equals the change in Gibbs free energy ( $\Delta G$ ) between the two states<sup>5</sup> and will be referred to as  $\Delta G$  hereafter. The inequality (24) holds true if the path from the initial state to the end state is the same for each measurement. The difference between  $\langle W \rangle$  and  $\Delta G$  is dissipated work.

The inequality (24) can be converted into an equality known as *Jarzynski's equality*:

$$\langle e^{-W/k_B T} \rangle = e^{-\Delta G/k_B T} \quad (25)$$

The consequences of equation (25) are profound. By calculating exponential averages of the work required in going from one state to another it is possible to calculate the free energy difference between the two states. The equality is independent of both the path between the states and the rate at which the change occurs. Consequently, the Jarzynski equality provides the following free energy estimator:

$$\Delta G = -k_B T \ln \left( \langle e^{-W/k_B T} \rangle \right) \quad (26)$$

The nature of this estimator is complicated and a thorough analysis of its behavior is provided by Gore et al. [Gore et al., 2003]. F. Ritort and C. Bustamante have tried to estimate the number of transitions required to estimate  $\Delta G$  within  $k_B T$ . They estimate that the number of transitions required increases exponentially with amount of dissipated work ( $W_{diss}$ ) and that more than 1,000 transitions are required for a  $W_{diss}$  of  $5k_B T$  [Ritort and Bustamante, 2002]. The Jarzynski equality has been validated experimentally for small values of  $W_{diss}$  and gave good results [Liphardt et al., 2002]. Also, the exponential average in the Jarzynski equality makes it sensitive towards rare events, i.e. it is sensitive towards very low and very high work-events.

<sup>5</sup>A=U-TS, G=H-TS and H=U+pV. Thus, dG=dU+dpV+p dV-dTS-TdS=dU-dTS-TdS=dA

### 7.5.2 Crooks Fluctuation Theorem

The large statistical uncertainty of the Jarzynski estimator in the far from equilibrium regime for finite sample sizes makes it less applicable for estimating free energy differences for pseudoknots. Relatively large values of  $W_{diss}$ , 10-50  $k_B T$ , are typically observed for mechanical unfolding of RNA structures [Green et al., 2008, Collin et al., 2005]. Another method for estimating free energy differences is through the Crooks Fluctuation Theorem (CFT) [Crooks, 1999]:

$$\frac{P_U(\beta W_U)}{P_R(-\beta W_R)} = e^{\beta(W - \Delta G)} \quad (27)$$

where  $\beta = (k_B T)^{-1}$ ,  $P_U(W_U)$  is the probability density function (PDF) for work during unfolding,  $P_R(W_R)$  is the PDF for work during refolding,  $W$  is reversible work, and  $\Delta G$  is the Gibbs free energy.

By creating two PDFs; one for the work used during the unfolding and one for the work returned from the system during refolding, the free energy difference between the two states can be estimated from equation (27). There is no formal limit for the value of  $W_{diss}$  and the CFT should provide good estimates even far from equilibrium as long as sufficient information about the overlap of the distributions exists. The CFT have been shown to provide good estimates of the free energy for unfolding of hairpins and pseudoknot [Green et al., 2008, Collin et al., 2005].

When the free energy ( $\Delta G_T$ ) is estimated from force-extension curves, using either Jarzynski's equality or the CFT, it has contributions from several different components, such as pseudoknot unfolding ( $\Delta G_{PK}$ ), stretch of RNA/DNA handles and stretch of ssRNA from the unfolded pseudoknot. Thus, these contributions must be subtracted from the total free energy difference in order to obtain the free energy of the unfolded structure ( $\Delta G_{PK}$ ):

$$\Delta G_{PK} = \Delta G_T - W_{corr} \quad (28)$$

where  $W_{corr}$  describes the work required for stretching of handles, ssRNA outside the structure, and ssRNA of the unfolded structure. The implementation of the CFT in this thesis is described in detail in Appendix D (page 159).

## 8 Materials and Methods

All solutions used in the single molecule experiments were filtered through a 0.2  $\mu\text{m}$  filter (Sartorius Stedim, Minisart, 0.2  $\mu\text{m}$ , cat. 16534K) before use.

### 8.1 Buffers

Buffer R	- 10 mM Tris-HCl pH 7.5 - 250 mM NaCl - 10 mM MgCl <sub>2</sub>
TE	- 10 mM Tris-HCl pH 8 - 1 mM EDTA

### 8.2 DNA oligonucleotides

Name	Sequence (5' → 3')	Use
TH408	ATAATTCGCGTCTGGCCTTC	R primer for downstream handle (5'-dig)
TH412	ATAATTCGCGTCTGGCCTTC	R primer for T7 RNAP template
TH414	TAATACGACTCACTATAGGGAGA GTATACCTCTCAGTTGGGTG	F primer for T7 RNAP template
oJT_1	TGAATCCGCGGTACCAGCAC	R primer for upstream handle (5'-Amino CX)
oJT_2	CTAATTCAGTGGCCGTCGTT	F primer for downstream handle (5'-phosphate)
oJT_3	GTATACCTCTCAGTTGGGTG	F primer for upstream handle (5'-phosphate)

“F” = Forward, “R” = Reverse. All TH-primers are from ref [Hansen et al., 2007]. All oligonucleotides were store in water at -20°C.

### 8.3 DNA Template for *in vitro* transcription

The DNA template was created through PCR using either DreamTaq™ or *Pfu* DNA polymerase (Fermentas, EP0702 or EP0502) according to manufactures recommendations. Primers were TH414 and TH412 and template was a pOFX302 derivative containing the pseudoknot-sequence of interest [Hansen et al., 2007]. PCR product was purified using GeneJet™ PCR purification kit (Fermentas, K0702).

### 8.4 Preperation of pseudoknot containing RNA

Run-off transcription using T7 RNA-polymerase was used to create pseudoknot containing RNA for single molecule experiments. The following protocol was used (final concentrations): 1x Transcription buffer (Fermentas), 2 mM rNTP mix (2 mM for each nucleotide, New England Biolabs or Fermentas),  $\approx$  1.3 ng/ $\mu\text{l}$  template DNA,  $\approx$  60 U RNasin (Promega) or RiboLock (Fermentas), and 40 U T7 RNA-polymerase (Fermentas)

Transcription was performed at 37°C for 2-3 hours after which the RNA was purified using phenol/chloroform extraction and ethanol precipitation: 150  $\mu\text{l}$  TE was added to the transcription reaction, RNA was extracted using 1 volume phenol and 1/2 volume of chloroform, precipitation at -20°C for 40 minutes with 200 mM NaCl, pelleted at 20,000 $\times$ g at 0°C for 1 hour, was washed with 100  $\mu\text{l}$  70% ethanol, and air dried. Finally the RNA was dissolved in autoclaved water and stored at -20°C for later use.

### 8.5 Single stranded DNA handles

The single molecule experiments required two types of single stranded DNA handles, one 5'-digoxigenin labeled handle and one 3'-biotin labeled handle. The dsDNA template was generated through a PCR reaction using TH408 and oJT\_2 (for the 5'-digoxigenin labeled ssDNA handle) and oJT\_1 and oJT\_3

(for 3'-biotin labeled ssDNA handle). The PCR product was purified using GeneJET™ PCR Purification Kit (Fermentas, #K0702).

### 5'-digoxigenin labeled ssDNA handle

The purified PCR product was digested with  $\lambda$ -exonuclease (NEB, M0262) in 100  $\mu$ l and purified using phenol/chloroform extraction in combination with size exclusion gravity columns (illustra NAP5, GE Healthcare): 1x  $\lambda$ -exonuclease buffer, 10  $\mu$ g dsDNA, 40 U  $\lambda$ -exonuclease. Incubation at 37°C for 30 minutes, 77°C for 10 minutes. DNA was extracted with 1 volume phenol and 1/2 volume chloroform. Supernatant was transferred to a NAP5 columns equilibrated with autoclaved water and DNA was eluted with 500  $\mu$ l autoclaved water. Concentration of DNA was quantified using the Thermo Scientific NanoDrop 1000

### 3'-biotin labeled ssDNA handle

The purified dsDNA template was biotinylated using terminal deoxynucleotidyl transferase (Fermentas, #EP0161) and Biotin-11-dUTP (Fermentas, #R0081): 1x Terminal deoxynucleotidyl transferase buffer, 72  $\mu$ M Biotin-11-dUTP, 10  $\mu$ g dsDNA template ( $\approx$  70 pmol 3'-ends), 60 U Terminal deoxynucleotidyl transferase. Incubation at 37°C for 30 minutes. The biotinylated dsDNA was subjected to  $\lambda$ -exonuclease digestion and size exclusion purification as described for the 5'-digoxigenin handle above.

The quality of the ssDNA purified handles was estimated from an ethidium bromide stained agarose gel and they were stored at -20°C.

## 8.6 RNA/ssDNA hybridisation

The RNA and ssDNA were hybridized in the following manner in a final volume of 40  $\mu$ l: 35  $\mu$ l 1x Buffer R,  $\approx$  60 ng 3'-biotin ssDNA,  $\approx$  60 ng 5'-digoxigenin ssDNA,  $\approx$  40 ng RNA. The mix was heated to 75°C for 6 minutes in a heating block after which the block was placed at 5°C. After 35-45 minutes the temperature was  $<$  15°C and the ssDNA/RNA hybrid was stored at -20°C. The quality of the hybridization was evaluated from a ethidium bromide stained agarose gel as described in [Hansen et al., 2007].

## 8.7 Sample chamber

The sample chamber was made from a microscope slide (Menzel-Gläser, 25 x 50 mm, #1.5) and a cover slip (VWR, 18 x 18 mm, #1, cat. 631-1567) separated by double sticky tape (tesa, 12 mm, cat. 57912)). The tape was cleaved in such a way that the walls of the chamber consisted of 6 mm tape. The height of the sample chamber was in the region of 100-150 nm and its volume was approximately 25-30  $\mu$ l.

## 8.8 Pulling experiments

The bead and ssDNA/RNA mix was prepared: 5  $\mu$ l 1x Buffer R, 0.3-0.5  $\mu$ l 0.5% (w/v) 3.05  $\mu$ m streptavidine coated beads (Spherotech, SVP-30-5), 0.3-0.5 ng ssDNA/RNA hybrid. Incubation for 30-45 minutes at room temperature after which the following was added: 250  $\mu$ l 1x Buffer R and 0.5-0.8  $\mu$ l 0.1% (w/v) 2.03  $\mu$ m or 2.17  $\mu$ m anti-digoxigenin coated beads (Spherotech, DIGP-20-2). This suspension was injected into a freshly made sample chamber and the experiment was conducted as follows.

1. The bottom of the sample chamber was located
2. The Köhler illumination was adjusted
3. Laser power was set to 200 mW and power sharing was set to 25%/75% (trap1/trap2)
4. A 2.03  $\mu$ m bead and a 3.05  $\mu$ m bead was trapped in trap1 and trap2 respectively
5. The beads were held near the bottom for approximately 5 minutes while allowing other beads to precipitate to the bottom of the chamber

6. A clean area (no beads on bottom) was located and the trapped beads were moved up to a height of approximately 20-25  $\mu\text{m}$  from the bottom.
7. The laser power was raised to 2300 mW
8. A power spectrum was recorded for each direction (x, y and z) through the JPK NanoTracker™ software (average of 10 scans) and saved as text-files (these files contains frequency- and PSD data)
9. Drag-force calibration was conducted by move the piezo 95  $\mu\text{m}$  at 50  $\mu\text{m/s}$ , 100  $\mu\text{m/s}$ , 150  $\mu\text{m/s}$  and 200  $\mu\text{m/s}$  while recoding the QPD signal at 10 kHz
10. The beads were aligned in the x-direction and brought closer together in the y-direction
11. The beads were brought closer in the y-direction in increments of 20 nm while checking if a tether had formed - if a tether had formed i.e. the beads appeared to be linked together - a pulling experiment was conducted:
  - The QPD signal and trap positions where sampled at 10 kHz
  - (a) The bead in trap2 was moved 0.5-0.8  $\mu\text{m}$  away from trap1 at 100 nm/s in the y-direction
  - (b) The bead in trap2 was moved back to the starting position
  - (c) The tether was pulled repeatedly until it broke

## 8.9 Data Analysis

All data from single-molecule experiments were analyzed using custom made software written in iGor Pro (<http://www.wavemetrics.com/>). The software converted the recorded QPD-signal and trap-position into a force-extension curve as described in appendix B.

The constructed force-extension curves were filtered in order to reduce the probability of accepting invalid traces for later analysis. The filter was subjective, to say the least, but there is no true objective way of validating a force-extension curve. Only force-extension curves complying to the following rules were used: Initial slope must be close to zero and curve must contain both unfolding and refolding events.

## 8.10 Estimation of expected rip-lengths

Expected rip-lengths can be estimated by finding the appropriate roots for the EWLC for the force immediately before the rip. However, this method will overestimate the rip-length as it fails to account for the change in exerted force during the unfolding process. A more precise estimate can be obtained by computer simulations of the experiments where the decrease in exerted force during the unfolding is taken into account. The details of these simulations are shown in Part IV and the parameters used is shown in Table 2.

Paramter	DNA/RNA hybrid	ssDNA/ssRNA
Persistence length ( $L_p$ )	45 nm	1 nm
Contour length ( $L_c$ )	0.28 nm/bp	0.59 nm/nt
Stretch modulus ( $K$ )	1000 pN	800 pN

Table 2: Tether parameters. Parameters used in all simulations of optical tweezer experiments. RNA/DNA hybrid was treated as dsDNA and ssRNA was treated as ssDNA [Wang et al., 1997, Liphardt et al., 2001, Milman and Langridge, 1967, Smith et al., 1996]

## 9 Results

The analysis of single molecule experimental data is complicated, and the following section contains a lot of information. The complexity arises from the many possible structures, folding intermediates and unfolding pathways. In the end of this section, a summary of all data is listed and through the discussion I will try to combine all the information into a more readable format.

For each pseudoknot investigated, an ssDNA/RNA hybrid was made as described in materials and methods in which the pseudoknot containing RNA was annealed to two labelled ssDNA handles. This construct was suspended between two trapped polystyrene beads in the JPK NanoTracker™ and the mechanical investigation was conducted. The setup is shown in Figure 21 (not drawn to scale).

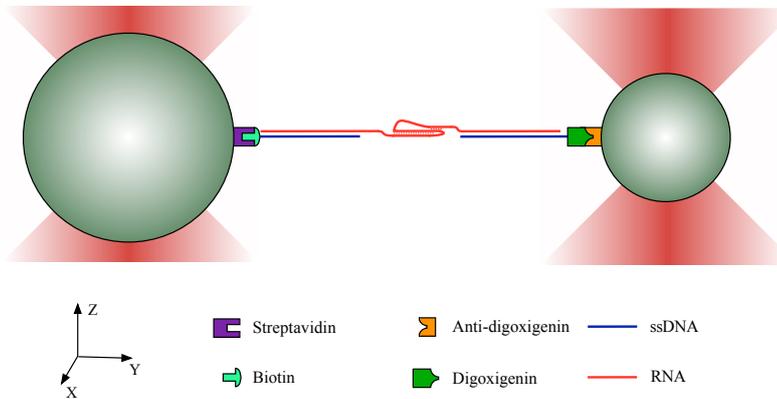


Figure 21: Schematic representation of the setup in which the mechanical properties of our pseudoknots were investigated. Two polystyrene beads (green) were trapped in two separate laser beams. One bead was coated with streptavidin (magenta) while the other was coated with  $\alpha$ -digoxigenin (orange). The ssDNA/RNA hybrid (blue and red respectively) was attached between the beads via biotin (turquoise) and digoxigenin (green) on the ssDNA handles. The figures is not drawn to scale as the beads are approximately  $2 \mu\text{m}$  and  $3 \mu\text{m}$  while the ssDNA/RNA has a length of approximately  $0.3 \mu\text{m}$

In the following section the term “extension” and “retraction” will be used to describe the two domains of a force-extension experiment in which the bead-bead distance is increased or decreased, respectively. Theoretical unfolding lengths for our various RNA structures are calculated through the simulation framework described in Part IV with no  $z$ -level misalignment and the average trap stiffness from the experiments for each pseudoknot.

### 9.1 Validation of Conversion Factor $\beta$

To estimate the accuracy of the estimate of calibration parameter  $\beta$ , we recorded the QPD signal while moving the trapped particle a known distance within the optical trap. This is not entirely trivial. One way to move the trapped particle by a known distance is to

immobilize it on the surface of the sample chamber and move the piezo stage to drive the bead away from the optical trap. This procedure does, however, suffer from a number of uncertainties: the bead must be absolutely stable relative to the surface of the chamber as any movements would hamper the validation and the bead must be located at the same  $z$ -position within the optical trap as it would if it were trapped in solution.

Another way to validate the estimate of  $\beta$  is to trap two beads (one in each trap), perform the calibration to estimate  $\beta$ , push the trapped beads together in order to get them to stick inelastically to one another, and move one trap relative to the other trap. This way, we know that if the beads are stuck together, the sum of the bead movement in both traps must equal the distance we move one of the traps relative to the other. If we increase the trap-trap distance by 100 nm and the bead-bead distance is constant, the sum of the bead movements in the two traps must equal 100 nm - if our estimate of  $\beta$  is correct.

This was performed for three sets of beads in three different sample chambers and the sum of the recorded bead movement was compared with the relative trap-trap distance. The result of this analysis is shown in Figure 22A.

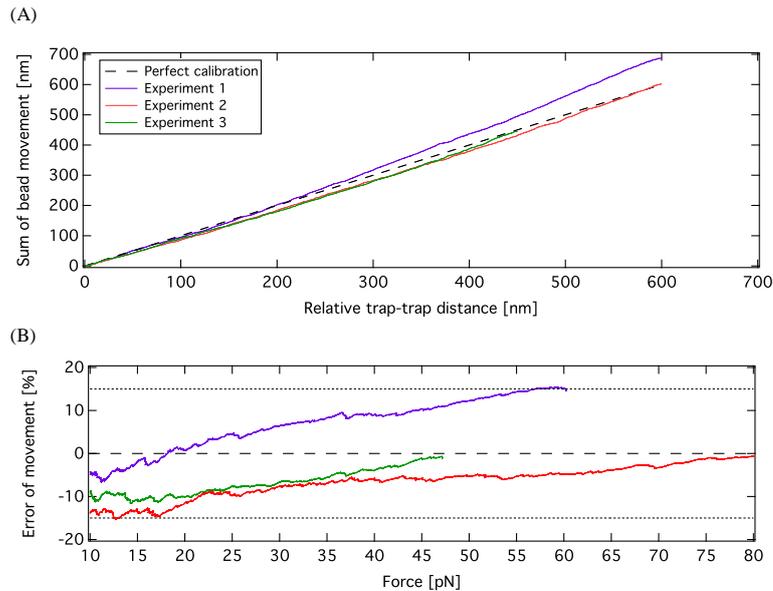


Figure 22: Error analysis of  $\beta$ . **A**: Analysis from three independent experiments where two beads were trapped (one in each trap). The two beads were pushed together to make them stick to one another (inflexible link). One of the traps was moved the “Relative trap-trap distance” away from the other and the sum of the bead displacement in both traps was calculated from the QPD signal and the two  $\beta$  values. For a perfect calibration (dashed line) the sum of bead movement should equal the relative trap-trap distance. **B**: Estimation of error from (A). The difference between the sum of bead movement and relative trap-trap distance was calculated and plotted as a function of force (calculated using  $\xi$  and the QPD signal).

In Figure 22B the error of the movement is plotted as a function of force. From the

figure it is clear that the error of recorded movement is within 15% for any force above 10 pN. This level of uncertainty is acceptable for our investigation and consequently we must expect that any unfolding lengths (e.g. from unfolding of RNA structures) are within 15% of the theoretical estimate.

Interestingly, the non-zero slope of the errors in Figure 22B indicate that additional factors, other than an error in  $\beta$ , may contribute to the overall error. One such factor could be “cross-talk” between the two traps due to their close proximity. This could possibly be caused by one trap affecting the bead in the other, or by light from one trap being recorded by the positioning system of the other trap.

## 9.2 Instrumental Noise

From the PSD shown in Figure 20 (page 61) it is clear the trapped particles are subjected to JPK NanoTracker™-induced vibrational noise over a broad range of frequencies. The PSD, however, does not provide information about the absolute movement of the trapped particles caused by these vibrations. To assess how the noise affects the absolute fluctuations of the trapped particles, the actual fluctuations were compared with the expected fluctuations. Using a Monte Carlo simulation as described by Gong et al. the theoretical Brownian motion of a particle in a harmonic potential was simulated [Gong et al., 2006]. This was compared to the actual positions of the particle detected by the JPK NanoTracker™. The result for one of our experiments is shown in Figure 23, where the consequences of the noise is clearly visible.

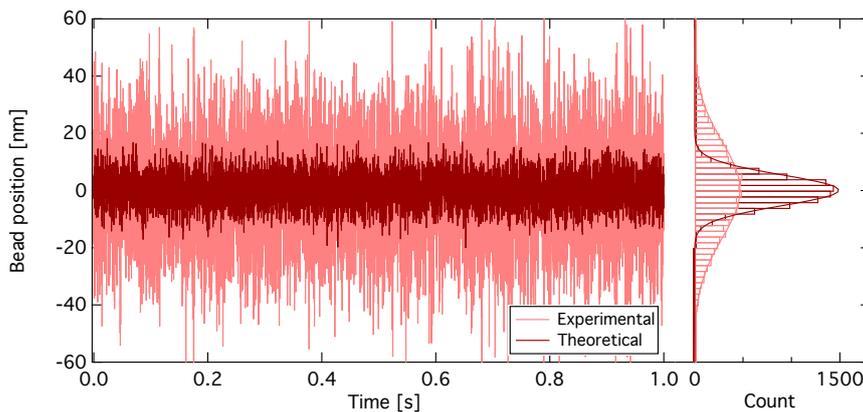


Figure 23: Brownian motion in harmonic potential. Actual bead positions detected by the JPK NanoTracker™ (light red) and theoretical bead positions from Monte Carlo simulation (dark red). Visited positions is shown in histogram (right panel). Sample rate was 10 kHz for experimental data. Theoretical positions was calculated for intervals of  $10^{-8}$  seconds and subsequently sampled at 10 kHz [Gong et al., 2006].

The standard deviations of bead movements in Figure 23 are:

$$\sigma_{Exp} = \sqrt{\sigma_{Exp}^2} = \sqrt{286 \text{ nm}^2} = 16.9 \text{ nm}$$

$$\sigma_{The} = \sqrt{\sigma_{The}^2} = \sqrt{29 \text{ nm}^2} = 5.4 \text{ nm}$$

where  $\sigma_{Exp}$  and  $\sigma_{The}$  is the experimental and theoretical standard deviation respectively. The theoretical estimate is in excellent agreement with the equipartition theorem which predicts  $\sigma_{The} = 5.3 \text{ nm}$  for the relevant trap stiffness [Molloy and Padgett, 2002]. Thus, it seems that the noise from the galvanometric mirrors increases the standard deviation of bead movement by a factor of three.

Although these vibrations made the calibration difficult, they do not pose an imminent problem for our mechanical unfolding experiments. By choosing a sufficiently wide window for our smoothing algorithm, we can still measure the position of the trapped particle with a sufficiently high degree of accuracy. This is also illustrated in the good agreement with experimental and theoretical unfolding lengths for our RNA structures presented later (Table 4, Table 11, and Table 16).

The level of noise in our force-extension curves dictate how short unfolding events we will be able to reliably detect. From the two representative force-extension curves shown in Figure 24, it appears that we are able to reliably detect unfolding events which increase the length of the tether by more than 5 nm.

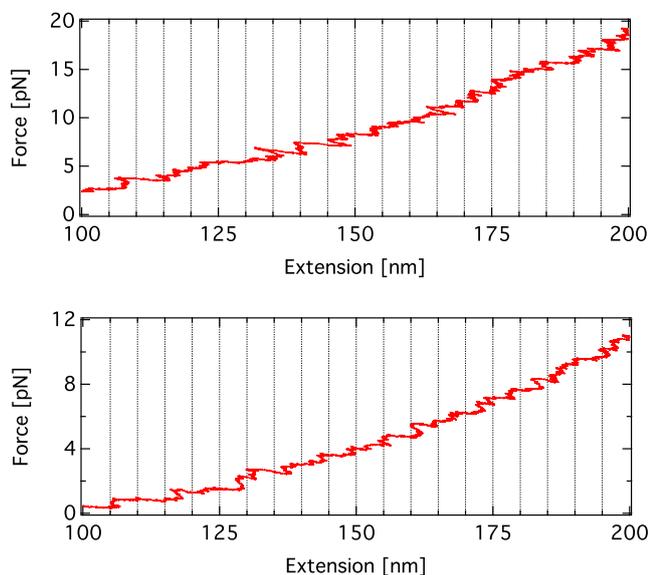


Figure 24: Force-Extension curves - Noise. Two representative force-extension curves with 5 nm grid lines on the abscissa. Reliable detection of unfolding events will require unfolding events longer than 5 nm.

### 9.3 Negative Control

Due to the rather crude definition of a tether as some “connection” which creates a mechanical coupling between the two trapped particles it is important to investigate how the presence of a pseudoknot changes this coupling.

First, the ability of the  $\alpha$ -digoxigenin-beads and the streptavidine-beads to form unspecific tethers was limited. Although no direct experimental evidence is provided, it was observed each time a new batch of ssDNA/RNA was to be used. At low concentrations of ssDNA/RNA, tethers between the trapped beads were very rare and the few unspecific tethers which did form between the trapped beads was significantly less flexible and unpredictable than DNA/RNA tethers and thus relatively easy to remove from experimental data. An example of a force-extension curve for such an unspecific tether is shown in Figure 25. This figure also confirms that our estimate of  $\beta$  is sufficiently accurate for the experiments conducted here.

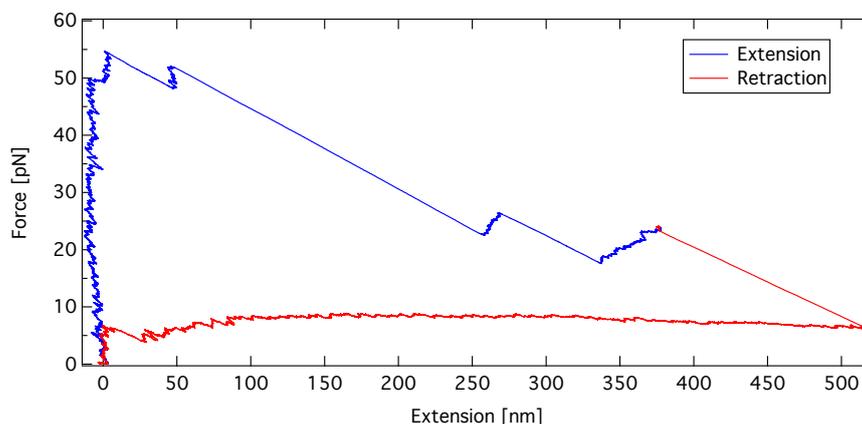


Figure 25: Force-extension curve of unspecific tether

Second, to ensure that the unfolding and refolding events obtained from our pseudoknot-containing constructs truly reflect physical properties of the RNA structures, a negative control was used in which the RNA did not contain a pseudoknot. The control used was the PK421 construct in which the pseudoknot sequence was replaced with 5’-GCGC-3’ [Hansen et al., 2007]. Using this construct, a total of 214 pulls were made on a DNA/RNA tether of which 2.3% showed distinct unfolding events somewhat similar to those expected for unfolding RNA structures. However, none of these “false” traces exhibited any refolding events. Based on these results, it was decided that only force-extension curves containing both unfolding and refolding events should be accepted in the analysis of the mechanical properties of our pseudoknots.

### 9.4 Pulling Statistics

Several of the pseudoknots used for *in vivo* experiments were subjected to mechanical unfolding, of which some were conducted by Rebecca Bolt Ettliger [Ettliger, 2012] as

part of her Master thesis work. A brief summary of the experimental output from our unfolding experiments are listed in Table 3.

	Pseudoknot						
	10/6	11/6	6/11	16/6	22/6a	22/6b	22/6c
Number of pulls <sup>a</sup>	1195	685	473	1105	1505	1459	1229
Non-breaking <sup>b</sup>	437	429	306	398	405	572	250
Unique tethers	31	20	9	11	6	14	26
Pulls with unfold+refold <sup>c</sup>	100	171	64	22	6	121	36
Experiment conducted by	JT	RBE	RBE	JT	JT	JT	JT

Table 3: Overview of optical tweezers experiments. For each pseudoknot, the experiments were conducted either by Jesper Tholstrup (JT) or by Rebecca Bolt Ettliger (RBE) as part of her Master thesis. a: total number of pulling experiments made on ssDNA/RNA tethers, b: number of pulling experiments in which the tether did not break, c: number of pulling experiments made where the force-extension trace contained both an unfolding and a refolding.

The nomenclature of our pseudoknots is such, that the first number describes the number of base pairs in stem1, the second number describes the number of base pairs in stem2 and any subsequent letters (e.g. a, b, ...) indicate sequence variations. For example, pseudoknot 10/6 contains 10 bp in stem1 and 6 bp in stem2 (see Figure 26)

## 9.5 Pseudoknot 10/6

### 9.5.1 Possible structures

In addition to the designed pseudoknot 10/6, our RNA sequences can form a folding intermediate hairpin of this pseudoknot, consisting of a 10 bp stem and a 12 nt loop (10hp). This intermediate is essentially pseudoknot 10/6 without stem2. In addition to the expected pseudoknot, another pseudoknot can form using part of the loop2 sequence to form an alternative stem2 where one of the six base pairs is a G-U wobble. This alternative pseudoknot is called 10/6U. The three structures are illustrated in Figure 26.

### 9.5.2 Mechanical unfolding

Two main types of unfolding were observed for pseudoknot 10/6: unfolding through a single event (One-step unfolding) or unfolding through two separate events (Two-step unfolding). Representative force-extension curves for each of these unfolding pathways are shown in Figure 27. Typically, refolding followed the same pattern, occurring in either a single step or in two separate steps, albeit gradual refolding was observed in 17 force-extension curves.

#### One-step unfolding

In the single molecule experiments conducted here, the sole descriptor of the folded structure is the length of unfolding. We can, using a theoretical framework described in Part IV, estimate how much the length of the tether should increase for unfolding of any structure

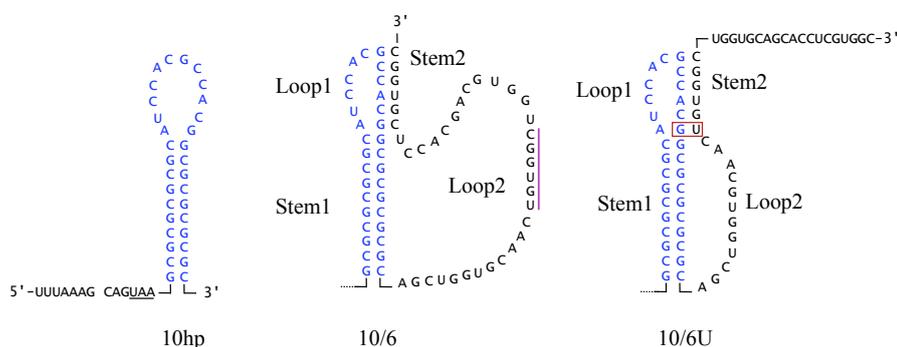


Figure 26: Possible structures for pseudoknot 10/6. The folding intermediate hairpin 10hp and the two pseudoknots 10/6 and 10/6U. 10hp is a folding intermediate of both pseudoknots (blue). Magenta line denote alternative stem2 sequence used in 10/6U. Red box denote G·U wobble base pair. Slippery sequence and in-frame stop codon (underline) is also shown upstream from 10hp. The downstream sequence is omitted in the depiction of 10hp.

at any force. The initial classification of our data was, therefore, based on the unfolding length. All data for tethers which unfolded in a single step was pooled and a cumulative distribution of the unfolding length was constructed. By fitting the model for one or more normal cumulative distribution functions (CDF) to the unfolding length distribution, the number of distinct distributions present in the data was estimated. This analysis of the unfolding data for pseudoknot 10/6 is shown in Figure 28.

From Figure 28 it is clear that a single normal distribution does not describe the data particularly well. A double normal CDF describes the data fairly well but fails to describe the long unfolding events. A triple normal CDF describes the data very well, and although this could be a case of “over fitting”, later observations indicate that we very well could have three underlying distributions, reflecting the three structures shown in Figure 26.

Three overlapping distributions makes the subsequent analysis more tricky as we have to extract information from each of the overlapping distributions. In accordance with Figure 28, three disjoint subsets A, B, and C were defined based on unfolding length ( $l$ ) where the limits of each subset was set to the intersection of the three distributions shown in Figure 28:

$$\begin{aligned}
 A &= \{l \mid 0 \text{ nm} \leq l \leq 15.1 \text{ nm}\} \\
 B &= \{l \mid 15.1 \text{ nm} < l \leq 20.3 \text{ nm}\} \\
 C &= \{l \mid 20.3 \text{ nm} < l\}
 \end{aligned}$$

Creating these subsets is somewhat crude as we loose data in the tail of each distribution, however, given the limited number of data points it is difficult to divide them more gently.

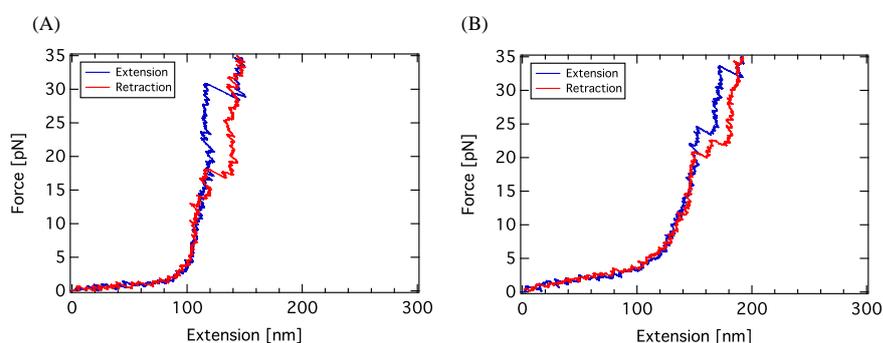


Figure 27: Examples of force-extension curve for pseudoknot 10/6. Of the 100 force-extension curves with unfolding and refolding obtained for pseudoknot 10/6, 74 curves showed unfolding in a single event (A) and 26 curves showed unfolding in two events (B).

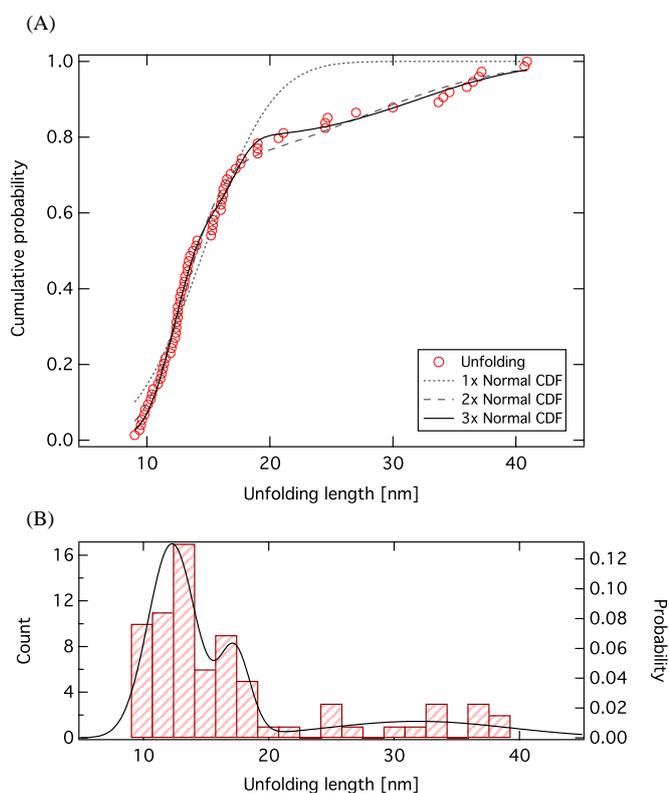


Figure 28: Unfolding lengths observed for unfolding of pseudoknot 10/6. **A:** The experimental data (red circles) was modeled with either a single normal CDF (black dotted), a double normal CDF (black dashed) or a triple normal CDF (black solid). **B:** Histogram of unfolding length with triple normal probability density function from A overlaid using right ordinate.  $n = 74$

Unfolding force versus unfolding length for the three subsets shown in Figure 29 where theoretical unfolding length versus force for the three possible structures shown in Figure 26 is also shown.

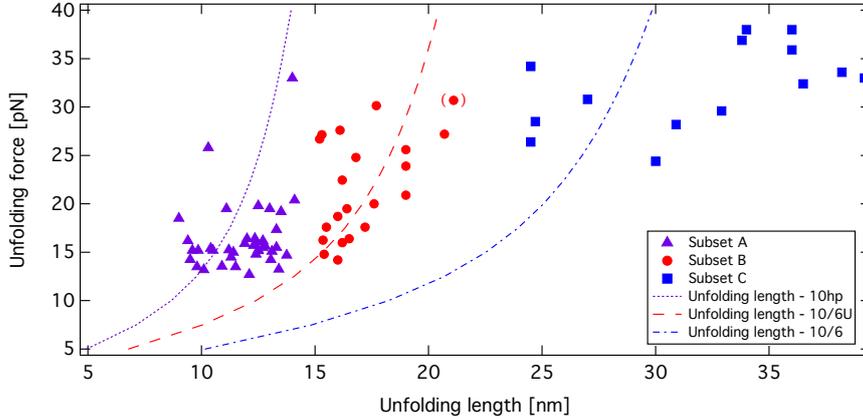


Figure 29: Unfolding force versus unfolding length for unfolding of pseudoknot 10/6. Data for subset A (magenta triangles,  $n = 39$ ), subset B (red circles,  $n = 21$ ) and subset C (blue squares,  $n = 14$ ). Theoretical unfolding lengths for 10hp (magenta, dotted), 10/6U (red, dashed) and 10/6 (blue, dash-dot) are also shown. *Please note:* The data point in subset B marked by parenthesis was manually move from subset C to subset B. It was at the intersection between the two distributions.

Possibly with the exception of the *very* wide distribution of unfolding lengths for subset C, the three distributions appear to describe unfolding of three different structures. The experimental data correlates very well with the expected unfolding lengths of the three indicated structures. Table 4 lists the observed unfolding lengths for subset A, B, and C and the theoretical unfolding length for the three possible structures. The observed unfolding length for subset A is very close to the expected unfolding length of the folding intermediate 10hp. The observed unfolding length for subset B is in good agreement with the expected unfolding length expected from the alternative pseudoknot 10/6U. For subset C the observed unfolding length is closest to the expected unfolding length of pseudoknot 10/6.

	$n$	Experimental values		Theoretical lengths		
		$F$	$l$	10/6	10/6U	10hp
Subset A	39	$16.5 \pm 0.6$ pN	$11.8 \pm 0.2$ nm	23.4 nm	16.0 nm	11.0 nm
Subset B	21	$21.8 \pm 1.1$ pN	$17.1 \pm 0.4$ nm	25.7 nm	17.6 nm	12.0 nm
Subset C	14	$32.1 \pm 1.2$ pN	$32.0 \pm 1.4$ nm	28.5 nm	19.5 nm	13.3 nm

Table 4: Experimental data and theoretical predictions for unfolding of pseudoknot 10/6. Unfolding force ( $F$ ) and length ( $l$ ) observed for the three subsets. Theoretical unfolding lengths for the three possible structures (10/6, 10/6U and 10hp) at the indicated unfolding force. Gray cells indicate best match between observed and predicted value.

The refolding pattern was more complex than the unfolding pattern as a combination of gradual<sup>6</sup>, complete<sup>7</sup>, partial+gradual<sup>8</sup>. The refolding observations are listed in Table 5.

	$n$	$F_r$	$l_r$
Gradual	24 (32%)	N.D.	N.D.
Gradual+partial	13 (18%)	12.3±0.8 pN	9.7±0.8 nm
Complete	37 (50%)	15.4±0.8 pN	11.1±0.5 nm

Table 5: Refolding data for 10/6. Summary of the three different types of observed refolding. Refolding force ( $F_r$ ) and refolding length ( $l_r$ ). Values are mean±SEM. N.D.: Not determined.

The average refolding force for what appeared to be complete refolding was 15.4 pN yielding an average refolding length of 11.1 nm. This refolding length is close to the expected refolding length of 11.0 nm at 15.4 pN for the folding of intermediate hairpin 10hp. It should be mentioned, however, that locating the position where refolding initiates is more difficult than finding the place where unfolding starts. As seen in Figure 27A, the retraction trace typically seems to “bend” towards the extension trace *before* the long and sharp refold step occurs. Such behavior makes it difficult to get an accurate measure of where the actual refolding initiates.

For refolding with occurred through a combination of gradual and partial refolding, the observed partial refolding had an average length of 9.7 nm at 12.3 pN. This is in good agreement with the expected refolding length of 10.1 nm for 10hp. This may indicate that 10hp can refold in a single event while the complete folding into a pseudoknot occurs gradually. Gradual refolding was observed in 30% of the traces and may reflect complete refolding of pseudoknot 10/6 and/or 10/6U.

## Two-step unfolding

For curves with two unfolding events the event occurring at the lowest force will be referred to as the “first” event while the unfolding which occurs at highest force will be referred to as the “second” event.

Analyzing data from force-extension curves where unfolding is a two-step process is complicated because the two unfolding lengths may describe a mix of several structural changes occurring at different forces. For a pseudoknot for example, the two unfolding events could describe unfolding of stem1 followed by unfolding of stem2 - or the other way around - or a mix of the two.

As for one-step unfolding, the length of the unfolding was used as descriptor of the original folded structure and was used for structural classification. There was no statistical difference between the unfolding length (p-value = 0.5878, two-tailed Student’s t-test) or

<sup>6</sup>Refolding where the retraction trace is gradually folded back on to the extension trace

<sup>7</sup>Refolding through a single step bringing the retraction trace back on top of the extension trace

<sup>8</sup>Refolding through a single step and gradual refolding

the unfolding force (p-value = 0.1104, two-tailed Student's t-test) between the two steps. If we consider only the two pseudoknots 10/6 and 10/6U (it seems less likely that a hairpin would unfold in two distinct events), we have four possible first steps and four possible second steps. Figure 30 show scatter plots of the first event (30A) and the second event (30B) along with theoretical unfolding lengths for the two possible unfolding pathways of pseudoknot 10/6 and 10/6U.

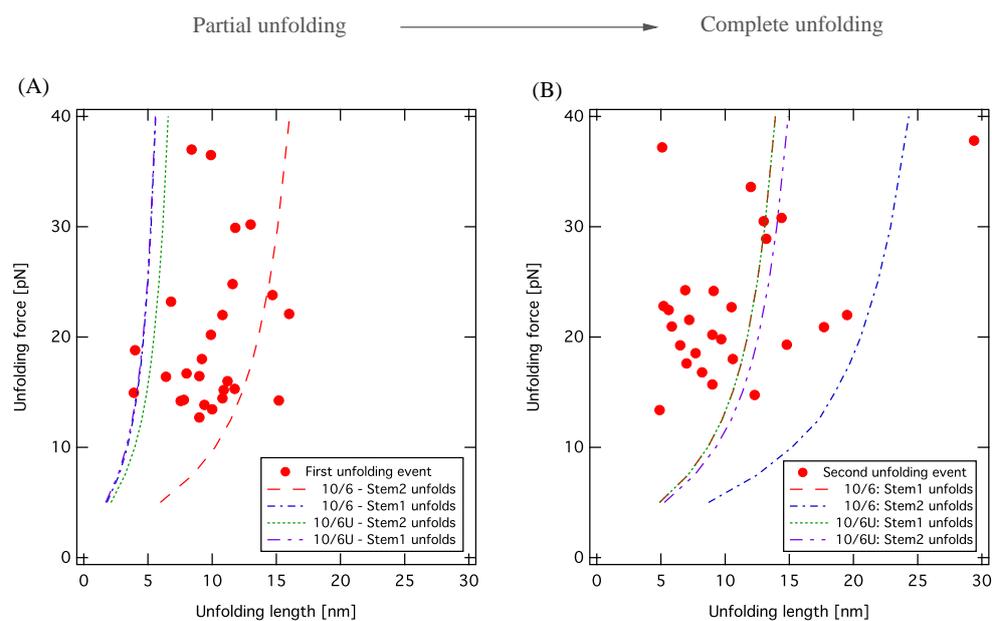


Figure 30: Two step unfolding of pseudoknot 10/6. **A**: Data for the first unfolding event. Theoretical unfolding length for partial unfolding of 10/6 (stem2 unfolds [red, long dash], stem1 unfolds [blue, dash-dot]) and partial unfolding of pseudoknot 10/6U (stem2 unfolds [green, dot] and stem1 unfolds [magenta, dash-dot-dot]). **B**: Data for the second unfolding event. Theoretical unfolding lengths for complete unfolding of partially unfolded pseudoknot 10/6 (stem1 unfolds [red, long dash], stem2 unfolds [blue, dash-dot]) and for complete unfolding of partially unfolded pseudoknot 10/6U (stem1 unfolds [green, dot], stem2 unfolds [magenta, dash-dot-dot]).

The best match for both the first and second unfolding event is the unfolding of pseudoknot 10/6 through the following pathway: First stem2 followed by stem1 (Figure 30).

Although there is a difference between the observed unfolding length and the expected unfolding length, most of the observed unfolding lengths were within 3-4 nm of the expected length. The observed length was systematically smaller than the expected length which may indicate that for these short unfolding events we are unable to accurately locate the unfolding start and end, possibly due to noise.

As for one-step refolding, assigning a specific structure to each refolding is tricky and the dataset is even more limited in this case. The 26 refolding events were a mix of gradual- (19%), complete- (23%), partial + gradual (35%) and two-step refolding (23%),

which makes the number of experimental observations in each group too small to allow structural assignments. A summary of the observed unfolding and refolding for two-step unfolding is listed in Table 6.

	Unfolding (experimental)		Refolding (experimental)			
			One-step		Two-step	
	First	Second	Partial+Gradual	Complete	First	Second
Force [pN]	$19.8 \pm 1.4$	$22.8 \pm 1.3$	$12.9 \pm 0.4$	$15.3 \pm 0.9$	$19.6 \pm 2.0$	$18.1 \pm 2.0$
Length [nm]	$9.9 \pm 0.6$	$10.6 \pm 1.1$	$8.9 \pm 0.6$	$12.8 \pm 0.7$	$11.4 \pm 1.9$	$8.6 \pm 1.1$
	$n = 26$		$n = 9$	$n = 6$	$n = 5$	

Table 6: Statistics for two-step unfolding of pseudoknot 10/6. Unfolding/refolding force and unfolding/refolding length for each event. See main text for description of refolding types. In addition to the listed refolding events a total of 6 refolded gradually. Values are mean  $\pm$  SEM.

It is worth mentioning that the two step unfolding traces appeared at random time points during the experiments, e.g. a single tether could unfold in a single step, refold, unfold in two steps, refold, unfold in a single step and so forth. This excludes the possibility that the two step unfolding events are caused by a single erroneous tether.

### 9.5.3 Kinetics of 10hp, 10/6U, and 10/6

Only traces where unfolding occurred in a single step could be used in the kinetic analysis as outlined in section 7.4. To investigate the structural kinetics from force-ramp experiments like those conducted here, we need to construct a probability distribution from the unfolding force distribution for each structure. The data indicates that we have three different structures, making it impossible to create the exact probability distribution. However, by using the three disjoint subsets A, B, and C, defined on page 74, we are able to get a reasonable estimate.

Hence, for the kinetic analysis of pseudoknot 10/6 we used the three disjoint subsets A, B, and C and the unfolding probability distributions were constructed as if these subsets contained all observed unfolding events for their particular structures. The result of the kinetic analysis is shown in Figure 31 and the results are listed in Table 7.

	$n$ (fit)	$X^\ddagger$	$k_0$
A	30	$3.3 \pm 0.28$ nm	$1.9 \cdot 10^{-5} \pm 1.9 \cdot 10^{-5}$ s <sup>-1</sup>
B	20	$0.7 \pm 0.09$ nm	$3.0 \cdot 10^{-2} \pm 1.4 \cdot 10^{-2}$ s <sup>-1</sup>
C	13	$0.9 \pm 0.07$ nm	$1.3 \cdot 10^{-3} \pm 7.3 \cdot 10^{-4}$ s <sup>-1</sup>

Table 7: Results of kinetic analysis of pseudoknot 10/6. Values for subsets are fitting coefficients  $\pm$  95% confidence interval (from fits in Figure 31).  $X^\ddagger$ : Distance to transition state,  $k_0$ : zero force rate constant

The kinetic parameters support the hypothesis from the analysis of unfolding lengths

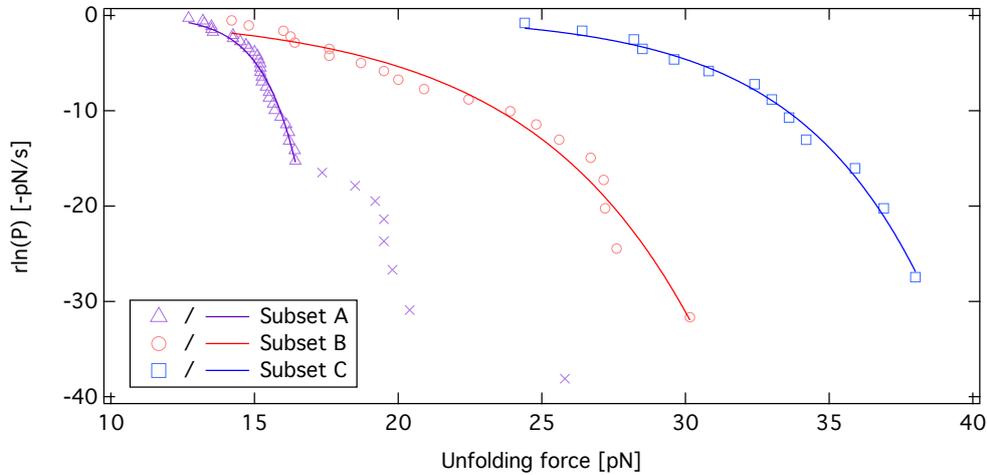


Figure 31: Kinetics of pseudoknot 10/6. Analysis for subset A (magenta, triangles), subset B (red, circles) and subset C (blue, squares) with equation (21) fitted (solid lines). *Please note:* Outliers for subset A was omitted from the fit (magenta, cross,  $n=8$ ).

above, namely that subset A describes a hairpin while subsets B and C each describe pseudoknots. The distance to transition state determined for subset A is much larger than the same parameter for subset B and C. This supports the notion that the data in subset A originates from the unfolding of a hairpin. Hairpins typically have a  $X^\ddagger$  of several nanometers, while pseudoknots tend to have a much smaller value of  $X^\ddagger$  (in a buffer containing  $\text{Mg}^{2+}$ ) [Green et al., 2008, Li et al., 2006b, Liphardt et al., 2001]. The value of  $k_0$  for subset A is small, albeit with a wide confidence interval, further supporting the hairpin hypothesis. The kinetic parameters for subset B and C both indicate that the data originates from unfolding of brittle structures, like pseudoknots. The distance to transition state is significantly smaller than for subset A and the value of  $k_0$  is larger than for subset A. Both of these observations have previously been observed for pseudoknots [Green et al., 2008]. The kinetic analysis support the notion that subset A describes unfolding of a folding intermediate hairpin of pseudoknot 10/6(U) and that subset B and C describes unfolding of pseudoknots.

#### 9.5.4 Thermodynamics of 10hp, 10/6U and 10/6

As described in Section 7.5 it is possible to extract equilibrium information from the non-equilibrium experiments conducted here. This analysis was performed by calculating the work used during unfolding and the work returned during refolding. The work was calculated as described elsewhere [Collin et al., 2005] using the CFT (details about how the work was calculated are provided in Appendix D).

As for the kinetic analysis, we are forced to look at the three different structures based on the three disjoint subsets A, B, and C separately. Unfolding and refolding work was extracted from as many traces as possible from each subset and the correction term ( $W_{corr}$ ,

equation (28)) was calculated under the assumption that subset A, B, and C each describe 10hp, 10/6U and 10/6 respectively. It was not possible to extract work information from every trace as some unfolding events occurred during the retraction phase. Also, drift was a problem in some traces - although the drift was not too severe, it was enough to prevent proper overlap between the extension and retraction after refolding.

The mix of structures complicates calculation of the correction term. This term is used to subtract the work used for stretching the RNA/DNA handles, the ssRNA outside the structure and the ssRNA of the unfolded structure. Thus, the correction assumes a particular structure as the nature of the structure determines how many RNA nucleotides exist outside and within the folded structure and hence how the RNA nucleotides are stretched when the structure unfolds. This is illustrated in Table 8 where the correction term (i.e. energy) used for stretching the indicated structures from 10 pN to 15 pN are listed.

	10hp	10/6U	10/6
RNA/DNA handles	9.7 $k_B T$	9.7 $k_B T$	9.7 $k_B T$
ssRNA <sub>out</sub>	8.5 $k_B T$	6.1 $k_B T$	3.6 $k_B T$
ssRNA <sub>struct</sub>	11.5 $k_B T$	22.2 $k_B T$	30.6 $k_B T$
Total work	29.7 $k_B T$	38.0 $k_B T$	43.9 $k_B T$

Table 8: Correction term - examples. “RNA/DNA handles”: work required for stretching the RNA/DNA handles from 10 pN to 15 pN. “ssRNA<sub>out</sub>”: work required for stretching ssRNA outside the pseudoknot structure from 10 pN to 15 pN. “ssRNA<sub>struct</sub>”: the work required to stretch the ssRNA of the unfolded structure from 0 pN to 15 pN.

Despite of this structural assumption in the correction term, the unfolding length in the disjoint subsets A, B, and C allows us to predict from which structure the unfolding/refolding work is most likely to originate. The correction term was therefore calculated for subsets A, B, and C independently under the assumption that they each describe 10hp, 10/6U and 10/6 respectively.

The unfolding and refolding work distributions for the three subsets are shown in Figure 32. For each subset the unfolding and refolding work were separated into equally spaced bins. As there is no formal requirement for the nature of these distributions, i.e. they do not have to be Gaussian, the intersect between the bins were used as done in refs [Collin et al., 2005, Green et al., 2008].

The Gibbs free energy was calculated independently for each subset based on Figure 32 and the average correction term for each subset. The result from this analysis is listed in Table 9.

The estimates of the Gibbs free energy obtained from the CFT are in excellent agreement with the expected free energies for the three structures (Table 9). The small differences between subset A and subset B may indicate that we are unable to separate the two distributions to a sufficient degree.

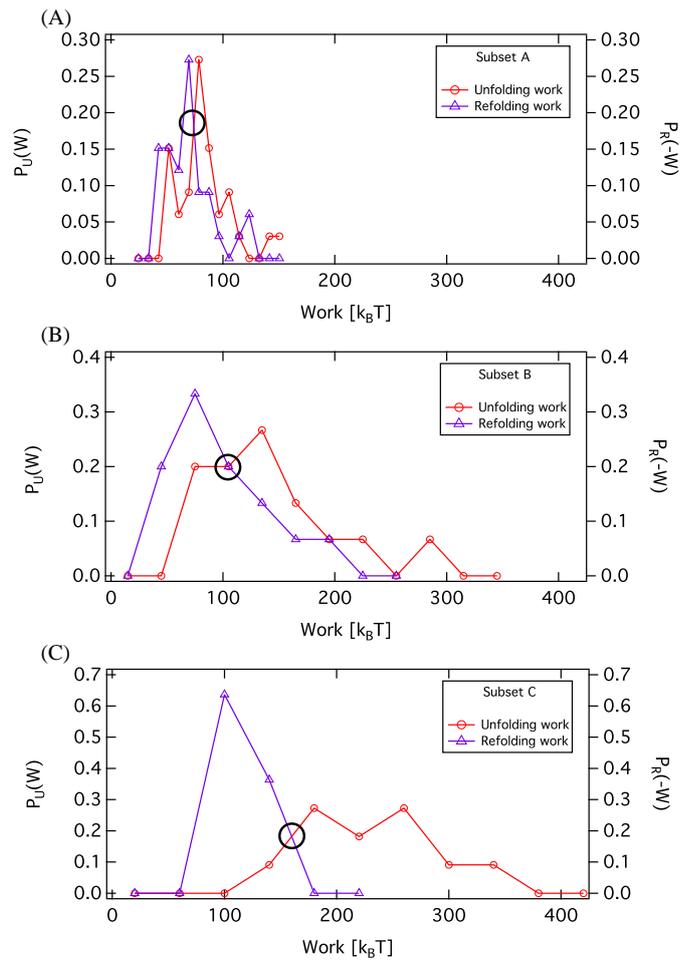


Figure 32: Work distributions for 10hp, 10/6U, and 10/6. **A:** Data for hairpin 10hp (subset A),  $n = 33$ . Histogram of unfolding work (red, circles) and refolding work (magenta, triangles). **B:** Pseudoknot 10/6U (Subset B),  $n = 15$ . **C:** Pseudoknot 10/6 (Subset C),  $n = 11$ . Black circle indicate where  $P_U(W) = P_R(-W)$ .

Expected structure	Subset A	Subset B	Subset C
	10hp	10/6U	10/6
$\langle P_U(W) \rangle$	$81.3 \pm 3.7 \text{ k}_B\text{T}$	$123.4 \pm 11.1 \text{ k}_B\text{T}$	$220.0 \pm 20.8 \text{ k}_B\text{T}$
$\langle P_R(-W) \rangle$	$65.8 \pm 3.4 \text{ k}_B\text{T}$	$88.5 \pm 11.2 \text{ k}_B\text{T}$	$114.3 \pm 4.3 \text{ k}_B\text{T}$
$\langle W_{corr} \rangle$	$21.6 \pm 1.4 \text{ k}_B\text{T}$	$47.0 \pm 4.2 \text{ k}_B\text{T}$	$86.0 \pm 7.5 \text{ k}_B\text{T}$
$P_U(W) = P_R(W)$	$74.0 \text{ k}_B\text{T}$	$105.0 \text{ k}_B\text{T}$	$160.0 \text{ k}_B\text{T}$
$\Delta G_{PK} / \Delta G_{HP}$	$-31.1 \pm 0.8 \text{ kcal/mol}$	$-34.7 \pm 2.6 \text{ kcal/mol}$	$-43.9 \pm 4.5 \text{ kcal/mol}$
$\Delta G_{theory}$	$-23.2 \text{ kcal/mol}$	$-35.2 \text{ kcal/mol}$	$-41.6 \text{ kcal/mol}$

Table 9: Gibbs free energy of 10hp, 10/6U and 10/6. Average work required for unfolding ( $\langle P_U(W) \rangle$ ), average work returned during refolding ( $\langle P_R(-W) \rangle$ ), intersect between the two work distributions ( $P_U(W) = P_R(-W)$  [ $\Delta G_T$ ]) and the average correction for stretching of handles and RNA ( $\langle W_{corr} \rangle$ ). Estimated Gibbs free energy difference between folded and unfolded structure using Crooks Fluctuation Theorem ( $\Delta G_{HP} / \Delta G_{PK}$ ). Theoretical Gibbs free energy ( $\Delta G_{theory}$ ) was estimated using mfold [Zuker, M., 2003] (10hp) and pknotsRG [Reeder et al., 2007] (10/6U and 10/6). Values are mean  $\pm$  SEM. Number of data points: 33, 15 and 11 for subset A, B, and C respectively.

### 9.5.5 Unfolding of 10hp, 10/6U, and 10/6 - Summary

A summary of the observed unfolding and refolding events are shown in Figure 33 where unfolding is shown in blue (one-step) and green (two-step) and possible refolding is shown in red.

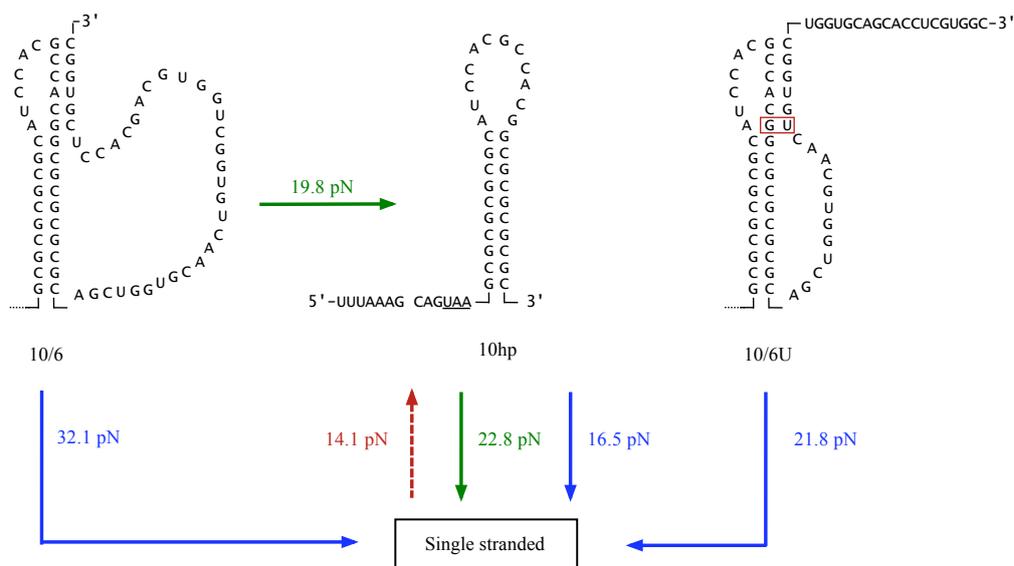


Figure 33: Observed unfolding and refolding for 10/6 structures. One-step unfolding (blue), two-step unfolding (green) and refolding (red) was observed at the indicated forces. The folding intermediate 10hp could possibly refold from other species than single stranded RNA.

For the one-step unfolding events there was excellent agreement between the experimental and the theoretical unfolding length for 10hp and 10/6U - and a reasonable agreement between the experimental and theoretical unfolding length for pseudoknot 10/6 (see Figure 29). There was also excellent agreement between the experimental refolding length and the theoretical refolding length of 10hp. However, it was not always clear when refolding initiated.

The average unfolding length of the two-step unfolding events deviated slightly from the expected unfolding lengths. However, from Figure 30A and 30B it does not seem unlikely that the two-step unfolding events originate from unfolding of pseudoknot 10/6 through 10hp.

## 9.6 Pseudoknot 11/6 and 6/11

Mechanical unfolding of pseudoknot 11/6 and pseudoknot 6/11 was performed by Rebecca Bolt Ettliger as part of her Master thesis work [Ettliger, 2012] and the structures of these two pseudoknots are shown in Figure 34 along with folding intermediates and an alternative pseudoknot (11/6U). Pseudoknot 11/6 was able to induce -1 frameshift *in vivo* [Tholstrup

et al., 2012], while pseudoknot 6/11 was unable to induce detectable levels of -1 frameshift (not shown).

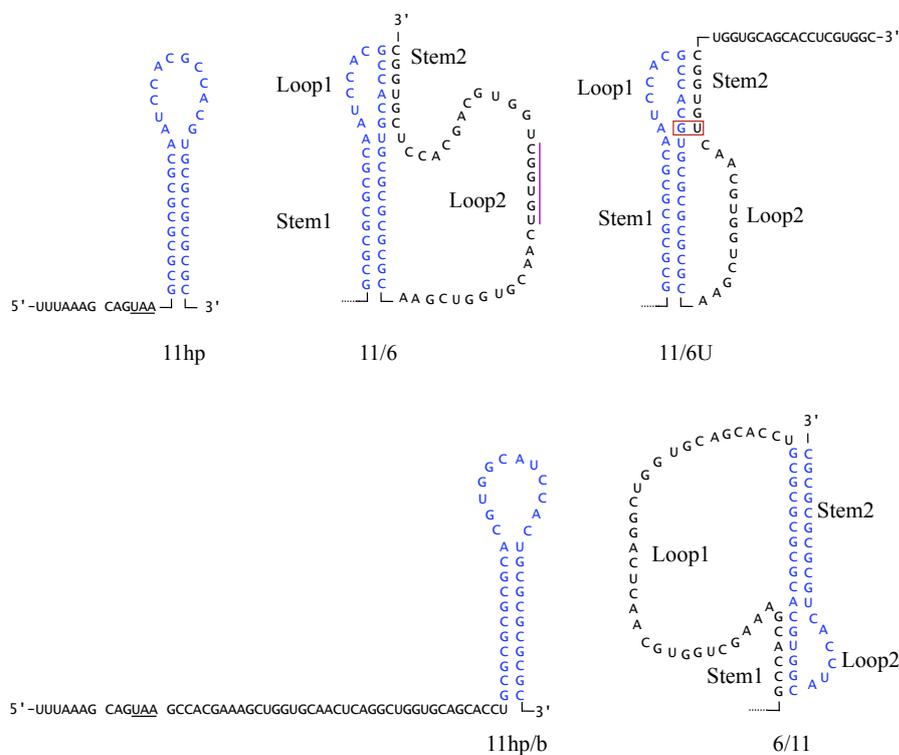


Figure 34: Possible structures for pseudoknot 11/6 (top) and 6/11 (bottom). **Top:** The folding intermediate “11hp” and the two pseudoknots 11/6 and 11/6U. 11hp is a folding intermediate of both pseudoknots (blue). Magenta line denotes alternative stem2 used in 11/6U. Red box denotes G·U wobble base pair. **Bottom:** Folding intermediate “11hp/b” is a folding intermediate of pseudoknot 6/11 consisting of stem2 (blue). Slippery sequence and in-frame stop codon (underline) is also shown upstream from 11hp and 11hp/b. *Please note:* The downstream sequence is omitted in the depiction of 11hp and the upstream sequence is omitted for 11/6, 11/6U and 6/11.

Rebecca Bolt Ettlinger found that the unfolding pattern of pseudoknot 11/6 was very similar to that of pseudoknot 10/6 investigated above, namely that both the expected pseudoknot 11/6, pseudoknot 11/6U, and the folding intermediate hairpin (11hp) was able to form. Unfortunately, it was not possible to separate the unfolding data for pseudoknot 11/6U from unfolding data for the 11hp hairpin.

For pseudoknot 6/11, a “reverse” pseudoknot where the lengths of stem1 and stem2 are swapped compared to the 11/6 pseudoknot, the unfolding data showed that the sequence most likely does *not* form the desired pseudoknot and that the structural population consisted almost exclusively of unknown structures.

Rebecca Bolt Ettlinger also estimated the kinetic parameters and found estimates of

$X^\ddagger$  and  $k_0$  similar to those presented for pseudoknot 10/6 above. A summary of the unfolding data and kinetic parameters are listed in Table 10.

	$n$	Experimental values		Estimated values	
		$F$	$l$	$X^\ddagger$	$k_0$
Pseudoknot 11/6U+11hp	145	19.4±0.5 pN	14.8±0.3 nm	0.75±0.02 nm	3.0·10 <sup>-2</sup> ±2.0·10 <sup>-2</sup> s <sup>-1</sup>
Pseudoknot 11/6	8	28.9±4.4 pN	28.2±0.3 nm	0.32±0.06 nm	7.0·10 <sup>-2</sup> ±2.0·10 <sup>-3</sup> s <sup>-1</sup>
6/11: Unknown structure(s)	51	15.4±0.5 pN	13.8±0.5 nm	1.56±0.08 nm	1.0·10 <sup>-3</sup> ±4.0·10 <sup>-4</sup> s <sup>-1</sup>

Table 10: Data from [Ettlinger, 2012]. **Top two rows:** most probable structures recovered for the sequence designed to form pseudoknot 11/6: an 11 bp folding intermediate hairpin+pseudoknot 11/6U and the full pseudoknot 11/6. **Bottom row:** It was not possible to find structures corresponding to the observed unfolding lengths for the sequence designed to fold into pseudoknot 6/11.  $F$ : unfolding force,  $l$ : unfolding length,  $X^\ddagger$ : Distance to transition state,  $k_0$ : zero-force rate constant. The structure of pseudoknot 11/6 and 6/11 are shown in Figure 34. Values are mean±SEM for  $F$  and  $l$ , and mean±asymptotic standard error for  $X^\ddagger$  and  $k_0$ .

## 9.7 Pseudoknot 16/6

We found that pseudoknot 16/6 was able to stall translating ribosomes but despite rigorous efforts with several different RNA preparations and different ssDNA/RNA hybridizations it was not possible to obtain a sizable dataset for pseudoknot 16/6. Possible explanations for the lack of observed unfolding events will be discussed later. Figure 35 shows the observed one-step unfolding events and Figure 36 shows the observed two-step unfolding events. As for pseudoknot 10/6, pseudoknot 16/6 is able to form both a hairpin (16hp) and an alternative structure (16/6U) in addition to pseudoknot 16/6 - these are identical to those showed for 10/6 in Figure 26 (page 74) except that stem1 consists of 16 bp with the sequence shown in [Tholstrup et al., 2012].

From Figure 35 it seems that most of the observed unfolding events could be explained by unfolding of the three expected structures, 16hp, 16/6U and 16/6. Due to the limited dataset it was not possible to assign a specific structural change to each event. From Figure 36 it seems that the recorded two-step unfolding events can be explained by unfolding of pseudoknot 16/6 and 16/6U, both in the pathway: stem2 followed stem1.

As a consequence of the very limited dataset for unfolding of pseudoknot 16/6 it is not possible to make any quantitative conclusions. However, the observed unfolding events indicate that the expected structures are able to form.

## 9.8 Pseudoknot 22/6a

Pseudoknot 22/6a is the most stable of the pseudoknots designed in here, and we found that it was able to stall translating ribosomes *in vivo* [Tholstrup et al., 2012]. However, as for pseudoknot 16/6, we were unable to obtain tangible unfolding data for pseudoknot

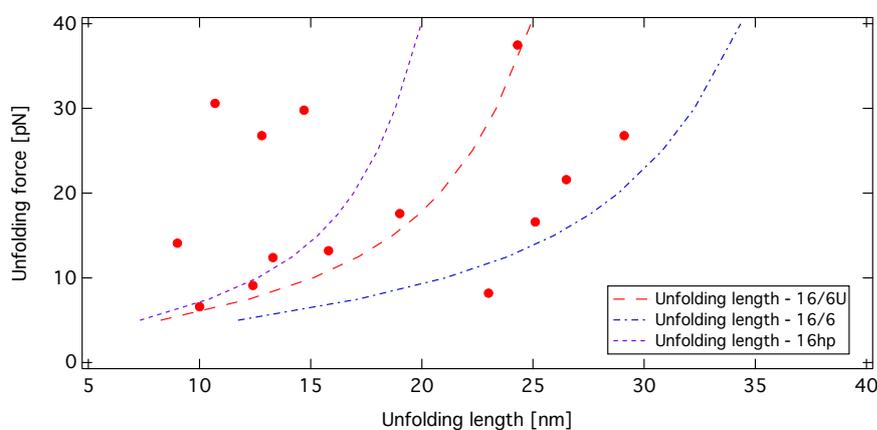


Figure 35: Unfolding force versus unfolding length of pseudoknot 16/6. Theoretical unfolding length for 16hp (magenta, dotted), 16/6U (red, dashed), and for 16/6 (blue, dash-dot) have been added.

22/6a despite rigorous attempts. Of the 405 pulls on tethers that did not break only 6 unfolding/refolding traces were observed. Each of these 6 unfolding/refolding traces was observed for different tethers, i.e. we were not able to get repeated unfolding/refolding data from the same tether.

The unfolding force for one-step unfolding was  $16.7 \pm 3.4$  pN and the unfolding length was  $19.0 \pm 4.0$  nm (Mean  $\pm$  SEM,  $n = 5$ ). This unfolding length is consistent with unfolding of a 22bp hairpin which could exist as a folding intermediate of pseudoknot 22/6a. A single two-step unfolding was observed with unfolding forces of 13.2 pN and 11.8 pN and unfolding lengths of 20.0 nm and 8.2 nm. These unfolding lengths are not consistent with unfolding of known structures.

As for pseudoknot 16/6, we have no bases for making any statements about the mechanical properties of pseudoknot 22/6a.

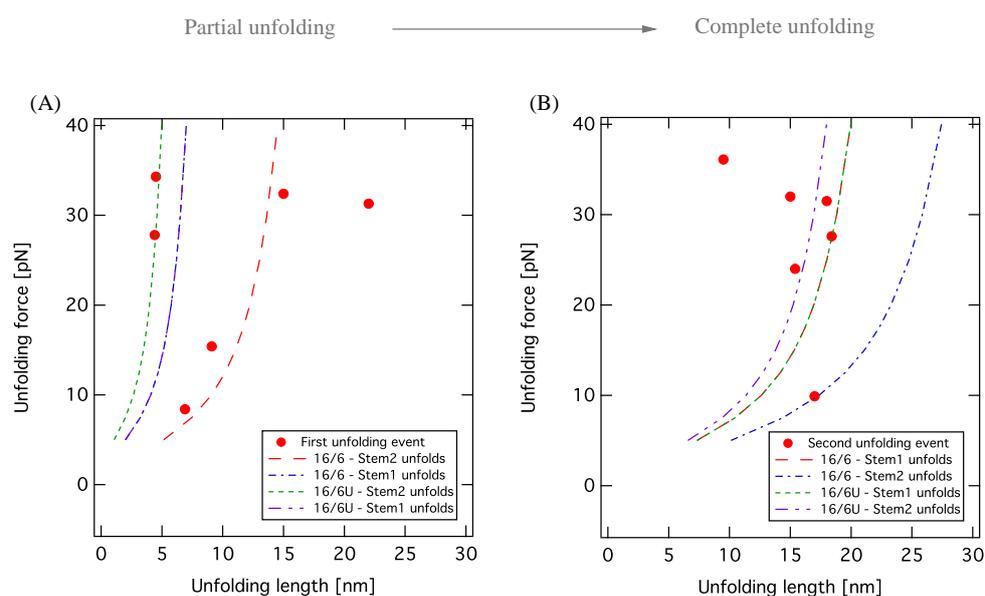


Figure 36: Two step unfolding of pseudoknot 16/6. **A:** Data for the first unfolding event. Theoretical unfolding length for partial unfolding of 16/6 (stem2 unfolds [red, long dash], stem1 unfolds [blue, dash-dot]) and partial unfolding of pseudoknot 16/6U (stem2 unfolds [green, dot] and stem1 unfolds [magenta, dash-dot-dot]). **B:** Data for the second unfolding event. Theoretical unfolding lengths for complete unfolding of partially unfolded pseudoknot 16/6 (stem1 unfolds [red, long dash], stem2 unfolds [blue, dash-dot]) and for complete unfolding of partially unfolded pseudoknot 16/6U (stem1 unfolds [green, dot], stem2 unfolds [magenta, dash-dot-dot]). *Please note:* The theoretical unfolding length for unfolding of stem1 is identical between 16/6 and 16/6U in both A and B.

## 9.9 Pseudoknot 22/6b

As for pseudoknot 10/6, the sequence designed to form pseudoknot 22/6b is able to form an alternative pseudoknot (22/6bU) in addition to the 22bp folding intermediate hairpin (22hp/b). These three structures are shown in Figure 37. This pseudoknot is structurally identical to pseudoknot 22/6a, but stem1 contains destabilizing G-C to A-U substitutions.

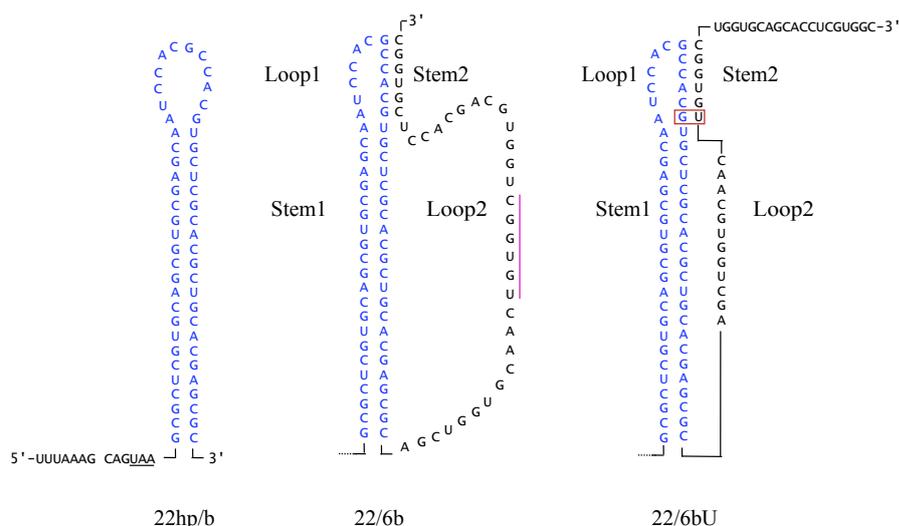


Figure 37: Possible structures for pseudoknot 22/6b. The folding intermediate “22hp/b” and the two pseudoknots 22/6b and 22/6bU. 22hp/b is a folding intermediate of both pseudoknots (blue). Magenta line denote alternative stem2 used in 22/6bU. Red box denote G·U wobble base pair in stem2 of 22/6bU. *Please note:* The downstream sequence is omitted in the depiction of 22hp/b.

### 9.9.1 Mechanical Unfolding

Several different types of unfolding and refolding was observed for pseudoknot 22/6b, of which the majority is illustrated in Figure 38. Unfolding could occur in a single step, in two steps or in a combination of multiple steps and gradual unfolding. Refolding could occur in a single step, in two steps, gradually or with partial refolding combined with gradual full refolding. In addition to the examples shown here, a total of 13 force-extension curves showed gradual unfolding in one or more “steps”.

#### One Step Unfolding

The majority of unfolding events occurred in a single step (75%) as shown in Figure 38A and B. As for the analysis of pseudoknot 10/6, the unfolding length was used as descriptor for the folded structure and a cumulative distribution of unfolding length is shown in Figure 39.

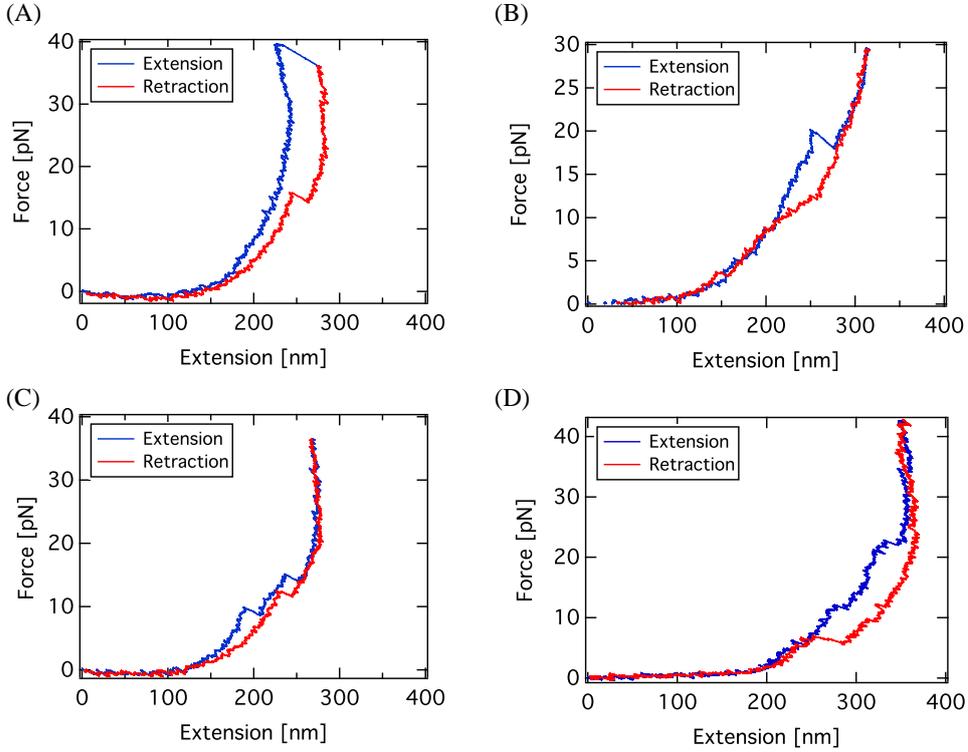


Figure 38: Examples of force-extension curves for pseudoknot 22/6b. **A**: Unfolding in one event with either full refolding during retraction or partial combined gradual refolding (shown). **B**: Unfolding in one event and gradual refolding during retraction. **C**: Unfolding in two distinct events with either full, gradual or partial combined with gradual refolding during retraction. **D**: Unfolding through several small steps with possible gradual unfolding and either full or partial combined with gradual refolding.

From Figure 39 it is clear that we have at least two separate distributions for unfolding length of pseudoknot 22/6b. Consequently, the dataset was separated into two disjoint subsets separated at the intersection between the unfolding length ( $l$ ) for two distributions:

$$\begin{aligned} \text{DE} &= \{l \mid 0 \text{ nm} \leq l \leq 34.4 \text{ nm}\} \\ \text{F} &= \{l \mid 34.4 \text{ nm} < l\} \end{aligned}$$

Although the unfolding length distribution was best described by a bimodal normal distribution, it is still possible that the data in subset DE describes two different structures. As we saw for pseudoknot 10/6, the possibility for an alternative pseudoknot also exists for pseudoknot 22/6b (called hairpin 22hp/b). However, unlike the situation for pseudoknot 10/6U and hairpin 10hp, it is much more difficult to separate the unfolding of hairpin 22hp/b from the unfolding of pseudoknot 22/6bU due to their similar unfolding length.

The scatter plot of unfolding force versus unfolding length for the subset DE and

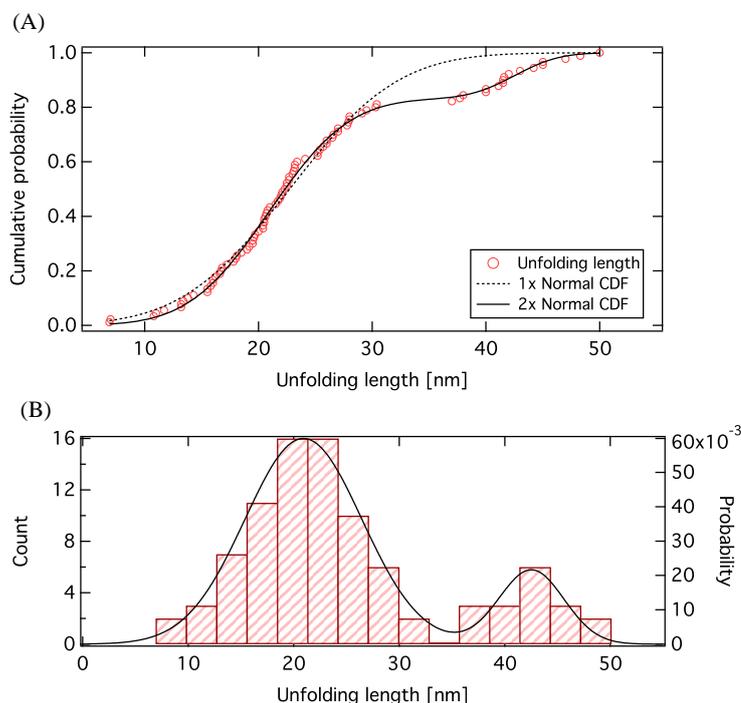


Figure 39: **A**: Cumulative probability for unfolding length of pseudoknot 22/6b (red circles). Data was fitted with either a single normal CDF (black, dotted) or a double normal CDF (black, solid). **B**: Normal PDF from fit in A with histogram of unfolding length for pseudoknot 22/6b overlaid using right ordinate.  $n = 90$

F shown in Figure 40A illustrates this problem. The difference between the expected unfolding length of 22hp/b and 22/6bU is quite small. Even so, one can argue that two clusters are visible within subset DE. One is a few nanometers shorter than the theoretical unfolding length of 22hp/b and the other is a few nanometers longer than the theoretical unfolding length for 22/6bU. Subset DE was therefore separated into two disjoint subsets, subset D and subset E, by minimizing the horizontal distance to the theoretical unfolding length for 22hp/b and 22/6bU. A scatter plot of unfolding force versus unfolding length for subset D, E, and F is shown in Figure 40B.

A summary of unfolding data is listed in Table 11 where the observed unfolding lengths have been used to assign a structure to each subset.

The refolding pattern for the three subsets was comprised of four types of refolding: complete, gradual, gradual+partial and two-step. There was no difference in the refolding pattern between the three subsets, indicating once again that refolding is independent of unfolding. The refolding data for subset D, E, and F is listed in Table 12.

Assigning a structural change to the gradual refolding is difficult as the process occurs continuously (gradually) over a wide range of forces. However, gradual refolding was always complete in the sense that the retraction trace either ended on top of the extension

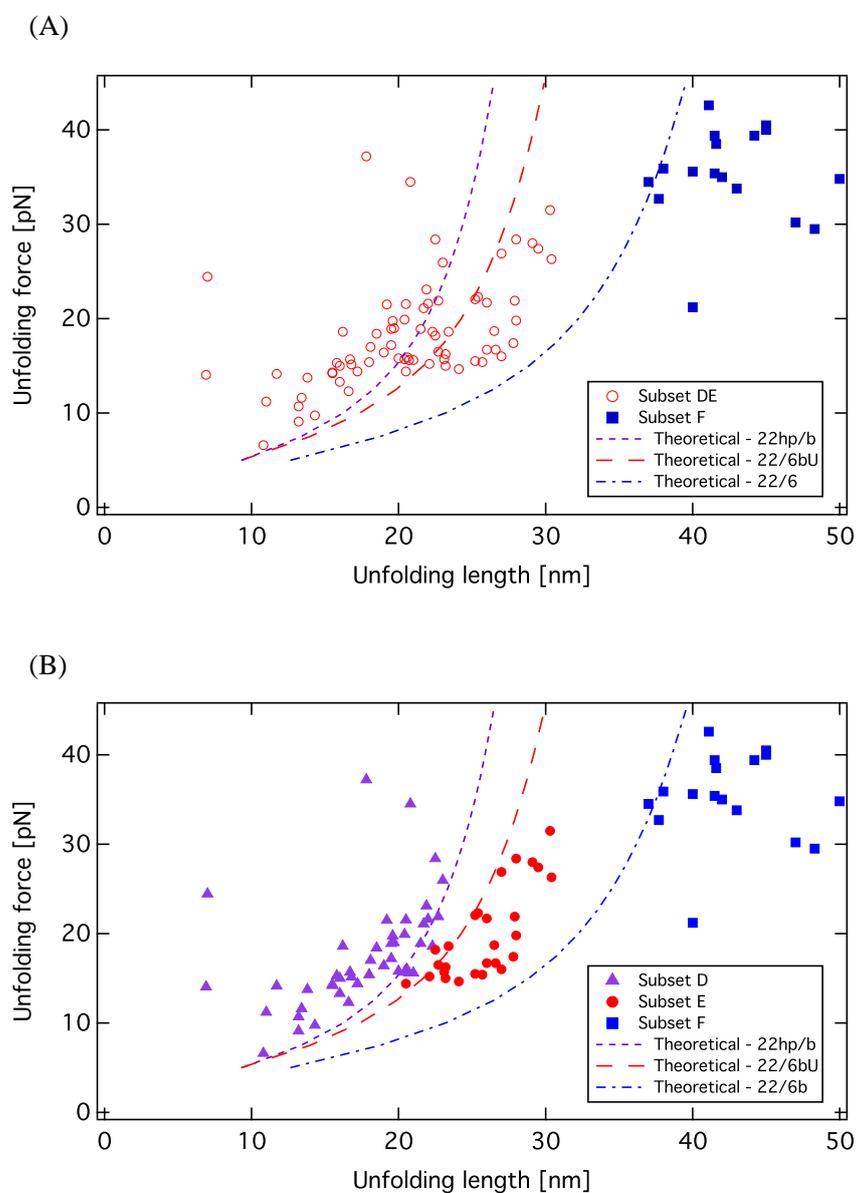


Figure 40: Unfolding force versus unfolding length of pseudoknot 22/6b. **A:** Data for subset DE (red, open circles,  $n = 73$ ) and subset F (blue squares,  $n = 17$ ). **B:** Data for subset D (magenta, triangles,  $n = 46$ ), subset E (red, circles,  $n = 27$ ) and subset F (blue, squares,  $n = 17$ ). Theoretical unfolding lengths for 22hp/b, 22/6bU and 22/6b are also shown.

trace after refolding or was slightly “shorter”, i.e. having a slightly shorter extension after refolding was complete. This indicates that gradual refolding can result in the formation of different structures. This is in good agreement with the unfolding data which indicate

	$n$	Observed values		Theoretical lengths		
		$F$	$l$	22/6b	22/6bU	22hp/b
Subset D	46	$17.5 \pm 0.9$ pN	$17.6 \pm 0.6$ nm	30.7 nm	23.1 nm	21.0 nm
Subset E	27	$19.9 \pm 1.0$ pN	$25.8 \pm 0.5$ nm	32.2 nm	24.2 nm	21.9 nm
Subset F	17	$35.2 \pm 1.2$ pN	$42.5 \pm 0.9$ nm	37.6 nm	28.4 nm	25.2 nm

Table 11: Observed and theoretical unfolding data. Unfolding force and length observed for the three subsets D, E, and F. Theoretical unfolding lengths for three possible structures (22/6b, 22/6bU and 22hp/b) at the indicated unfolding force. Gray cells indicate best match between observed value and predicted value. Values are mean $\pm$ SEM.

	$n$	$F_r$	$l_r$
Gradual	45 (50%)	N.D.	N.D.
Gradual+partial	37 (41%)	$9.3\pm 0.5$ pN	$11.1\pm 0.5$ nm
Complete	5 (6%)	$11.6\pm 1.4$ pN	$15.3\pm 1.7$ nm
Two-step	3 (3%)	$15.0\pm 3.6$ pN	$19.5\pm 1.6$ nm
		$14.7\pm 3.7$ pN	$10.1\pm 2.7$ nm

Table 12: Refolding data for 22/6b. Summary of the four different types of observed refolding. Values are mean $\pm$ SEM. N.D.: Not determined.

that the folded structure alternates between 22/6bU and 22/6b (and possibly 22hp/b) on the same tether.

The refolding length of the distinct refolding event in tethers with both gradual and partial refolding was 11.3 nm. This refolding length is shorter than any known structure and may indicate that refolding occurs either through partial formation of stem1 or through an unknown structure about 40% of the time. Complete refolding was observed in 5% of our retractions with a refolding length of 15.3 nm. The limited number of data points available makes it difficult to make solid statements about these events, however, this refolding length is close to the expected refolding length of 18.9 nm of 22hp/b.

Interestingly, both the unfolding and refolding length of 22hp/b, was shorter than expected. A possible explanation is that the stem of 22hp/b is only partially folded. This possibility is discussed in detail later.

Refolding in two separate steps was observed for three tethers, and as for complete refolding, the limited number of data points and the relative large uncertainty limits our ability to make structural assignments. Two step refolding of 22/6b or 22/6bU cannot explain the observed refolding lengths and these refolding event does most likely not result in the formation of any predicted structures.

### Two-Step Unfolding

A fraction of the unfolding events for pseudoknot 22/6b occurred in two distinct events (15%). Refolding for all of these was either complete, gradual or gradual+partial. A scatter plot of unfolding force versus unfolding force for the first (A) and the second (B) unfolding event is shown in Figure 41.

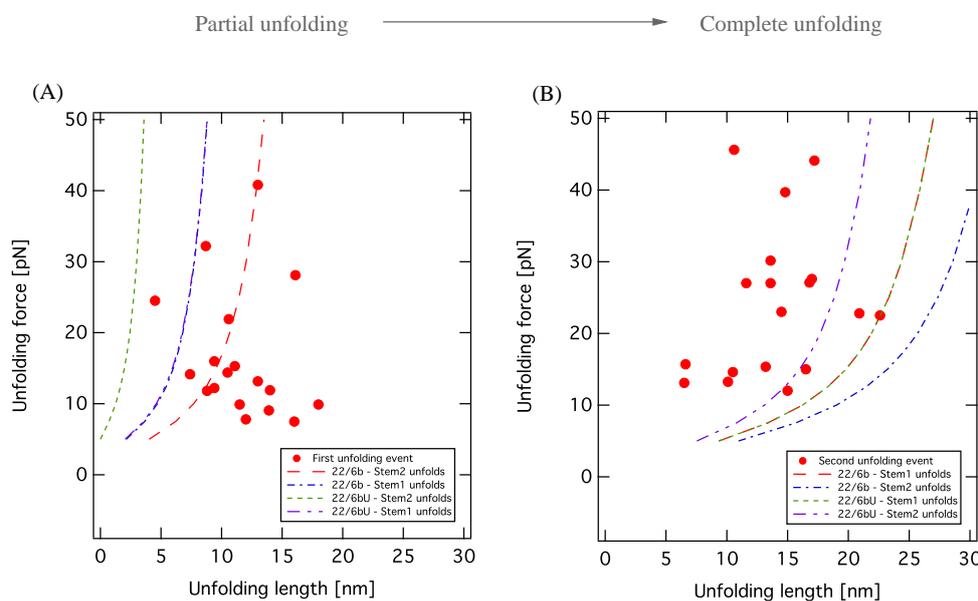


Figure 41: Two step unfolding of pseudoknot 22/6b. **A:** Data for first unfolding event (red circles). Theoretical unfolding length for partial unfolding of pseudoknot 22/6b (stem2 unfolds [red, long-dash], stem1 unfolds [blue, dash-dot]) and partial unfolding of 22/6bU (stem2 unfolds [green, short-dash], stem1 unfolds [magenta, dash-dot-dot]). **B:** Data for second unfolding event (red circles). Theoretical unfolding length for complete unfolding of partially unfolded pseudoknot 22/6b (stem1 unfolds [red, long-dash], stem2 unfolds [blue, dash-dot]) and for pseudoknot 22/6bU (stem1 unfolds [green, short-dash], stem2 unfolds [magenta, dash-dot-dot]). *Please note:* The theoretical unfolding length for unfolding of stem1 is identical between 22/6b and 22/6bU in both A and B.

From Figure 41 it is clear that neither unfolding pathway of the two pseudoknots 22/6b and 22/6bU is able to explain the observed two-step unfolding events particularly well. It was not possible to find alternative structures which could explain the observed unfolding events.

The pattern of refolding for two-step unfolding tethers was similar to that of one-step unfolding with the exception that no two-step refolding was observed. A summary of unfolding and refolding data is listed in Table 13.

The refolding length for the two complete refolding events ( $19.8 \pm 2.0$  nm) is close to the expected refolding length of 16.3 nm for complete refolding of pseudoknot 22/6bU. However, the limited number of data points makes it difficult to exclude other possibilities. As for the one-step unfolding data, gradual+partial refolding was observed in approximately 50% of the retractions. The observed refolding length for the partial refolding event is very close to the value observed for the one-step unfolding data, and they probably reflect the same structural change. Gradual refolding was also observed in approximately 50% of the retractions and probably encompasses complete refolding of all three structures (22hp/b,

	Unfolding (Experimental)		Refolding (Experimental)		
	First	Second	Gradual	Gradual+Partial	Complete
Force [pN]	$16.7 \pm 2.2$	$24.2 \pm 2.5$	N.D.	$12.1 \pm 2.4$	$7.6 \pm 1.4$
Length [nm]	$11.5 \pm 0.8$	$14.0 \pm 1.0$	N.D.	$11.7 \pm 1.3$	$19.8 \pm 2.0$
	$n = 18$		$n = 8$	$n = 8$	$n = 2$

Table 13: Statistics for two-step unfolding of pseudoknot 22/6b. Unfolding/refolding force and unfolding/refolding length for each event. Values are mean  $\pm$  SEM.

22/6bU and 22/6b).

### Multi-Step unfolding

Multi-step unfolding refers to the situation shown in Figure 38D, where unfolding occurred through more than two distinct events. Only 13 such unfolding-traces was observed - mostly from the same tether. No attempts were made to assign structural changes to the individual unfolding steps. The refolding events observed during retraction showed that four out of six made complete refolding with a refolding length of 19.5 nm at 7.6 pN. This refolding length is quite close to the expected refolding length of 16.8 nm for complete refolding of pseudoknot 22/6bU. It is possible that these unfolding events reflect unfolding of unknown structures which subsequently fold into the expected structures.

#### 9.9.2 Kinetics of 22hp/b, 22/6bU, and 22/6b

We attempted to extract kinetic information for 22hp/b (subset D), 22/6bU (subset E), and 22/6b (subset F) from our force-ramp experiments. The analysis was conducted independently on data from the three disjoint subsets D, E and F (from Figure 40B) and the result is shown in Figure 42. The kinetic parameters from this analysis are listed in Table 14.

	$n$ (fit)	$X^\ddagger$	$k_0$
Subset D	40	$0.8 \pm 0.03$ nm	$4.6 \cdot 10^{-2} \pm 5.9 \cdot 10^{-3} s^{-1}$
Subset E	26	$0.5 \pm 0.10$ nm	$6.6 \cdot 10^{-2} \pm 3.1 \cdot 10^{-2} s^{-1}$
Subset F	16	$1.1 \pm 0.22$ nm	$1.4 \cdot 10^{-4} \pm 2.7 \cdot 10^{-4} s^{-1}$

Table 14: Results of kinetic analysis of subset D, E, and F. Values are fitting coefficients  $\pm$  95% confidence interval from fits in Figure 42.  $X^\ddagger$ : distance to transition state,  $k_0$ : zero force rate constant.

The parameters from the kinetic analysis listed in Table 14 provide some surprising information about the structures. For 22hp/b (subset D) the distance to transition state is very short indicating that it is a brittle structure. The value of  $k_0$  is also quite large compared to what has been previously observed for hairpins [Green et al., 2008]. These findings are somewhat surprising and in sharp contrast to the findings for the 10hp structure. The kinetic model does not describe the data for 22/6bU (subset E) particularly

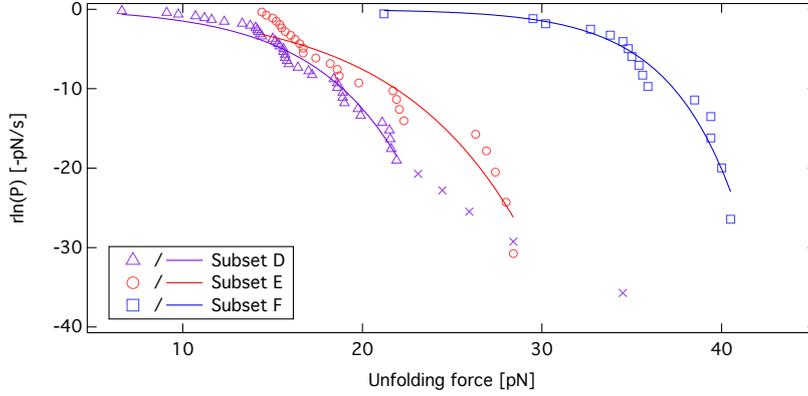


Figure 42: Kinetics of subset D (magenta, triangles), subset E (red, circles), and subset F (blue, squares).  $r \ln(P)$  plotted as a function of unfolding force. Equation (21) was fitted to data (solid lines). Loading rate was  $9.3 \pm 1.2 \text{ pN/s}$  (mean  $\pm$  SEM). Five data points was omitted from subset D (magenta, cross).

well but the parameters are within the range expected for a brittle structure such as a pseudoknot. Interestingly, although the difference is quite small, pseudoknot 22/6b (subset F) was identified as the least brittle structure of the three as described by  $X^\ddagger$ . The uncertainty for the estimate of  $k_0$  for subset F was unfortunately relatively large, but it appears that  $k_0$  for 22/6b could be smaller than for both 22hp/b (subset D) and 22/6bU (subset E).

### 9.9.3 Thermodynamics of 22hp/b, 22/6bU, and 22/6b

As previously, equilibrium information was extracted for subset D, E, and F from our unfolding experiments. The result from the CFT is shown in Figure 43 where work distributions for unfolding and refolding are depicted in red and magenta respectively. The data in the three subsets were analyzed independently of each other and the results are listed in Table 15

For 22hp/b (subset D) the estimated Gibbs free energy is slightly lower than the theoretical estimate from mfold. This is not surprising as the unfolding length for this subset was shorter than expected, which indicates that 22hp/b may exist primarily as a partially melted structure as indicated in Figure 44. For 22/6bU (subset E) the estimated free energy is also slightly smaller than the expected value. The free energy estimate of 22/6b (subset F) is much larger than expected. This could be due to the very limited information about the overlap between the two work distributions.

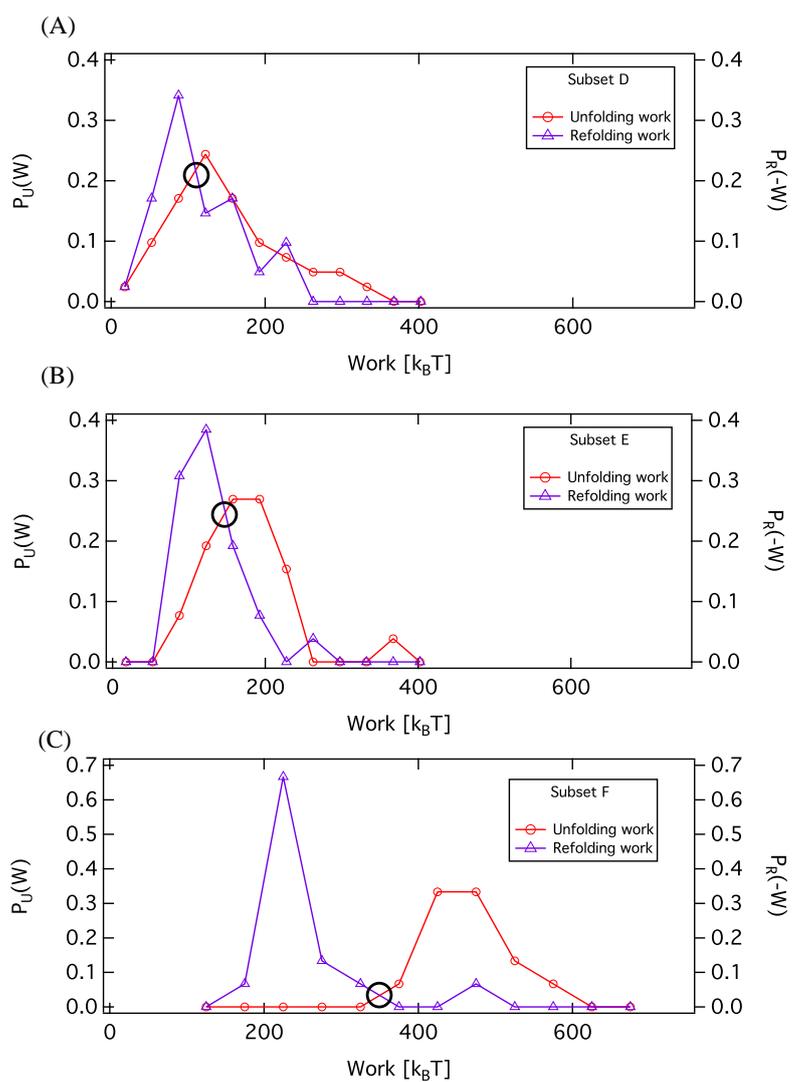


Figure 43: Work distributions for subset D, E and F. **A:** Data for 22hp/b (subset D),  $n = 41$ . **B:** Data for 22/6bU (subset E),  $n = 26$ . **C:** Data for 22/6b (subset F),  $n = 15$ .

Expected structure	Subset D	Subset E	Subset F
	22hp/b	22/6bU	22/6b
$\langle P_U(W) \rangle$	$147.5 \pm 11.6 \text{ k}_B\text{T}$	$174.0 \pm 10.5 \text{ k}_B\text{T}$	$480.9 \pm 20.6 \text{ k}_B\text{T}$
$\langle P_R(-W) \rangle$	$115.8 \pm 8.7 \text{ k}_B\text{T}$	$131.8 \pm 8.2 \text{ k}_B\text{T}$	$252.0 \pm 16.0 \text{ k}_B\text{T}$
$\langle W_{corr} \rangle$	$47.5 \pm 2.6 \text{ k}_B\text{T}$	$57.2 \pm 2.9 \text{ k}_B\text{T}$	$161.3 \pm 6.7 \text{ k}_B\text{T}$
$P_U(W) = P_R(-W)$	$109.8 \text{ k}_B\text{T}$	$147.5 \text{ k}_B\text{T}$	$350.0 \text{ k}_B\text{T}$
$\Delta G_{PK} / \Delta G_{HP}$	$36.9 \pm 1.5 \text{ kcal/mol}$	$53.6 \pm 1.8 \text{ kcal/mol}$	$111.9 \pm 4.0 \text{ kcal/mol}$
$\Delta G_{theory}$	$-47.5 \text{ kcal/mol}$	$-60.0 \text{ kcal/mol}$	$-66.9 \text{ kcal/mol}$

Table 15: Crooks analysis of subset D ( $n = 41$ ), subset E ( $n = 26$ ), and subset F ( $n = 15$ ). Mean work required for unfolding ( $\langle P_U(W) \rangle$ ) and mean work returned during refolding ( $\langle P_R(-W) \rangle$ ), intersect between the two work distributions ( $P_U(W) = P_R(-W)$  [ $\Delta G_T$ ]) and the mean correction for stretching of handles and RNA ( $\langle W_{corr} \rangle$ ). Estimated Gibbs free energy difference between folded and unfolded structure using Crooks Fluctuation Theorem ( $\Delta G_{PK}$  or  $\Delta G_{HP}$ ). Theoretical Gibbs free energy ( $\Delta G_{theory}$ ) was estimated using mfold [Zuker, M., 2003] (22hp/b) and pknotsRG [Reeder et al., 2007] (22/6b and 22/6bU). Mean  $\pm$  SEM.

### 9.9.4 Unfolding of 22hp/b, 22/6bU, and 22/6b - Summary

A summary of the observed unfolding and refolding events are shown in Figure 44 where unfolding is shown in blue. The complexity of the refolding pattern made it impossible to assign specific structural changes to the observed refolding events.

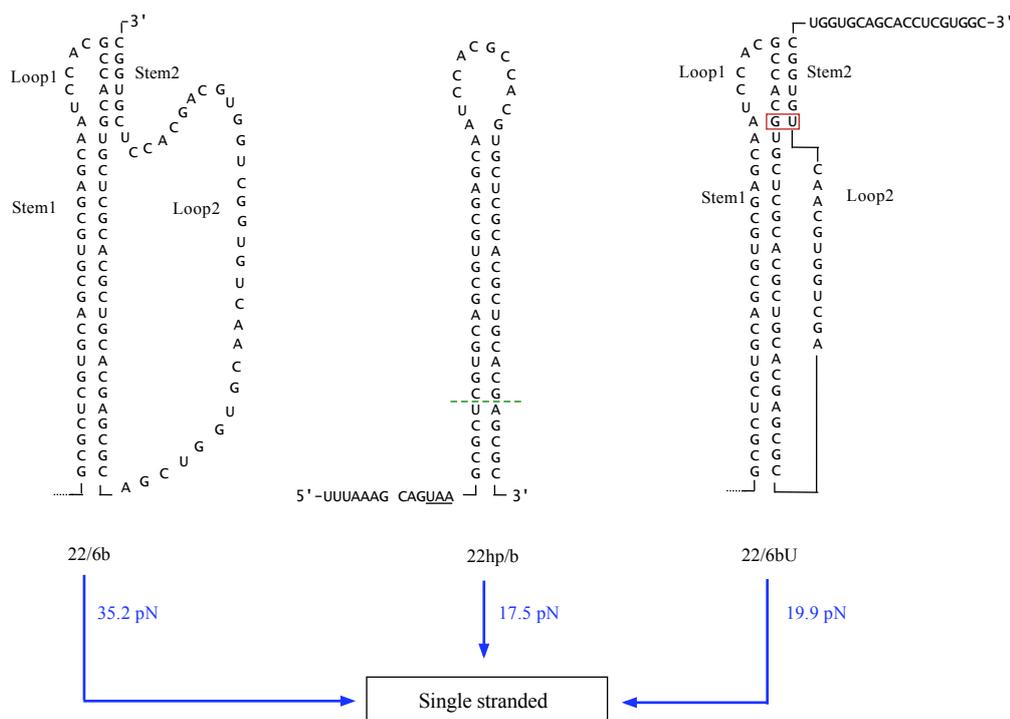


Figure 44: Observed unfolding of 22hp/b, 22/6bU and 22/6b. One-step unfolding (blue) was observed at the indicated forces. The folding intermediate 22hp/b could possibly exist in a partially unfolded state where the bottom of the stem, up to and including the first A-U base pair, is melted (green dashed line).

For the one-step unfolding there was generally a good agreement between the expected unfolding length and the observed unfolding length. However, the unfolding length of 22hp/b (subset D) was slightly shorter than the expected unfolding length. The observed unfolding length of  $17.6 \pm 0.6$  nm corresponds to unfolding of a partially melted version of 22hp/b lacking the bottom 5 bp (see Figure 44). The expected unfolding for this hairpin at 17.5 pN is 16.9 nm - in excellent agreement with the experimental data. This may also explain why the estimated Gibbs free energy was lower than expected - the theoretical Gibbs free energy for this 17bp hairpin is -33.8 kcal/mol. Thus, both the unfolding length and the estimated Gibbs free energy support the existence of this 17 bp hairpin.

The unfolding length observed for 22/6bU (subset E) was slightly longer than the expected unfolding length. Although the difference is quite small it could be explained by a deformation/bending of stem1 caused by the short loop2 sequence. Brierley et al. showed that a loop of 8 nucleotides was able to span an 11 bp stem while a 5 nucleotide

loop was unable to do so [Brierley et al., 1991]. This indicates that the 12 nucleotide loop2 of 22/6bU is stretched to its limit when spanning a 22 bp stem. This may cause a slight bending of stem1 - resulting in a longer than expected unfolding length.

## 9.10 Pseudoknot 22/6c

Pseudoknot 22/6c is structurally identical to pseudoknot 22/6a and 22/6b but has a further weakened stem1. As for pseudoknot 10/6 and 22/6b, there is a possibility of a folding intermediate hairpin and an alternative pseudoknot formed by the supposed loop2 sequence and the stem2 sequence - these three structures are shown in Figure 45.

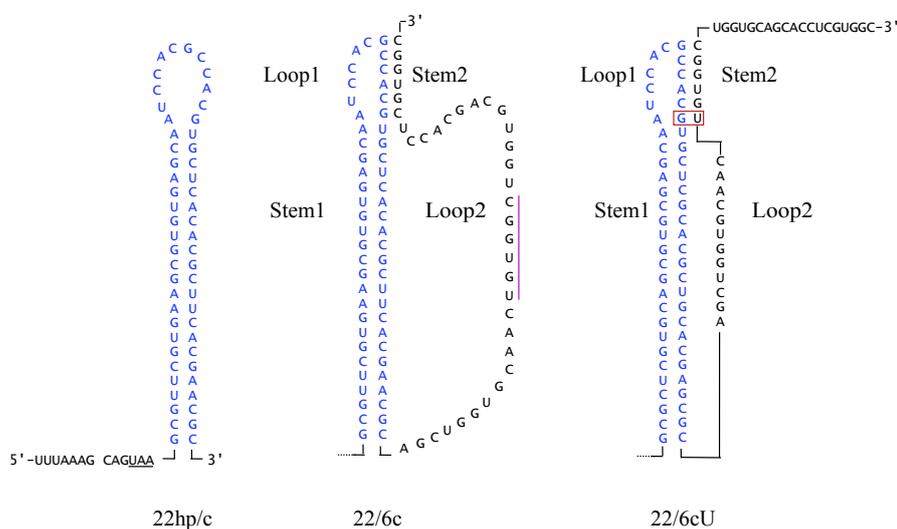


Figure 45: Possible structures for pseudoknot 22/6c. The folding intermediate “22hp/c” and the two pseudoknots 22/6c and 22/6cU. 22hp/c is a folding intermediate of both pseudoknots (blue). Magenta line denote alternative stem2 used in 22/6cU. Red box denote G·U wobble base pair. *Please note:* The downstream sequence is omitted in the depiction of 22hp/c.

### 9.10.1 Mechanical Unfolding

The data for unfolding of pseudoknot 22/6c, the member of the 22/6 structure family with the weakest predicted stem1, was similar to that of pseudoknot 22/6b except that the two-step unfolding events were more frequent. About 30% of the observed unfolding event occurred through two steps. Figure 46 show examples of the three major unfolding and refolding curves observed for pseudoknot 22/6c.

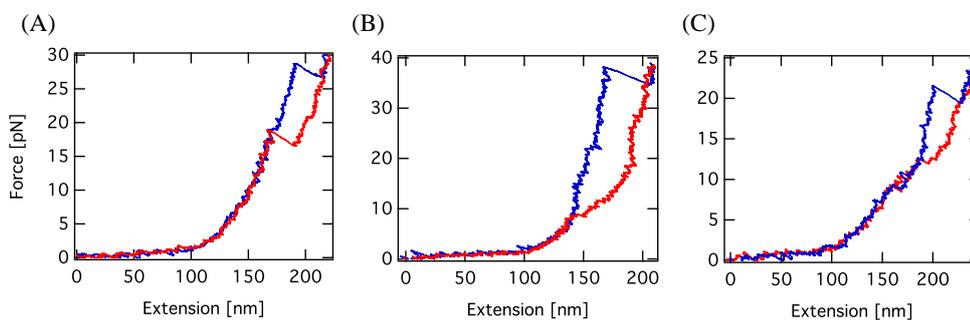


Figure 46: Examples of force-extension curves for pseudoknot 22/6c. Extension (blue) and retraction (red) **A**: Unfolding in one event with either partial or full refolding during retraction. **B**: Unfolding in one event and gradual refolding during retraction. **C**: Unfolding in two distinct events with mixed refolding during retraction.

### One-step Unfolding

The distribution of unfolding length is shown in Figure 47 and it appears that unfolding length was best described by a bimodal distribution.

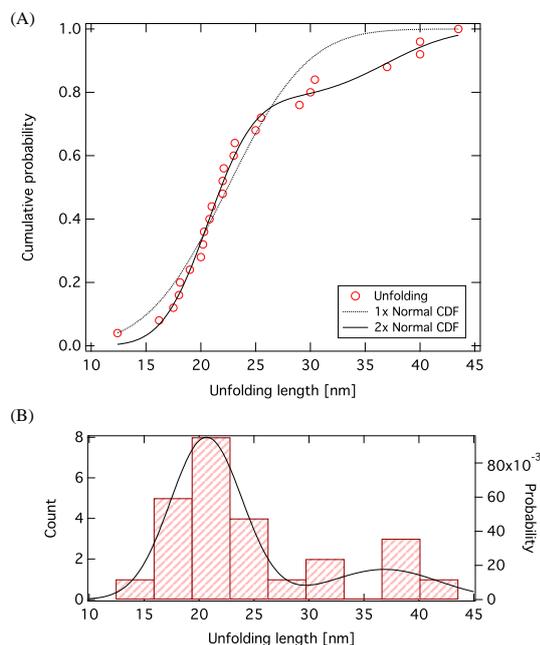


Figure 47: One step unfolding of pseudoknot 22/6b. **A**: Unfolding length (red, circles) displayed as a cumulative distribution. Solid lines are fit to a single normal CDF (black, dotted) and a double normal CDF (black, solid). **B**: Histogram of unfolding lengths with probability density function for a double normal distribution (black, solid) constructed from fit in A (right ordinate).  $n = 25$

It was decided to treat the unfolding lengths as originating from two underlying distributions and the dataset was divided in two subsets, subset G and subset H. Once again the separation was based on the intersection of the unfolding length ( $l$ ) distributions:

$$G = \{l \mid 0 \text{ nm} \leq F \leq 27.5 \text{ nm}\}$$

$$H = \{l \mid 27.5 \text{ nm} < F\}$$

A scatter plot of unfolding force versus unfolding length for each of the two subsets is shown in Figure 48 along with theoretical unfolding lengths for 22hp/c, 22/6cU and 22/6c.

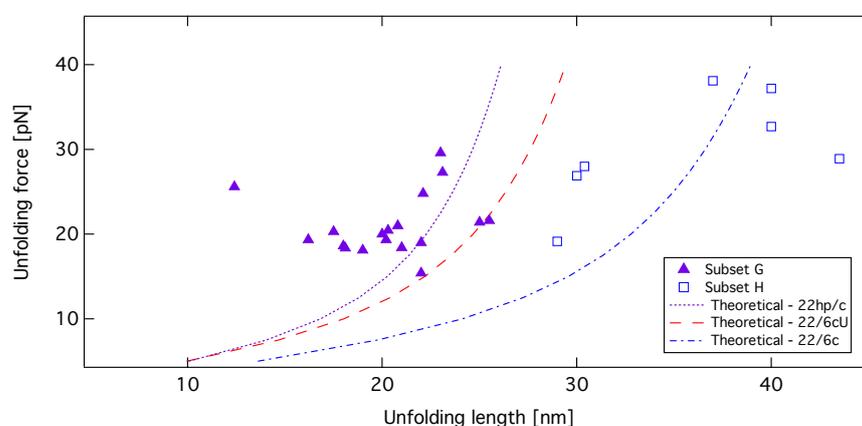


Figure 48: Unfolding force versus unfolding length of pseudoknot 22/6c. Data for subset G (magenta triangles,  $n = 18$ ) and subset H (blue squares,  $n = 7$ ). Lines indicate theoretical unfolding length versus force for 22hp/c, 22/6cU and 22/6c

As a consequence of the limited number of data points it was decided to treat subset G as originating from unfolding of 22hp/c and subset H as originating from 22/6cU and 22/6c. No attempts were made to separate unfolding events occurring from 22/6cU and 22/6c, and subset H was therefore not used in the subsequent analysis. However, the data in subset H indicates that the expected pseudoknots (22/6c and possibly 22/6cU) are able to form. Table 16 lists the observed unfolding force and lengths for subset H and subset G.

	$n$	Observed values		Theoretical lengths		
		$F$	$l$	22/6b	22/6cU	22hp/c
Subset G	18	$21.0 \pm 0.8$ pN	$20.3 \pm 0.8$ nm	33.3 nm	25.1 nm	22.6 nm
Subset H	7	$30.1 \pm 2.5$ pN	$35.7 \pm 2.2$ nm	36.6 nm	27.6 nm	24.7 nm

Table 16: Observed unfolding force and unfolding length for subset G and subset H along with theoretical unfolding lengths for the three structures 22/6c, 22/6cU and 22hp/c. Values are mean $\pm$ SEM.

The observed unfolding length for subset G is in good agreement with the expected unfolding length for 22hp/c, although it is a few nanometers shorter than expected. As for 22hp/b it is possible that this is caused by a lack of base pairing in the bottom of the stem. In fact, if this is true for 22hp/b it would most likely also be the case for 22hp/c as the stability of lower part of the stem is even further reduced compared to 22hp/b. The unfolding length of subset H is very close to the expected unfolding length for pseudoknot 22/6c.

The refolding events for pseudoknot 22/6c was a mix of several different refolding types as observed for the other pseudoknots. Refolding could occur gradually, partially+gradually or in a single step. The refolding observations are listed in Table 17.

	$n$	$F_r$	$l_r$
Gradual	5 (20%)	N.D.	N.D.
Gradual+partial	11 (44%)	11.7±0.9 pN	13.8±0.9 nm
Complete	9 (36%)	13.1±0.5 pN	17.6±1.0 nm

Table 17: Refolding data for 22/6c. Summary of the three different types of observed refolding. Values are mean±SEM for both subset G and subset H. N.D.: Not determined.

The refolding length for complete refolding was 17.6±1.0 nm which is quite close to the expected refolding length of 20.0 nm for complete refolding of 22hp/c. However, from the unfolding length of subset G we expect that the hairpin exists in a partially unfolded conformation. If we assume that complete refolding recreates this 17bp hairpin, we would expect a refolding length of 16.1 nm - in good agreement with the observed length.

Partial + gradual refolding was the predominant refolding pattern yielding a refolding length of 13.8±0.9 nm. This refolding length is insufficient to account for complete refolding of any known structure and it could describe partial refolding of stem1 (partial refolding of 22hp/c).

### Two-step Unfolding

About 30% of the tethers for 22/6c unfolded in two distinct events, a higher fraction than observed for pseudoknot 10/6 and 22/6b. As for the other pseudoknots, a scatter plot of unfolding force versus unfolding length was created for the first and the second event. Theoretical unfolding lengths for the two possible unfolding pathways for 22/6c and 22/6cU were added to the scatter plot to find the most probable explanation for the observed data. This is illustrated in Figure 49.

It seems that unfolding of pseudoknot 22/6c in the pathway: stem2 followed by stem1 is able to explain the observed unfolding lengths quite well. Each of the two sets of possible outliers indicated with parenthesis were observed in the same pull, i.e. the outliers denoted “1” were observed for one extension and the outliers denoted “2” were observed in another extension. It seems reasonable to conclude that the origin of these two pairs of unfolding data originates from an unknown structure. A summary of the two-step unfolding data is listed in Table 18.

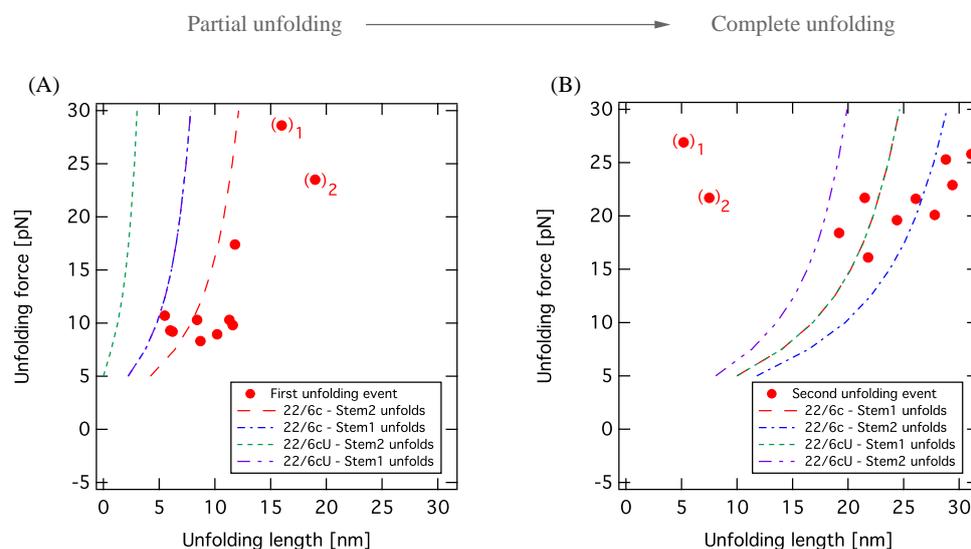


Figure 49: Two step unfolding of pseudoknot 22/6c. **A:** Data for first unfolding event (red circles). Theoretical unfolding length for partial unfolding of pseudoknot 22/6c (stem2 unfolds [red, long-dash], stem1 unfolds [blue, dash-dot]) and partial unfolding of 22/6cU (stem2 unfolds [green, short-dash], stem1 unfolds [magenta, dash-dot-dot]). **B:** Data for second unfolding event (red circles). Theoretical unfolding length for complete unfolding of partially unfolded pseudoknot 22/6c (stem1 unfolds [red, long-dash], stem2 unfolds [blue, dash-dot]) and for pseudoknot 22/6cU (stem1 unfolds [green, short-dash], stem2 unfolds [magenta, dash-dot-dot]). *Please note:* The theoretical unfolding length for unfolding of stem1 is identical between 22/6c and 22/6cU in both A and B. Two possible outliers are indicated with parenthesis and a subscript. The subscript indicate paired data, i.e. the two points denoted “1” originates from the same extension and the two points denoted “2” originates from the same extension.

	Refolding (Experimental)							
	Unfolding (Experimental)		One-step				Two-step	
	First	Second	Gradual	Partial	Complete	First	Second	
Force [pN]	$13.3 \pm 2.1$	$21.8 \pm 1.0$	N.D.	$12.4 \pm 0.4$	16.6	$12.8 \pm 0.6$	$9.2 \pm 0.8$	
Length [nm]	$10.4 \pm 1.3$	$22.1 \pm 2.6$	N.D.	$11.3 \pm 3.2$	20.0	$9.9 \pm 0.9$	$10.8 \pm 2.0$	
	$n = 11$		$n = 2$	$n = 2$	$n = 1$	$n = 6$		

Table 18: Statistics for two-step unfolding of pseudoknot 22/6c. Unfolding/refolding force and unfolding/refolding length for each event. Values are mean  $\pm$  SEM.

The observed complete refolding of 20.0 nm is in good agreement with the expected refolding length of 21.7 nm expected for complete refolding of 22hp/c. The partial refolding is not compatible with complete refolding of any of the two folding intermediates of pseudoknot 22/6c or 22/6cU. However, it could represent partial refolding of stem1 forming

a part of 22hp/c. The two-step refolding lengths do not correspond to complete refolding of any known structures.

### 9.10.2 Kinetics of hairpin 22hp/c

As for the other pseudoknots attempts were made to extract kinetic information about the unfolded structure. The dataset for pseudoknot 22/6c was so limited that the kinetic analysis was isolated to the 22hp/c hairpin (subset G). A plot of  $r \ln(P)$  versus unfolding force is shown in Figure 50.

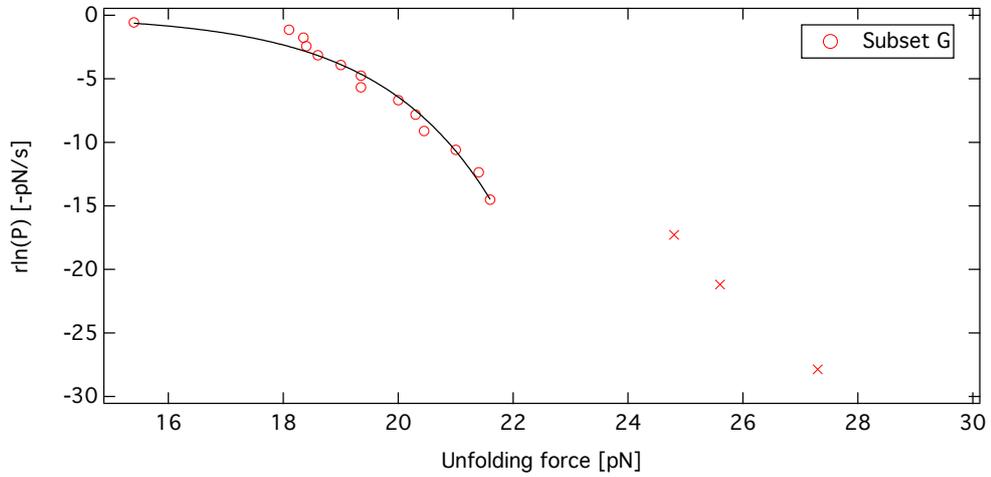


Figure 50: Kinetics of 22hp/c (subset G).  $r \ln(P)$  plotted as a function of unfolding force for subset G. Three points were excluded from the analysis (red cross). Equation (21) was fitted to data for traces with a single unfolding event (solid line). Loading rate was  $9.6 \pm 0.1 \text{ pN/s}$  (mean  $\pm$  SEM),  $n = 17$ .

The fitting parameters with 95% confidence intervals are listed in Table 19 and the kinetic parameters for 22hp/c are comparable to those for 10hp. Although the zero force rate constant ( $k_0$ ) is comparable to that of 10hp, the uncertainty of this parameter is quite large. The distance to transition state for 22hp/c is larger than those observed for our pseudoknots, this supports the notion that this is a hairpin structure.

	$n$ (fit)	$X^\ddagger$	$k_0$
Subset G	14	$2.1 \pm 0.26 \text{ nm}$	$1.3 \cdot 10^{-4} \pm 1.6 \cdot 10^{-4} \text{ s}^{-1}$

Table 19: Results of kinetic analysis of 22hp/c (subset G). Values are fitting coefficients  $\pm$  95% confidence interval from Figure 50.

### 9.10.3 Thermodynamics of hairpin 22hp/c

Once again the CFT was employed to extract equilibrium information from the unfolding and refolding data. The number of data points was very limited, only a total of 15 work values was obtained from the 18 unfolding/refolding traces assigned to 22hp/c (subset G). A histogram of work used during unfolding and work returned during refolding is shown in Figure 51.

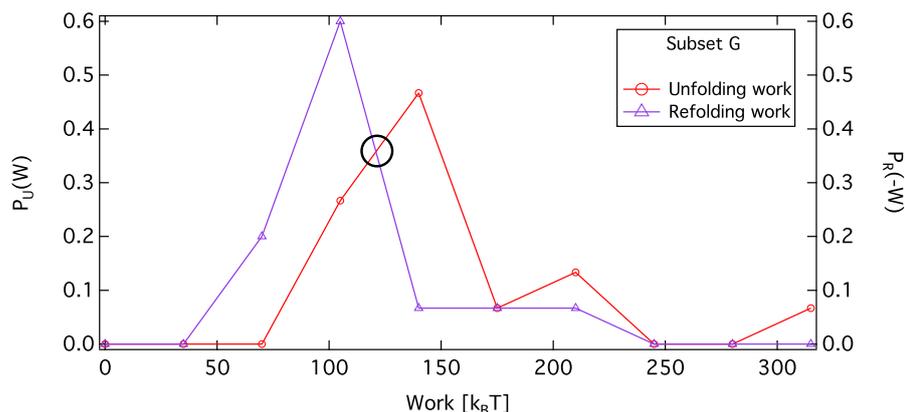


Figure 51: Work distributions for 22hp/c (subset G). Probability of work used during unfolding (red, circles) and probability of work returned during refolding (magenta, triangles).  $n = 15$

As for the same analysis of the other structures the unfolding and refolding work was binned into equally spaced bins and the intersection between the two was used as  $W = \Delta G_T$ . The result is listed in Table 20 and show that the estimate from the CFT is in excellent agreement with the theoretical free energy of 22hp/c.

Expected structure	Subset G
	22hp/c
$\langle P_U(W) \rangle$	$167.5 \pm 13.7 \text{ k}_B\text{T}$
$\langle P_R(-W) \rangle$	$126.8 \pm 9.0 \text{ k}_B\text{T}$
$\langle W_{corr} \rangle$	$52.1 \pm 4.5 \text{ k}_B\text{T}$
$P_U(W) = P_R(W)$	$120.9 \text{ k}_B\text{T}$
$\Delta G_{HP}$	$-40.8 \pm 2.7 \text{ kcal/mol}$
$\Delta G_{theory}$	$-41.4 \text{ kcal/mol}$

Table 20: Crooks analysis of subset G. Mean work required for unfolding ( $\langle P_U(W) \rangle$ ) and mean work returned during refolding ( $\langle P_R(-W) \rangle$ ), intersect between the two work distributions ( $P_U(W) = P_R(-W) [\Delta G_T]$ ) and the mean correction for stretching of handles and RNA ( $\langle W_{corr} \rangle$ ). Estimated Gibbs free energy of 22hp/c using Crooks Fluctuation Theorem ( $\Delta G_{HP}$ ). Theoretical Gibbs free energy ( $\Delta G_{theory}$ ) was estimated using mfold [Zuker, M., 2003]. Mean  $\pm$  SEM.  $n = 15$ .

### 9.10.4 Unfolding of Hairpin 22hp/c - Summary

The analysis of this pseudoknot was hampered by the limited number of data points available, however, we did observe unfolding events in good agreement with the expected RNA structures. Only unfolding data for subset G was used in the kinetic analysis and for the estimate of free energy. The results obtained from subset G were all consistent with 22hp/c or a partially folded version of 22hp/c. A summary of the observed structural changes and the forces at which they occur is shown in Figure 52.

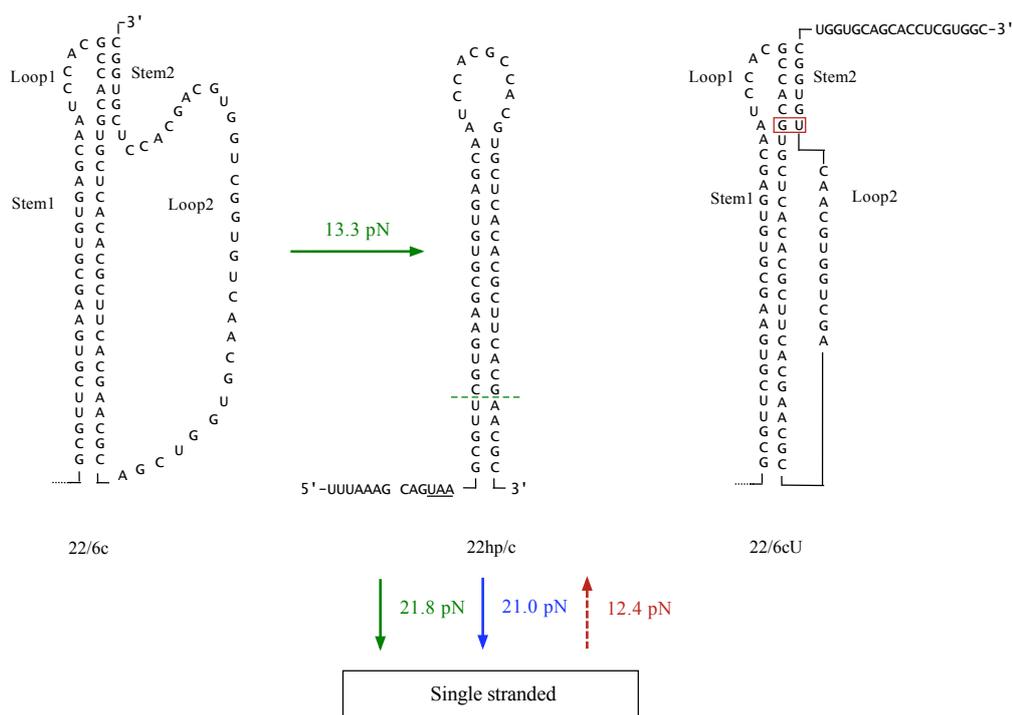


Figure 52: Observed unfolding and refolding for 22/6c structures. One-step unfolding (blue) and two-step unfolding (green) was observed at the indicated forces. Possible partial refolding of 22hp/c was observed at the indicated force (red). The folding intermediate 22hp/c could possibly exist in a partially unfolded state where the base of the stem, up to and including the second A-U base pair is melted (green dashed line).

Although subset H contains unfolding of 22/6c, and possibly 22/6cU, they have not been included in Figure 52 as the number of data points makes correct classification impossible.

The unfolding length of 22hp/c was shorter than expected (as observed for 22hp/b in Figure 40 (page 92)). This supports the notion that the first 5 bp of the stem is melted and the stem consists of only 17 bp.

## 9.11 Summary - Single Molecule Experiments

The following section contains a quick recap of the results presented above (the refolding data have been omitted - see discussion).

### 9.11.1 Unfolding data

A scatter plot of unfolding force versus unfolding length for all the unfolded structures is shown in Figure 53. As expected, the unfolding length of 10hp is shorter than the unfolding length of 22hp/b and 22hp/c. The unfolding length of 10/6U is shorter than the unfolding length of 22/6bU and the unfolding length of pseudoknot 10/6 is shorter than that of pseudoknot 22/6b. The unfolding data for pseudoknot 11/6, from Rebecca Bolt Ettliger, has also been included in the figure [Ettliger, 2012].

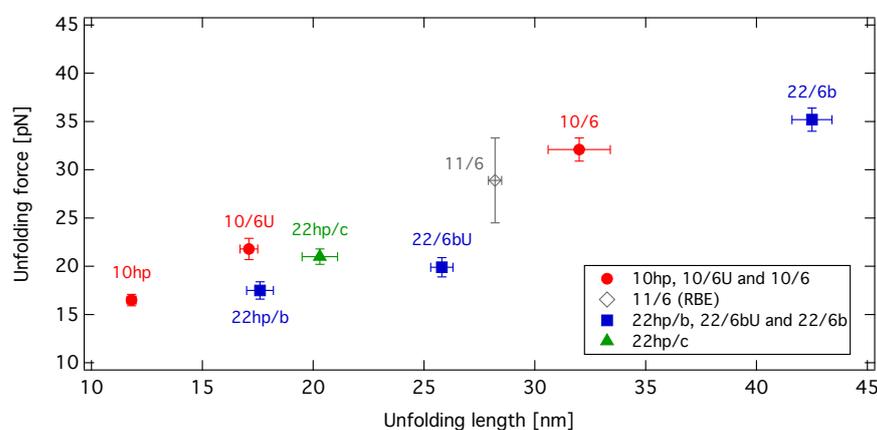


Figure 53: Summary of one-step unfolding. Unfolding force versus unfolding length for 10hp ( $n = 39$ ), 10/6U ( $n = 21$ ), 10/6 ( $n = 14$ ), 11/6 ( $n = 8$ ), 22hp/b ( $n = 46$ ), 22/6bU ( $n = 27$ ), 22/6b ( $n = 17$ ) and 22hp/c ( $n = 18$ ). Data for 11hp and 11/6 are from [Ettliger, 2012]. Points are mean  $\pm$  SEM.

In general the experimental unfolding lengths were in good agreement with the theoretical estimates obtained through simulations (the simulations are described in Part IV). A plot of experimental unfolding length versus theoretical unfolding length is shown in Figure 54. It is possible that the underestimation of unfolding length for 22hp/b and 22hp/c is caused by melting of the bottom 5 bp in the hairpin, as illustrated in Figure 44 and Figure 52.

### 9.11.2 Structural kinetics

Table 21 lists the observed kinetic parameters determined from the one-step unfolding data presented above. It was shown that both 10hp and 22hp/c were less brittle than all pseudoknots while 22hp/b seemed to resemble pseudoknots in brittleness and zero force rate constant.

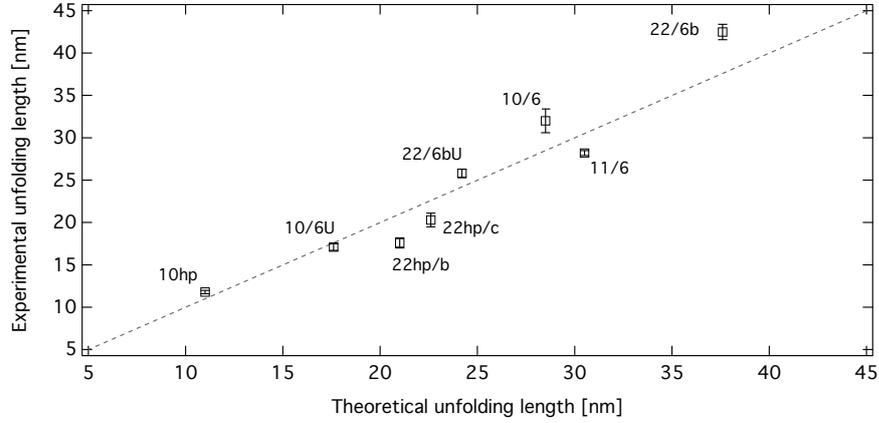


Figure 54: Experimental unfolding length plotted as a function of theoretical unfolding length. Dotted line indicate perfect match between experimental and theoretical value. Data for 11hp and 11/6 are from [Ettlinger, 2012]. Experimental data are mean $\pm$ SEM.

	Structure	$n$ (fit)	$X^\ddagger$	$k_0$
Subset A	hairpin 10hp	30	$3.3 \pm 0.28$ nm	$1.9 \cdot 10^{-5} \pm 1.9 \cdot 10^{-5} s^{-1}$
Subset B	pseudoknot 10/6U	20	$0.7 \pm 0.09$ nm	$3.0 \cdot 10^{-2} \pm 1.4 \cdot 10^{-2} s^{-1}$
Subset C	pseudoknot 10/6	13	$0.9 \pm 0.07$ nm	$1.3 \cdot 10^{-3} \pm 7.3 \cdot 10^{-4} s^{-1}$
Subset D	hairpin 22hp/b	40	$0.8 \pm 0.03$ nm	$4.6 \cdot 10^{-2} \pm 5.9 \cdot 10^{-3} s^{-1}$
Subset E	pseudoknot 22/6bU	26	$0.5 \pm 0.10$ nm	$6.6 \cdot 10^{-2} \pm 3.1 \cdot 10^{-2} s^{-1}$
Subset F	pseudoknot 22/6b	16	$1.1 \pm 0.22$ nm	$1.4 \cdot 10^{-4} \pm 2.7 \cdot 10^{-4} s^{-1}$
Subset G	hairpin 22hp/c	14	$2.1 \pm 0.26$ nm	$1.3 \cdot 10^{-4} \pm 1.6 \cdot 10^{-4} s^{-1}$
	10/6U + 11hp	145	$0.75 \pm 0.02$ nm	$3.0 \cdot 10^{-2} \pm 2.0 \cdot 10^{-2} s^{-1}$
	Pseudoknot 11/6	8	$0.32 \pm 0.06$ nm	$7.0 \cdot 10^{-2} \pm 2.0 \cdot 10^{-3} s^{-1}$
	6/11: Unknown structures	51	$1.56 \pm 0.08$ nm	$1.0 \cdot 10^{-3} \pm 4.0 \cdot 10^{-4} s^{-1}$

Table 21: Summary of structural kinetics. Subset A-G: Kinetic parameters from fits in Figure 31, Figure 42 and Figure 50. Values are fitting parameter  $\pm$  95% confidence interval. Bottom three rows: Data from Rebecca Bolt Ettlinger [Ettlinger, 2012]. Values are mean $\pm$ asymptotic standard error

### 9.11.3 Thermodynamics

The results obtained through the CFT are summarized in Figure 55. In general there was excellent agreement between the theoretical Gibbs free energy and the free energy obtained from the experimental data, however, the free energy estimate of 22/6b was 70% higher than expected.

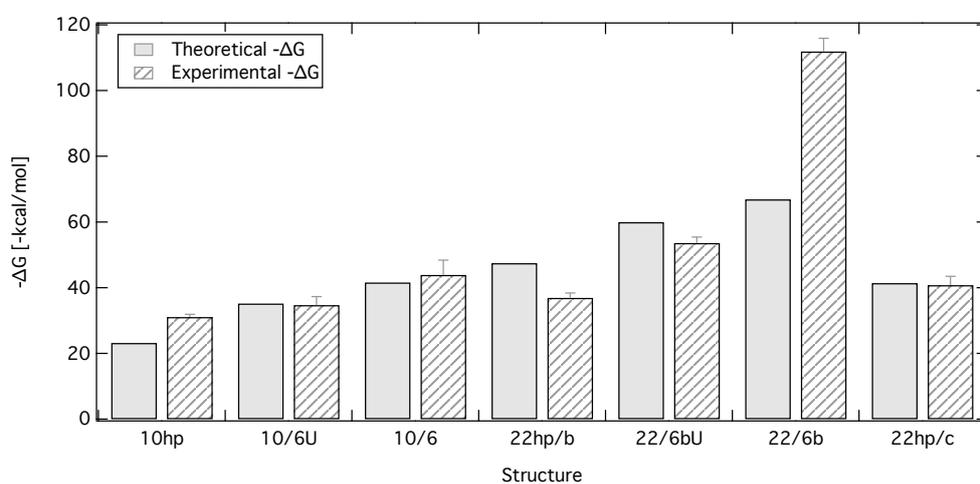


Figure 55: Summary of results from the Crooks Fluctuation Theorem. Theoretical Gibbs free energy (gray) and experimental Gibbs free energy (hatched) for the indicated structures. Experimental data are mean  $\pm$  SEM.

## 10 Discussion - Single Molecule Experiments

### 10.1 Unfolding Length and Unfolding Force

In general, the observed unfolding lengths were in excellent agreement with the theoretical unfolding lengths, typically deviating less than 15% - corresponding to a few nano-meters. This is very satisfying considering the difficulties we had in obtaining a reliable estimate of the distance-per-voltage conversion factor ( $\beta$ ) and the level of bead fluctuations caused by the galvanometric mirrors (Figure 23, page 70). This level of accuracy was also predicted from the error analysis shown in Figure 22 (page 69). The vertical behavior of the extension trace up to 50 pN in Figure 25 (page 72) also demonstrate that the estimate of  $\beta$  accurately describes the displacement of the beads in their respective traps. If the estimate of  $\beta$  was inaccurate, the extension trace would deviate from vertical as the beads are linked by an inflexible tether.

For the sequence designed to fold into pseudoknot 10/6, the unfolding data were divided into three subsets A, B, and C. Unfolding lengths corresponding to both hairpin 10hp (subset A) and pseudoknot 10/6 (subset C) were recovered from the dataset. Additionally, an alternative pseudoknot, called 10/6U, was also predicted from the unfolding data (subset B). This pseudoknot was not designed purposely, and is a result of a near-cognate stretch of nucleotides in loop2 of the original pseudoknot. The alternative pseudoknot (10/6U) was not predicted by `pknotsRG` unless the 3' hexa-nucleotide sequence designed to form the last strand of stem2 was omitted from the submitted sequence. This explains why the alternative pseudoknots were not identified in the design phase. Interestingly, we predicted the existence of this alternative pseudoknots from our single molecule experiments and subsequently found them using `pknotsRG`.

For the sequence designed to fold into pseudoknot 11/6 investigated by Rebecca Bolt Ettliger, both the expected pseudoknot, the alternative pseudoknot (11/6U), and the 11hp folding intermediate was recovered. Unfortunately, it was not possible to separate the data of 11/6U and 11hp and the listed kinetic parameters listed are consequently not reflecting the kinetic parameters of a single structure [Ettliger, 2012].

The sequence designed to fold into an "inverted" pseudoknot did not fold into the expected structure despite having a slightly higher predicted stability than pseudoknot 11/6 (due to possible base pairing in the loop). A mix of unknown structures was recovered for this sequence [Ettliger, 2012].

For the sequence designed to fold into pseudoknot 22/6b, the unfolding data were also divided into three subsets: D, E, and F. Although the final separation of subset D and E was performed with a structural assumption (Figure 40, page 92) the unfolding lengths were once again in good agreement with three possible structures. The unfolding length of 22hp/b (subset D) was, however, 3 nm shorter than expected. This difference could be caused by an inaccurate estimate of  $\beta$ , but since subset D is composed of data from several different experiments this is unlikely (assuming that we are not subject to systematic errors). Another explanation is that part of the 22hp/b stem (bottom 5 bp) is unfolded/melted as indicated with a green dashed line in Figure 44, yielding a 17 bp hairpin with a 12 nt loop. Such a structure would result in an unfolding length of 17.0 nm

at 17.5 pN - in excellent agreement with the experimental data. This could also explain why the experimental estimate of Gibbs free energy of subset D is smaller than expected for 22hp/b.

Our dataset for the sequence designed to form pseudoknot 22/6c is quite limited. A total of 25 one-step unfolding events were observed and they were divided into subset G and subset H (page 102). The majority of the data in subset G most likely describes unfolding of the folding intermediate hairpin 22hp/c. If the bottom of 22hp/b are melted under our experimental conditions as discussed above, we would expect the same for 22hp/c. Indeed, for 22hp/c we find an unfolding length which is 2 nm shorter than expected. This indicates that, at least during our single molecule experiments, the stem of 22hp/b and 22hp/c consists of only 17 bp. Subset H most likely describes a mix of unfolding of pseudoknot 22/6cU and 22/6c, but the limited number of data points made it impossible to separate the two. Subset H does, however, suggest that pseudoknot 22/6c and 22/6cU are able to form.

A few interesting features of the structures are visible from Figure 53 (page 108). First, the mechanical stability of the structures all followed the pattern hairpin < alternative pseudoknot < designed pseudoknot. Second, there was no difference in the mechanical stability between pseudoknot 10/6U and 22/6bU and between pseudoknot 10/6 and 22/6b. This is interesting as the free energy of the large 22/6b pseudoknot is larger than that of pseudoknot 10/6. This apparent lack of correlation between free energy and unfolding force and between free energy and frameshift efficiency has been observed previously [Chen et al., 2009, Napthine et al., 1999].

The similarity between the unfolding force of 10/6U and 22/6bU and between 10/6 and 22/6b could indicate a common point of failure within these two groups. Two-step unfolding of sequence 10/6 and sequence 22/6c identified stem2 as a potential common point of failure. If stem2 is the structural weak-point responsible for the structural failure leading to unfolding, one would expect that a weaker stem2 should result in a lower mechanical stability. Our results support this hypothesis - both 10/6U, 11/6U, and 22/6bU have lower unfolding forces than their larger counterparts: 10/6, 11/6, and 22/6b respectively. In pseudoknot 10/6U, 11/6U, and 22/6bU, stem2 is weakened by substitution of a G-C base pair with a G·U wobble relative to pseudoknot 10/6, 11/6, and 22/6b. It is possible that the wobble base pair weakens the stem not only as a consequence of reduced hydrogen bonding but also through a structural change [Varani and McClain, 2000]. However, as the length and composition of loop2 is also different between 10/6U and 10/6, between 11/6U and 11/6, and between 22/6bU and 22/6b, any loop2 contribution, such as tertiary interactions, cannot be excluded [Chen et al., 2009, Su et al., 1999]. Another possibly explanation is that the short loop2 in pseudoknot 10/6U and pseudoknot 22/6bU induces an internal structural strain which could explain why stem2 fails at a lower force for 10/6U and 22/bU compared to 10/6 and 22/6b. If this were true, however, one might expect the effect to be stem1-length dependent, but we observe no such correlation (compare 10/6U and 22/6bU).

The theoretical unfolding length predictions shown in Figure 30 (page 78), Figure 36

(page 88), Figure 41 (page 94), and Figure 49 (page 104) raises an important question: is the high prevalence of the stem1 folding intermediate (10hp, 22hp/b, and 22hp/c) caused by inadequate spatial resolution in our force-extension curves or does it reflect a true property of the RNA pool? As the predicted unfolding length for unfolding of stem2 in the alternative pseudoknots (10/6U, 16/6U, 22/6bU and 22/6cU) is less than 5 nm for forces below 40 pN it is reasonable to assume that we would not be unable to detect these short unfolding events. As a consequence, we are not able to directly transfer the structural distributions from our single molecule experiments to the structural distribution *in vivo*.

Despite intense efforts we were unable to obtain any tangible unfolding data for pseudoknot 16/6 and 22/6a, but from the *in vivo* experiments in Part I we know that some structure exists on the mRNA. Since the unfolding data presented in this section indicates that the expected pseudoknot structures for 10/6, 11/6, 22/6b, and 22/6c are able to form, we have no reason to assume that pseudoknot 16/6 and in particular pseudoknot 22/6a are unable to form. As we were able to reach the over-stretch regime (at around 65 pN, not shown) without observing structural unfolding, it is appealing to conclude that pseudoknots 16/6 and 22/6a are too strong to unfold in the force regime investigated here. Obviously, making such a conclusion based on negative results poses a severe problem. Consequently, we are currently performing the same experiments in a buffer without  $Mg^{2+}$ , ions which are known to increase the mechanical stability RNA structures [White et al., 2011, Green et al., 2008, Onoa et al., 2003, Liphardt et al., 2001].

## 10.2 Refolding Length and Refolding Force

The observed refolding typically consisted of a combination of a single well defined step and gradual refolding. Consequently, structural assignment for the refolding data is uncertain. Although it would have been elegant to show that each of our structures refolds through a single step, the observed pattern is neither surprising nor concerning. The pattern is similar to the refolding pattern observed by Chen et al. for both hairpins and pseudoknots derived from the human telomerase RNA [Chen et al., 2007], by Li et al. for the TAR RNA hairpin [Li et al., 2006b], and by Chen et al. for several variants of the  $\Delta U177$  pseudoknot (derived from the human telomerase RNA) [Chen et al., 2009].

There are many possible explanations for the refolding pattern - the simplest explanation is that refolding follows a more complex path and goes through different states than unfolding [Tinoco et al., 2006]. Green et al. obtained estimates of  $X^\ddagger$  for unfolding and refolding of both an IBV derived pseudoknot and a related hairpin (hpS12), and observed that the sum of the two  $X^\ddagger$ 's did not correspond to the total unfolding length of the molecule. This indicates either that the refolding process is not simply the "reverse" of the unfolding process or that  $X^\ddagger$  does not describe a real physical distance along the reaction coordinate [Green et al., 2008]. Also, Li et al. observed that the loading rate had an influence on the refolding length for refolding of the TAR RNA hairpin, adding to the complexity of interpreting refolding data [Li et al., 2006b]. The mechanism behind a clear

refolding event, is a continuous formation of base pairs which forces the beads away from the center of their traps. This suggests that the trap stiffness must have an impact on the probability of observing a refolding event and on the length of the observed refold event. For very stiff traps the formation of base pairs holds insufficient energy to pull the beads away from the trap centers and no clear refolding events should ever be observed (possibly with the exception of very fast retraction speeds). Other factors which could have an impact on refolding include buffer components and length of the RNA/DNA handles (long handles might add some flexibility to the setup).

### 10.3 Unfolding Kinetics

We extracted parameters describing structural kinetics from our force-ramp experiments by fitting equation (21) to an appropriately constructed probability distribution. In general we observed that hairpin structures were less brittle than pseudoknots, in agreement with the findings of others. Our estimates of  $X^\ddagger$  for pseudoknots are in excellent agreement with those determined by others: 0.2 nm (with  $\text{Mg}^{2+}$ ) [Hansen et al., 2007], 0.8 nm (no  $\text{Mg}^{2+}$ ) [Chen et al., 2007], and 1.4 nm (with  $\text{Mg}^{2+}$ ) [Green et al., 2008]. Our estimates of  $X^\ddagger$  for the folding intermediates (10hp, 22hp/b and 22hp/c) is generally somewhat smaller than that found for hairpins by others. We estimate  $X^\ddagger$  to be 3.3 nm, 0.8 nm and 2.1 nm for the three hairpins respectively while others have estimated  $X^\ddagger$  for varying hairpin structures to be 5-20 nm (no  $\text{Mg}^{2+}$ ) [Woodside et al., 2006], 7-11 nm (no  $\text{Mg}^{2+}$ ) [Chen et al., 2007], 12 nm (with and without  $\text{Mg}^{2+}$ ) [Liphardt et al., 2001, Green et al., 2008], and 8 nm (no  $\text{Mg}^{2+}$ ) [Li et al., 2006b]. Although our short estimate of  $X^\ddagger$  could reflect a real physical difference it cannot be excluded that other factors contribute to this discrepancy, e.g. pH, salts, breaking tethers, and loading rate.

The intrinsic difference between the direction of the applied force between hairpins and pseudoknots makes it difficult to compare the value of  $X^\ddagger$  between these structural classes. For hairpin structures the force is applied perpendicular to the stem while the force is applied parallel to the stems for pseudoknots. Thus, while it is possible that  $X^\ddagger$  describes an actual deformation (i.e. melting) of hairpin structures it is more difficult to interpret the value of  $X^\ddagger$  for pseudoknots. Furthermore, for different pseudoknots with different tertiary structures, the value of  $X^\ddagger$  may reflect completely separate structural features depending on what determines the “length” of the structure in the direction of the applied force. Hence, the current understanding of pseudoknots as brittle structures may not apply in general.

The parameter  $k_0$  is an estimate of the zero force rate constant and describes the lifetime of the structure in absence of force. Using this parameter along with the estimate of  $X^\ddagger$ , enables us estimate the lifetime of our structures at arbitrary forces using equation (18). With the exception of hairpin 22hp/b, the estimates of  $k_0$  indicate that while the hairpins have longer lifetimes at zero force, they quickly become less stable than pseudoknots under applied force. This is in excellent agreement with the findings of Green et al. for an IBV derived pseudoknot and a related hairpin [Green et al., 2008].

In our kinetic investigation we used equation (21) (page 62) to obtain the estimates of

$X^\ddagger$  and  $k_0$ . This equation was partially based on a version of the Arrhenius equation in which the *activation energy* ( $E_a$ ) was expressed as  $-FX^\ddagger$ . Although this makes sense from a classical mechanics perspective, it renders the interpretation of the parameters difficult. One of the assumptions of equation (21) is that force favors the unfolded structure - implying that force itself changes the energy landscape. Consequently, the values of  $X^\ddagger$  and  $k_0$  are, to some extent, force dependent [Green et al., 2008, Dudko et al., 2006]. The value of  $k_0$  is thus not a very good descriptor of the zero force rate constant and should only be used to estimate rate constants at forces near the detected unfolding force [Tinoco, 2004].

Additionally,  $k_0$  may contain components from the handles and other factors from the experimental setup [Li et al., 2006b, Liphardt et al., 2001] and it follows from equation (18) that only one-step transitions can be modeled. The latter assumption may be invalid for unfolding of complex structures like pseudoknots, where unfolding could occur through one or more intermediate steps.

Here we will limit the interpretation of  $k_0$  to state that the estimates of  $k_0$  for our pseudoknots are in good agreement with those found for IBV inspired pseudoknots, while our estimates of  $k_0$  for hairpins are much larger than those reported for the IBV-hairpin hps12 [Green et al., 2008, Hansen et al., 2007].

The unfolding and refolding rate constants could have helped explain the apparently high prevalence of folding intermediate hairpins (10hp, 11hp, 22hp/b, and 22hp/c). Unfortunately, the complex refolding pattern prevented the estimation of refolding rate constants for our pseudoknots. However, if we assume that our pseudoknots have similar refolding rate constants as those of the IBV pseudoknot determined by Green et al. it is possible to estimate the expected distribution of hairpin versus pseudoknot in solution. These rate constants indicate that almost all RNA will be folded into pseudoknots at zero force [Green et al., 2008]. It therefore seems reasonable, given the short unfolding length expected for stem2 unfolding of pseudoknots 10/6U, 11/6U, 22/6bU, and 22/6cU, that the apparent prevalence of folding intermediate hairpins observed here is caused by our limited spatial resolution. This is also supported by our estimates of  $\Delta G$ , from which one can estimate the equilibrium constant.

Although more sophisticated models have been developed by Olga K. Dudko and colleagues [Dudko et al., 2006] to extract kinetic information, the limited number of data points in each of our subsets makes it unfeasible to apply these complex models.

## 10.4 Thermodynamics

The large amount of dissipated work observed here, all but eliminates the applicability of Jarzynski's equality as the number of unfolding events required to obtain a reasonable estimate of  $\Delta G$  is extremely large. Consequently, we used CFT as described by Collin et al. to estimate the free energy of our structures [Collin et al., 2005] and a summary is shown in Figure 55 (page 110). In general there is excellent agreement between the theoretical and experimental estimates, which further supports our structural assignments. It also illustrates the applicability of the CFT even for datasets of very limited size ( $n < 20$ ), even for large values of dissipated work. The error-estimation of the free energy estimates

is somewhat rudimentary as it only takes variation of the correction term into account. The actual uncertainty of the estimates are larger than indicated, as the estimated work is affected by every element in the experiments (e.g. conversion factors, unfolding force, and unfolding length).

The excellent agreement between the estimates of  $\Delta G_{\text{PK}}$  and  $\Delta G_{\text{HP}}$  and the theoretical values, indicate that for the structures investigated here, tertiary interactions are not adding significantly to the overall stability as observed for other pseudoknots [Su et al., 1999, Chen et al., 2009]. The theoretical estimate from `pknotsRG` and `mfold` is based on the Turner rules (1M NaCl, 37°C) and they do not take possible tertiary interactions into account [Reeder et al., 2007, Zuker, M., 2003, Walter et al., 1994, Mathews et al., 1999].

The estimates of  $\Delta G_{\text{PK}}$  and  $\Delta G_{\text{HP}}$  were based on the intersection between the histograms with evenly distributed bins for unfolding and refolding work as no information about their true distribution is known (i.e. they do not necessarily follow a Gaussian distribution). This approach was also used by Green et al. and Collin et al. to estimate the free energy of hairpins and pseudoknots from similar experiments [Collin et al., 2005, Green et al., 2008]. Possibly, this introduces some bias depending upon bin width. The result, however, did not change substantially if the distributions were modeled as Normal distributions (not shown).

As mentioned in Section 9 there is a structural bias in the calculation of the correction term used to isolate  $\Delta G_{\text{PK}}$  and  $\Delta G_{\text{HP}}$  from  $\Delta G_T$ . As the correction term increases from hairpin to pseudoknot it is unlikely that the pattern of increasing free energy observed for 10hp to 10/6U to 10/6 and for 22hp/b to 22/6bU to 22/6b is an artifact caused by the correction term bias.

The estimate of  $\Delta G_{\text{HP}}$  for 22hp/b was lower than expected. However, as discussed above, it is possible that 22hp/b exists in a partially melted form. The theoretical free energy of this 17 bp version of 22hp/b is -33.8 kcal/mol - in excellent agreement with the experimental estimate of -36.9 kcal/mol. Although, the estimate of  $\Delta G_{\text{HP}}$  for 22hp/c was larger than the expected free energy of the 17 bp version of 22hp/c, which was the preferred structure from the unfolding length observations. However, as the estimate of  $\Delta G_{\text{HP}}$  is associated with a higher degree of uncertainty than the unfolding length, we will not change the structural assumption based solely on  $\Delta G_{\text{HP}}$ .

The average dissipated work observed for 10/6 and 22/6b is significantly larger than what has previously been observed when using the CFT to estimate free energy differences [Green et al., 2008, Collin et al., 2005]. Although the CFT has no formal upper limit for the dissipated work, the average dissipated work of almost 250  $k_B T$  observed for 22/6b appears to be close to the upper limit (for the width of the work distributions obtained here). The large value of dissipated work means that we have very little information about the overlap between the unfolding work distribution and the refolding work distribution (see Figure 43C, page 97). This may also explain why  $\Delta G_{\text{PK}}$  for 22/6b is much larger than the theoretical estimate.

Had the dissipated work been much smaller (close to zero), we could have estimated the free energy simply by integrating the area of the unfolding event (i.e. integrating from point A to point C in Figure 68, page 159) [Green et al., 2008, Li et al., 2006b, Tinoco et al., 2006].

The fact that we have a mixed structural population, consisting of at least two different pseudoknots for each structure investigated here, means that other methods to determine the Gibbs free energy, e.g. fluorescence competition assays, would result in erroneous estimates. Here we are able to separate the individual structures, and we can estimate the Gibbs free energy of each structure independently [Liu et al., 2009].

## 10.5 Mechanical Stability and Frameshifting Efficiency

The correlation between mechanical stability and frameshifting efficiency (*in vivo*) was first proposed by Hansen et al. based on results from two highly similar IBV inspired pseudoknots [Hansen et al., 2007]. A correlation between mechanical stability and frameshifting efficiency (*in vitro*) was also observed by Chen et al. for several mutations of the  $\Delta$ U177 pseudoknot [Chen et al., 2009]. In a thorough investigation using several different IBV derived pseudoknots Green et al. failed to observe such a correlation (*in vitro*) [Green et al., 2008]. Green et al. did, however, observe a correlation between the rate constant of unfolding in a narrow force window and frameshift efficiency [Green et al., 2008]. If we assume that the mechanical unfolding performed here, to some degree, resembles the mechanism of unfolding during translation, then the rate constant ( $k(F)$ ) could have a significant effect on the ability to induce frameshift. Not only does the rate constant describe for how long the structure is able to resist the forward pressure of the ribosome, but the refolding rate constant also describes how fast the unfolded structure refolds and is primed to induce frameshift in subsequent cycles of translation. Interestingly, the estimates of  $k_0$  obtained by Hansen et al. also suggests that the structural lifetime under applied force could be important for frameshift efficiency [Hansen et al., 2007] (supporting information).

The findings of this study does not contradict hypothesis of a correlation between mechanical stability and frameshifting efficiency between similar structures. We found that neither the mechanical stability nor the frameshift efficiency of pseudoknots 10/6 and 11/6 was statistically significantly different (p-value = 0.1261 [frameshift efficiency] and p-value = 0.3893 [mechanical stability], Two-tailed Student's t-test). As we were unable to obtain the mechanical stability of pseudoknots 22/6a and 22/6c it is not possible to verify the correlation for these structures. This correlation could also be elusive for these structures, as 22/6a is able to stall translating ribosomes.

## 10.6 Breaking Tethers

It is clear from Table 3 (page 73) that most of our tethers survive only a few cycles of extension and relaxation. It is difficult to asses if this is unusual compared to other studies, as those numbers are seldom published. However, Chen et al. encountered the same complication in their investigation of the  $\Delta$ U177 pseudoknot [Chen et al., 2009]. They obtained between 96 and 761 unfolding traces from between 2 and 13 different tethers ([Chen et al., 2009], supplementary information).

Many factors may contribute to the limited lifetime of our tethers and a combination of several factors are possible. Two factors, the limited mechanical strength of the digoxigenin- $\alpha$ -digoxigenin connection which has a propensity to break at forces above 45

pN, and the effect of oxygen damage are especially likely to contribute [Landry et al., 2009, Chen et al., 2007]. The link between digoxigenin and  $\alpha$ -digoxigenin is known to break at forces above 45 pN [Chen et al., 2007]. In addition Landry et al. have suggested that the high laser intensity near the polystyrene beads can create reactive oxygen species which leads to a decreased lifetime of the link [Landry et al., 2009]. Woodside et al. they have added oxygen scavenging components to their buffer to counteract this effect [Woodside et al., 2006]. If this hypothesis is true, the dual beam setup of the JPK NanoTracker™ would be especially vulnerable, as we have two beams and two polystyrene beads in very close proximity.

## Part IV

# Optical Tweezers - Simulations

## 11 Introduction - Nucleic Acid as Biopolymers

Having spent most of my education with molecular biology I had very little experience in the field of biophysics in general and optical tweezers specifically. I had numerous questions related to the practical work onto which I was about to embark and the equipment I was about to use. Some of these are listed below:

- How is the force-extension trace and unfolding length affected by misalignment in the  $z$ -level between the two traps?
- How to estimate the length of unfolding for any given RNA structure?
- Is there a difference in unfolding length between single-trap optical tweezers and dual-trap optical tweezers?
- How does the trap stiffness affect the unfolding length?
- How sensitive is the predicted unfolding length to variations in the polymer parameters ( $L_p$ ,  $L_c$  and  $K$ )?

As these questions were difficult to answer from existing literature, a simulation framework was constructed in which various optical tweezers scenarios could be tested. The framework was created *in silico* and centers around the extensible worm like chain model (EWLC) used to describe the elastic properties of biopolymers such as polynucleotides.

### 11.1 Stretching of Biopolymers

Like DNA and RNA, all biopolymers consists of monomers linked together to form a continuous tether and quite a few models have been developed to describe the stretching of such polymers. Two models are typically used to describe the stretching of RNA and DNA, the Worm Like Chain model (WLC) and the EWLC mentioned above. The WLC, the simplest of the two, is an explicit function which describes the entropic loss created when a polymer is stretched [Bustamante et al., 1994].

$$F = \left( \frac{k_B T}{L_p} \right) \left[ \frac{1}{4(1 - x/L_c)^2} - \frac{1}{4} + \frac{x}{L_c} \right] \quad (29)$$

where  $k_B$  is the Boltzmann constant,  $T$  the absolute temperature,  $x$  the end-to-end extension of the polymer,  $L_p$  the persistence length and  $L_c$  the contour length. The contour length is the length one must travel when traversing the entire polymer from one end to the other. According to the WLC the end-to-end distance of the polymer is a fraction of the contour length and will move towards the contour length at high force. The force originates from the loss of entropy created when reducing the number of possible conformations of

the monomers of the polymer. The persistence length is formally the distance over which orientation correlation is lost and is expressed at the bending rigidity in units of  $k_B T$ . In other words, the persistence length describes the flexibility of the tether - a large value signifies a stiff polymer, a small value signifies a flexible polymer.

The WLC provides a good description of the elastic properties of the polymer at forces below 5 pN above which the entropic loss is not enough to account for the structural changes in the polymer. At forces above 5 pN a significant portion of the applied force is used to stretch the chemical bonds in the polymer giving rise to an enthalpic change. To account for this effect the EWLC model was created [Wang et al., 1997]:

$$F = \left( \frac{k_B T}{L_p} \right) \left[ \frac{1}{4(1 - x/L_c + F/K)^2} - \frac{1}{4} + \frac{x}{L_c} - \frac{F}{K} \right] \quad (30)$$

where the stretch modulus  $K$  have been introduced. The stretch modulus describes the polymers intrinsic resistance towards strain in the longitudinal direction. A correlation between the stretch modulus and the persistence length is typically observed for biological polymers - a stiff polymer tends to resist strain to a higher degree [Wang et al., 1997]. It should be noted that for forces above  $\sim 40$  pN the twist-stretch coupling should be taking into account, however, the EWLC should be sufficiently accurate in the force regime investigated here [Gross et al., 2011]. A requirement for both the WLC and EWLC is that  $L_c \gg L_p$ .

## 11.2 Polymer Parameters

Although the WLC and the EWLC have been used in the investigation of both RNA hairpins and RNA pseudoknots it is not obvious what the exact values of the different physical parameters are [Collin et al., 2005, Green et al., 2008, Hansen et al., 2007, Liphardt et al., 2001].

In a thorough investigation of persistence length and stretch modulus Wang et al. used different molecules in various buffers and showed that both persistence length and stretch modulus varied with buffer and with the contour length of the polymer [Wang et al., 1997]. The persistence length was estimated to be around 47-42 nm and the stretch modulus was estimated to be 1 nN for a 600 nm double stranded DNA tether. More recently Wen et al. investigated the persistence length of RNA/DNA handles like those used in our experiments and found that a persistence length of 10 nm was obtained from WLC fits to DNA/RNA hybrids of 1 kbp [Wen et al., 2007]. Wen et al. estimated the  $L_p$  by fitting the WLC to a tether consisting of both RNA/DNA handles and single stranded RNA. As the following simulations will demonstrate it is, however, possible that this low value of  $L_p$  is caused by the presence of single stranded RNA.

It has been found that RNA/DNA hybrids form an A-type helix and not the typical B-form helix found in double stranded DNA, and the A-type helix  $L_c$  must be used to describe the RNA/DNA hybrid [Milman and Langridge, 1967].

For the simulations here I use  $L_p = 45$  nm for dsDNA,  $L_p = 1$  nm for single stranded RNA,  $L_c = 0.28$  nm/bp for RNA/DNA handles,  $L_c = 0.338$  nm/bp for dsDNA,  $L_c =$

0.59 nm/nt single stranded RNA and single stranded DNA,  $K = 1\text{nN}$  for RNA/DNA and  $K = 800\text{ pN}$  for ssRNA [Liphardt et al., 2001, Wang et al., 1997, Milman and Langridge, 1967, Smith et al., 1996].

## 12 Implementation

The code was implemented in the Python programming language and used as a command-line tool [Python Software Foundation, 2012]. The purpose was to create a simulation framework which, by moving in very small steps, could accurately describe how any biopolymer tether would behave during a force-extension experiment. In the following the term “force” refers to the length of force vector ( $F = |\vec{F}|$ ).

Consider the situation illustrated in Figure 56 where a tether (red) is suspended between two spherical particles aligned in the  $x$ -direction but not in the  $z$ -direction.

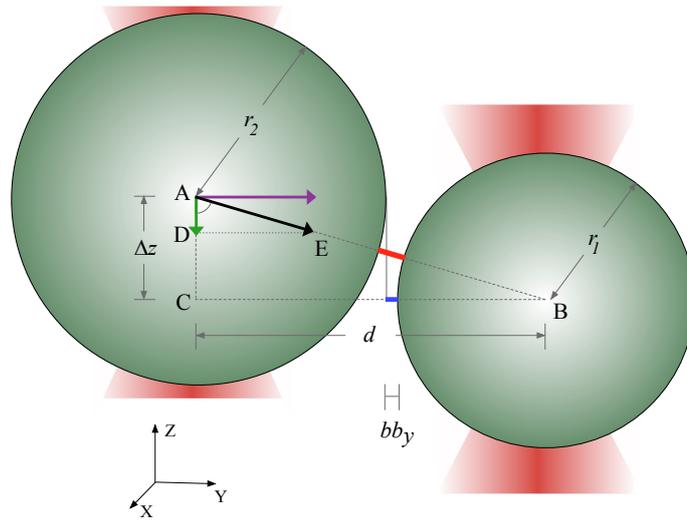


Figure 56: Geometry of misalignment. The centers of two beads with radii  $r_1$  and  $r_2$  are separated by a  $y$ -distance  $d$  and a  $z$ -misalignment of  $\Delta z$ . The actual extension of the tether (red) is larger than the apparent extension of the tether (blue,  $bb_y$ ). The force acting in the  $yz$ -plane (black,  $AE$ ) can be dissolved into two orthogonal components, one in the  $z$ -plane (green,  $AD$ ) and one in the  $y$ -plane (magenta,  $DE$ ).

The pulling geometry is defined by the triangle  $ABC$  which, in turn, is defined by the  $z$ -level misalignment ( $\Delta z$ ) and the distance between the center of the two particles. The force geometry is defined by  $ADE$ . As we increase the force on the tether its length is increased and  $d$  becomes larger. The angle  $a$  increases as a consequence of bead displacement in the  $z$ -direction ( $\Delta z$  decreases) and extension of the tether. The net effect is:

$$\lim_{|\vec{F}_{yx}| \rightarrow \infty} a = 90^\circ$$

Thus, if we want to model the illustrated situation for both single- and dual trap optical tweezers, we need two parameters for each trap (the spring constants in the  $y$  and  $z$  directions) and one parameter describing the  $z$ -level misalignment (set to zero for no misalignment). In addition we need parameters describing  $L_c$ ,  $L_p$ , and  $K$  of the tether, the radii of the particles and the distance between them.

The following pseudocode describes the simulation process for the described system:

```

Step 1. Initialize (i.e. enter z-level misalignment and bead sizes)
Step 2. Calculate trap-trap distance for zero extension of the tether
Step 3. Calculate y-directional surface distance between beads and pulling angle (a)
Step 4. Increase trap-trap distance - maintain bead-bead distance (increases force)
    - Calculate resulting y-directional movement in each trap
    - Increment y-directional force
    - Calculate movement in z-direction
    - Increment z-directional force
    - Update z-level misalignment
    - Increment yz-directional force
    - Calculate the yz-directional bead-bead surface distance (actual tether extension)
    - Calculate the tether length from the EWLC at the applied force
    WHILE EWLC length of tether > bead-bead yz-distance:
        - Move beads towards the center of their trap (decrease force)
        - Calculate new pulling angle
        - Calculate new force on tether
        - Calculate movement in z-direction
        - Calculate length of tether from EWLC
    WHILE EWLC length of tether < bead-bead yz-distance:
        - Move beads away from center of their trap (increase force)
        - Calculate new pulling angle
        - Calculate new force on tether
        - Calculate movement in z-direction
        - Calculate length of tether from EWLC
Step 5. Goto step 4.

```

Each step is described in detail in the following sections.

### 12.1 Step 2: Calculate trap-trap distance for zero extension of the tether

For a given  $z$ -level misalignment ( $\Delta z$ ) the distance between the center of the two beads ( $d$ ) was calculated assuming that the surface distance between the two beads in the  $yz$ -direction is 0 nm. This corresponds to zero extension of the tether (red in Figure 56).

$$d = \sqrt{(r_1 + r_2)^2 - \Delta z^2} \quad (31)$$

where  $r_1$  and  $r_2$  is the radii of the two trapped particles.

### 12.2 Step 3: Calculate surface distance and pulling angle

The  $y$ -directional surface distance ( $bb_y$ , blue in Figure 56) is then:

$$bb_y = d - (r_1 + r_2) \quad (32)$$

The pulling angle ( $a$ ) is:

$$a = \arcsin\left(\frac{d}{\sqrt{d^2 + \Delta z^2}}\right) \quad (33)$$

### 12.3 Step 4. Increase trap-trap distance - maintain bead-bead distance

As this takes place within a loop each equation describes the  $i$ -th step in the iteration. The trap-trap distance ( $d_{tt}$ ) is incremented with an appropriate small step size ( $l_{step}$ ):

$$d_{tt,i} = dd_{tt,i-1} + l_{step} \quad (34)$$

The force is increased as we move the trap center away from the bead center. The absolute movement of the two beads in the  $y$ -direction is determined by the spring constants for the two traps ( $\kappa_{y1}$  and  $\kappa_{y2}$ ). If we maintain the bead positions (constant  $d$ ) as we increase the trap-trap distance the total movement of the beads ( $x_t$ ) must be the sum of the movement of the two beads ( $x_1$  and  $x_2$ ):

$$x_t = x_1 + x_2 = l_{step} \quad (35)$$

The movement in the individual traps can now be calculated :

$$(l_{step} - x_{y2,i}) \kappa_{y1} = x_{y2,i} \kappa_{y2} \Leftrightarrow x_{y2,i} = l_{step} \left( \frac{\kappa_{y1}}{\kappa_{y1} + \kappa_{y2}} \right) \quad (36)$$

$$x_{y1,i} = l_{step} - x_{y2,i} \quad (37)$$

The  $y$ -directional force in each of the two traps ( $F_{y1}$  and  $F_{y2}$ ):

$$F_{y1,i} = F_{y1,i-1} + x_{y1,i} \kappa_{y1} \quad (38)$$

$$F_{y2,i} = F_{y2,i-1} + x_{y2,i} \kappa_{y2} \quad (39)$$

The force in the  $z$ -direction can be calculated from the  $y$ -directional force and the pulling angle:

$$F_{z,i} = \frac{F_{y1,i}}{\tan(a)} \quad (40)$$

this assumes that we move in sufficiently small steps to consider  $a$  as a constant. The total movement of the two trapped particles in the  $z$ -direction can be calculated from the force and the spring constants ( $\kappa_{z1}$  and  $\kappa_{z2}$ ). The movement of each bead is away from the focus of their respective traps and towards a common  $z$ -level:

$$x_{z1,i} = \frac{F_{z,i}}{\kappa_{z1}} \quad (41)$$

$$x_{z2,i} = \frac{F_{z,i}}{\kappa_{z2}} \quad (42)$$

The  $z$ -level misalignment can now be updated from the the movement in the two traps:

$$\Delta z_i = \Delta z_0 - (x_{z1,i} + x_{z2,i}) \quad (43)$$

The true force acting on the tether in the  $yz$ -plane is calculated from its components:

$$F_{yz,i} = \sqrt{F_{y1,i}^2 + F_{z1,i}^2} \quad (44)$$

The bead-bead surface distance in the  $yz$ -plane ( $bb_{yz}$ , the actual extension of the tether, red in Figure 56) can be calculated:

$$bb_{yz} = \sqrt{\Delta z^2 + (bb_y + r_1 + r_2)^2} - (r_1 + r_2) \quad (45)$$

Since we know the force on the tether ( $F_{yz}$ ) and the tethers composition, we can use the EWLC to calculate the theoretical length ( $e$ ) of the tether at the applied force. This was done by numerically solving the EWLC:

$$0 = \frac{k_B T}{L_p} \left( \frac{1}{4(1 - e/L_c + F_{yz}/K)^2} + \frac{1}{4} + \frac{e}{L_c} - \frac{F_{yz}}{K} \right) - F_{yz} \quad (46)$$

The appropriate root was found using a bisection algorithm. Although slower, the bisection algorithm proved far more stable at higher forces than the Newton-Raphson method. If the tether in question is a combination of single stranded and double stranded RNA/DNA the EWLC was solved for each component and the total length of the tether was the sum of the length for each component<sup>9</sup>.

### 12.3.1 WHILE EWLC length of tether > bead-bead yz-distance

We now know the actual extension of the tether and what the length should be at the applied force. If the theoretical extension from the EWLC is longer that the actual extension, we need to move the beads away from each other and towards the center of their traps. This increases the actual extension and lowers the force - and thus the theoretical extension using the EWLC. By moving back in very small steps we find a point where the actual extension is equal the the theoretical extension at the applied force. At this point the force and extension are in balance and we exit the while-loop and go back to step 4 for another cycle of trap movement. In the simulations conducted here, a step size of 2 pico-meters were used when incrementing the bead-bead distance.

If the applied force reached 0 pN (the two beads have reached the center of their traps) and the condition was still fulfilled - we exit and go to step 4 for another cycle.

<sup>9</sup>For short components the  $L_c \gg L_p$  condition is not met, however, the impact on the estimated length of the tether is very small

### 12.3.2 WHILE EWLC length of tether < bead-bead yz-distance

In this situation the actual length is longer than the theoretical extension and we need to decrease the actual extension. This is achieved by moving the beads towards each other, reducing the actual extension and increasing the force. Again, by moving the beads in very small steps we reach a point where the actual extension equals the theoretical extension at the applied force. As above the step size for decrementing the bead-bead distance was 2 pico-meters.

## 12.4 Summary of Implementation

For each simulation we get a multidimensional vector of points where the force and the length of the tether is in balance. The number of points are determined by the step-size ( $l_{step}$ ), the tether, and force-limit at which the simulation is stopped - the curves shown here consists of  $\sim 60,000$  points. At each of these points we get the following information:

Parameter	Description
$F_{yz}$	Applied force in the $yz$ -direction
$F_y$	Applied force in the $y$ -direction
$F_z$	Applied force in the $z$ -direction
$l_{tether}$	Length of tether from EWLC at the applied force ( $F_{yz}$ )
$bb_y$	Distance between bead surfaces in the $y$ -direction
$\Delta z$	$z$ -level misalignment
A	Pulling angle
$m_{z1}$	$z$ -directional movement of bead in trap1
$m_{z2}$	$z$ -directional movement of bead in trap2
$l_{dsDNA}$	Length of double stranded DNA (or RNA/DNA hybrid handle)
$l_{ssDNA}$	Length of single stranded DNA (or RNA)
$l_{dsDNA,PK}$	Length of double stranded DNA (or RNA) in pseudoknot
$l_{ssRNA,PK}$	Length of single stranded DNA (or RNA) in pseudoknot

Table 22: Parameters returned from simulations described in the text. At each point in the simulation we get the listed parameters.

If there is no  $z$ -level misalignment the value of  $F_y$  equals that of  $F_{yz}$ ,  $F_z$  is 0 pN,  $l_{tether}$  equals that of  $bb_y$ , pulling angle A equals  $\pi/2$  and  $m_{z1}$  and  $m_{z2}$  is 0 nm for all points.

## 13 Results

Unless stated otherwise the simulations conducted here are under the assumption of no  $z$ -level misalignment.

### 13.1 Force-extension Curve of a DNA Tether

To test the simulation, the extension of a 1 kb dsDNA tether was simulated using  $L_p = 45$  nm,  $L_c = 0.338$  nm/bp, and  $K = 1$  nN. The simulated curve along with the fitted EWLC (up to 30 pN) is shown in Figure 57

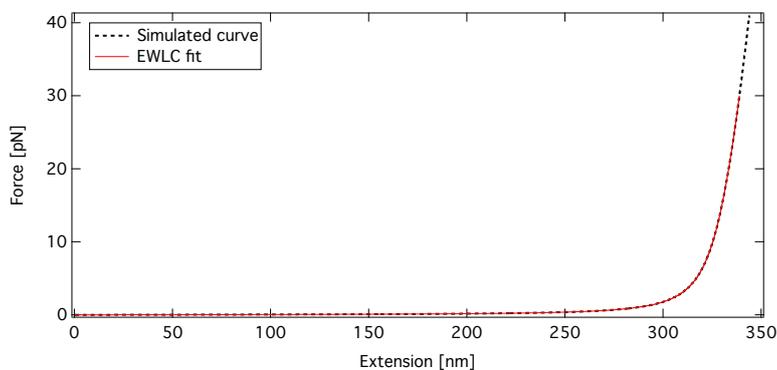


Figure 57: Simulated force-extension curve for a 1kb dsDNA tether (black, dashed) with fitted EWLC (red solid). Fitting parameters for the EWLC was  $L_p = 45.0$  nm,  $L_c = 338.0$  nm, and  $K = 1000.0$  pN.

It is clear that the simulated curve is in accordance with the EWLC as expected. The fitting parameters returned were exactly as entered in the simulation.

### 13.2 Force-extension Curve for a Compound Tether

In our experiments, and in all other similar experiments, the tether suspended between the trapped beads contains a mix of different elements. Specifically, the tether consists of both RNA/DNA hybrids, single stranded RNA and double stranded RNA (for pseudoknots). Wen et al. found that the persistence length of such a compound tether estimated from the WLC was significantly shorter than that of dsRNA and that the persistence length was smaller for shorter tethers [Wen et al., 2007]. Wen et al. state that they do not fully understand why the apparent persistence length changes with tether-length.

It is possible that this effect is caused by the single stranded RNA present in the tether. To test if the presence of single stranded nucleotides can change the apparent persistence length in a length dependent manner, the situation was simulated. Two double stranded DNA tethers of different length without or with single stranded nucleotides were “stretched” and the EWLC was fitted to the force-extension curves to obtain estimates of  $L_p$ ,  $L_c$  and  $K$ . The result of these simulations is shown in Figure 58.

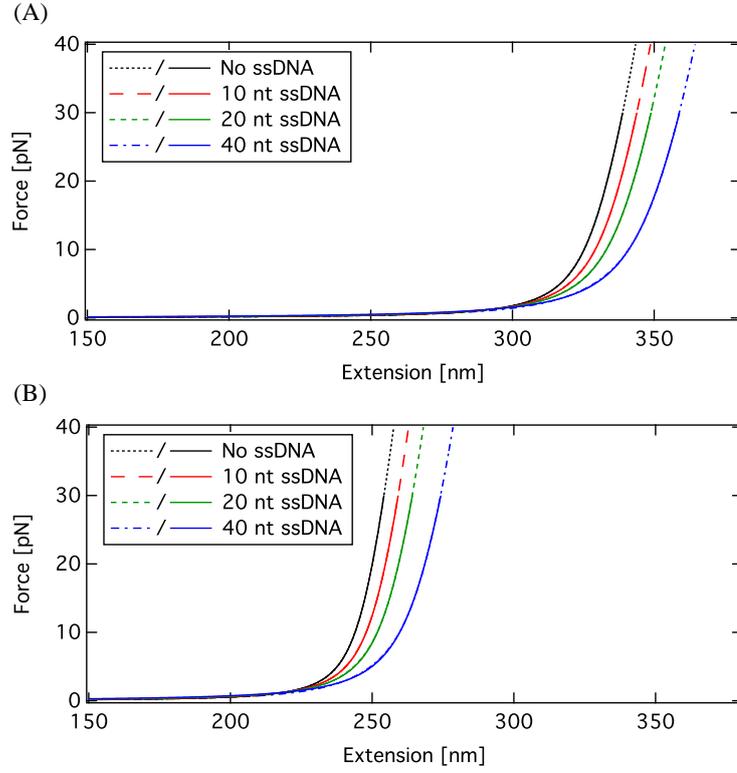


Figure 58: Impact of ssDNA on polymer parameters. **A**: Double stranded DNA tether of 1000 bp without ssDNA (black, dotted), with 10 nt ssDNA (red, long dashed), with 20 nt ssDNA (green, short dashed) and with 40 nt ssDNA (blue, dash-dot). Solid lines are fitted EWLC. **B**: Double stranded DNA tether of 750 bp without ssDNA (black, dotted), with 10 nt ssDNA (red, long dashed), with 20 nt ssDNA (green, short dashed) and with 40 nt ssDNA (blue, dash-dot). Solid lines are fitted EWLC.

The EWLC fitting parameters are listed in table 23 and show that both the number of ssDNA nucleotides and the length of the tethers have an impact on the predicted polymer parameters. Although, the exact parameters are not identical to those obtained by Wen et al. the pattern is identical [Wen et al., 2007].

ssDNA	750 bp dsDNA			1000 bp dsDNA		
	$L_p$	$L_c$	$K$	$L_p$	$L_c$	$K$
0 nt	45 nm	254 nm	1000 pN	45 nm	338 nm	1000 pN
10 nt	35 nm	259 nm	967 pN	37 nm	344 nm	977 pN
20 nt	29 nm	265 nm	938 pN	32 nm	349 nm	955 pN
40 nt	21 nm	276 nm	888 pN	24 nm	361 nm	916 pN

Table 23: Summary of EWLC parameters from Figure 58.

### 13.3 Practical Implications of $z$ -level Misalignment

The dual-beam setup used in this study, as with most other optical tweezers, is sensitive towards a misalignment of the two traps in the  $z$ -level. This is particularly important for short tethers as those used in Part III. The sensitivity originates from the way in which force and tether extension are calculated. Typically, both the force and the extension are calculated from the recorded bead-movement. This approach assumes that the two traps are aligned in the  $z$ -direction so that the pulling geometry does not affect the results.

It is relatively easy to obtain a sufficiently accurate alignment of the two traps in the  $xy$ -plane by visual manipulation of the trap positions. However, obtaining an accurate alignment of the traps in the direction of the propagating beams ( $z$ -direction) is more difficult and in reality there is always some degree of  $z$ -level misalignment.

To evaluate the consequences of  $z$ -level misalignment the problem was simulated. Four experiments were simulated in which a tether similar to that shown in Figure 57 was “stretched” while gradually increasing  $z$ -level misalignment. The simulation was performed assuming a dual trap setup with spring constants of 100 pN/ $\mu$ m for both traps in the  $y$ -direction and 30 pN/ $\mu$ m and 50 pN/ $\mu$ m in the  $z$ -direction<sup>10</sup>. The result is shown in Figure 59, where “apparent” force and extension describe the force and extension we would detect in the  $y$ -direction.

The appearance of the force-extension curves is quite intuitive: when the beads are misaligned in the  $z$ -direction and we record force and extension in the  $y$ -direction, we underestimate both the extension and the force. As we percentage-wise underestimate the extension more than the force, we record an artificially high force for the false short extension.

From figures 59B and 59C we also see that the angle increases as the force is increased. This makes intuitive sense as the  $z$ -level misalignment is reduced as the beads move away from the beam focus when  $F_z$  increases.

It follows from Figure 59A that for relative short tethers attention must be paid to  $z$ -level misalignment in order to estimate the polymer parameters correctly. For a  $z$ -level misalignment of 400 nm the persistence length decreases from 50 nm to 25 nm and the stretch modulus changes from 1000 pN to 484 pN (obtain by fitting the EWLC to the simulated data, not shown). It should be mentioned that this effect decreases as the length of the tether increases.

From the data used to construct Figure 59 we can also estimate the error in apparent force and apparent extension. From the simulations we know the force ( $F_{yz}$ ) and we know the full extension of the tether. By comparing these values to those of  $F_y$  and  $bb_y$  we can estimate the error caused by  $z$ -level misalignment. This is shown in Figure 60.

From Figure 60 we see that the error for force is small (<1%) even when we apply low force. Likewise, if we apply more than 5 pN the error in extension is less than 8%. Thus, even for relative large values of  $z$ -level misalignment we have good estimates for force and extension (if the traps are sufficiently weak in the  $z$ -direction).

<sup>10</sup>These are representative of those obtained from the JPK NanoTracker™ during our experiments.

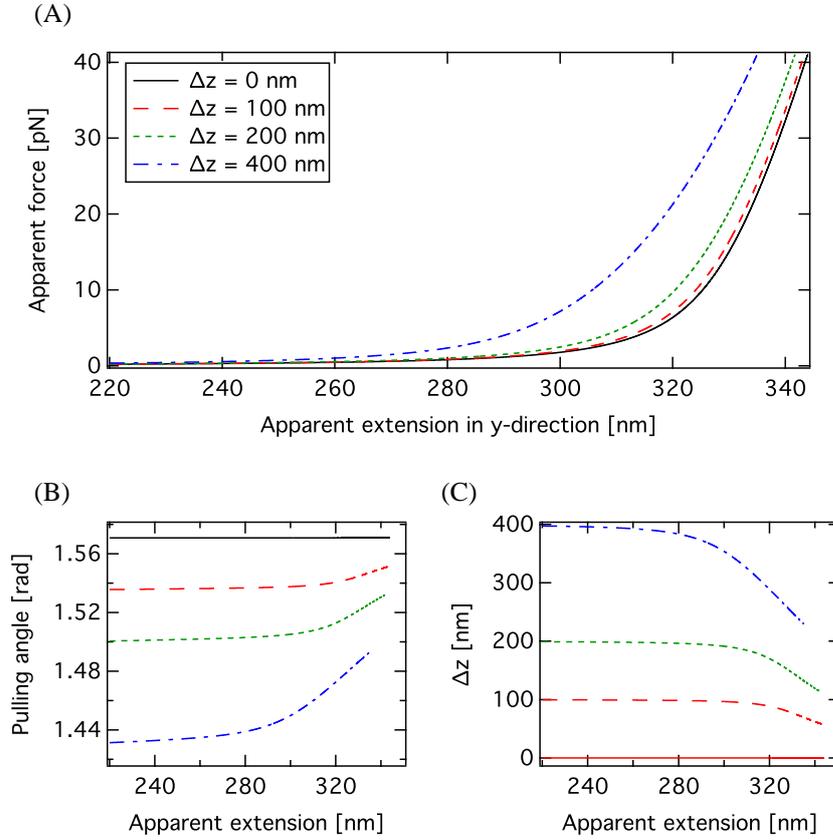


Figure 59: Effect of  $z$ -level misalignment on force-extension curves. Four experiments were simulated in which the misalignment was increased from 0 nm to 400 nm. **A**: Apparent force versus apparent extension. The term “apparent” refers to the fact that they do not reflect the true force and extension, but merely the force and extension detected in the pulling direction. **B**: Pulling angle versus apparent extension in  $y$ -direction. For any  $z$ -level misalignment the pulling angle increases as force is applied. Solid black line indicate  $\pi/2$ . **C**:  $\Delta z$  versus apparent extension in  $y$ -direction. For any  $z$ -level misalignment the misalignment decreases as force is applied. See Figure 56 for details about force and pulling angle. Simulated setup: Dual trap with spring constants in the  $z$ -direction of 30 pN/ $\mu\text{m}$  and 50 pN/ $\mu\text{m}$ , respectively, and 100 pN/ $\mu\text{m}$  in the  $y$ -direction.

### 13.4 Estimation of Unfolding Length

In single molecule experiments like those conducted in Part III, we need to estimate the expected unfolding length of the pseudoknots. Unfolding length refers to the change in extension during the “rip” which we interpret as unfolding of a RNA structure. What happens when the structure unfolds? Generally, the contour length of the tether is increased as single stranded nucleotides are released from the structure. When the contour length is increased abruptly, the length and force are no longer in balance. Consequently, the beads start to move towards the center of their traps resulting in an extension of the tether and

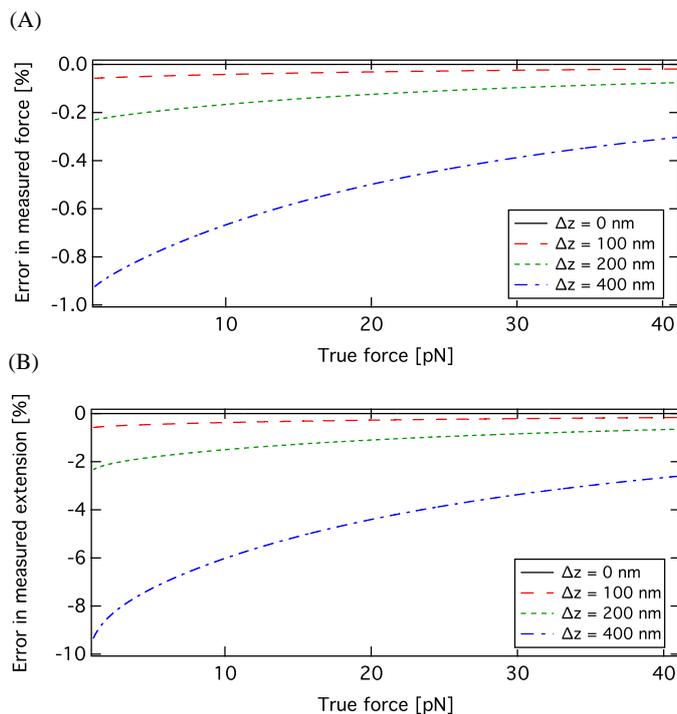


Figure 60: Error in apparent force and apparent extension. **A**: Error in estimated force for three  $z$ -level misalignments. **B**: Error in estimated extension for three different  $z$ -level misalignments. The error represents the error caused by detection of force and extension in the pulling direction. Simulated setup: Simulated setup: Dual trap with spring constants in the  $z$ -direction of  $30 \text{ pN}/\mu\text{m}$  and  $50 \text{ pN}/\mu\text{m}$  and in the pulling direction of  $100 \text{ pN}/\mu\text{m}$ .

a reduction of the force. The movement stops when the length of the tether at the applied force equals that of the bead-bead surface distance.

Estimating the unfolding length as the difference in theoretical extension between the two states at the unfolding force is erroneous, as this does not take into account the reduction in force during unfolding. However, these unfolding events can be simulated by changing the parameters of the tether at the unfolding force. Specifically, to simulate unfolding of our pseudoknots we change to composition of the tether at the force of unfolding. We substitute the two double stranded stems (stem1 and stem2) with an appropriate number of single stranded nucleotides. To compensate for the increase in length, the simulation will increase the bead-bead distance until the force and extension is once again in balance. Figure 61 shows the result of a simulation where unfolding of pseudoknot 22/6b (Figure 37, page 89) takes place at  $15 \text{ pN}$  resulting in a unfolding length of  $28.7 \text{ nm}$ .

#### 13.4.1 Effect of trap-stiffness on unfolding length

It follows from the argumentation in section 12, that immediately after the structure unfolds, the beads start to move towards the center of their traps, creating the distinctive

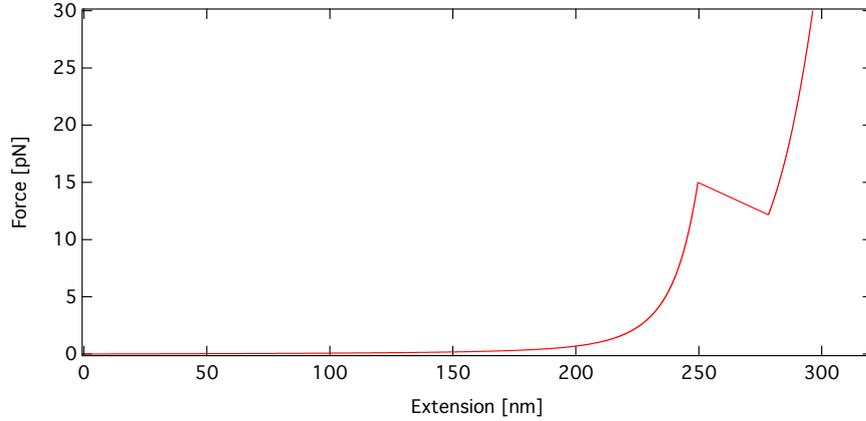


Figure 61: Simulated force-extension curve with unfolding of pseudoknot 22/6b at 15 pN. The unfolding “rip” of 28.7 nm is clearly visible. Simulated setup: single trap with stiffness 100 pN/ $\mu\text{m}$ .

unfolding “rip” as the extension is increased. As the beads move, the force exerted on the tether decreases. If the traps are stiff the force reduces rapidly, and if they are weak the force reduces slowly. The slope of the rip for single trap optical tweezer should theoretically equal the trap stiffness - if the bead moves by 1 nm towards the center of the trap, the force drops correspondingly in accordance with Hooke’s law [Tinoco et al., 2006, Li et al., 2006b]. For dual trap optical tweezer the picture is slightly more complicated where the slope ( $\alpha_F$ ) theoretically should equal:

$$\alpha_F = \frac{\Delta F}{\Delta x} = \frac{\Delta x_2 \kappa_{y2}}{\Delta x}$$

From (36) we rewrite:

$$\alpha_F = \frac{\Delta x_2 \kappa_{y2}}{\Delta x_2 \left( \frac{\kappa_{y1}}{\kappa_{y1} + \kappa_{y2}} \right)^{-1}} = \kappa_2 \left( \frac{\kappa_2}{\kappa_1} + 1 \right)^{-1}$$

If the slope depends on the stiffness of the traps so does the unfolding length - stiff traps will give rise to shorter unfolding lengths than weak traps. This is illustrated in Figure 62.

It is clear from Figure 62 that the trap stiffness has an impact on the unfolding length. The figure also illustrates why this is the case: the larger the trap stiffness the “faster” the beads follows the force-extension trace of a tether with the unfolded structure. It should be mentioned that the practical implications of this phenomenon is not as great as illustrated above, as the trap stiffnesses in literature (as judged by the slope of unfolding rips) typically is between 50 pN/ $\mu\text{m}$  and 200 pN/ $\mu\text{m}$ .

To verify that the simulation is able to capture the physics of unfolding events observed in practice, results from the literature were investigated. In the literature, however, parameters like trap stiffness often are not listed. Thus, for validation one must estimate

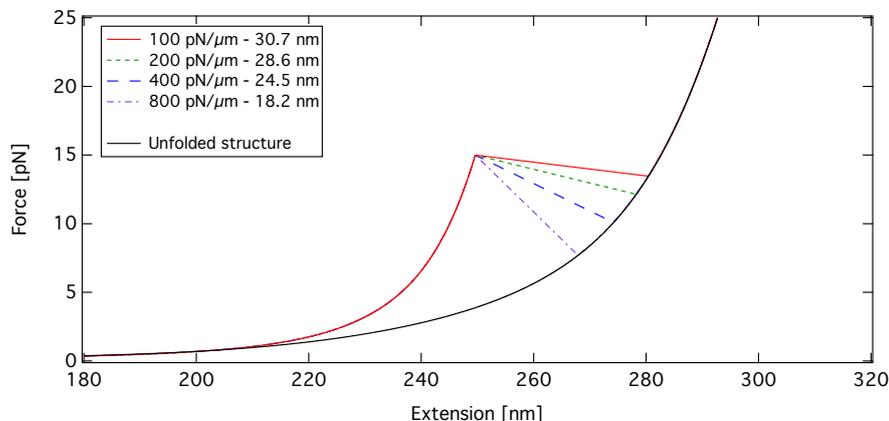


Figure 62: Simulated unfolding length of pseudoknot 22/6b in a dual trap optical tweezers with traps of identical stiffness. The trap stiffness in each simulation is listed in the legend along with the unfolding length. The black curve is a force-extension trace of a tether containing the unfolded structure.

this parameter. In a study by Green et al. mechanical unfoldings of both hairpins and pseudoknots were conducted and in a study by Chen et al. various pseudoknots were unfolded [Green et al., 2008, Chen et al., 2009]. A comparison between the unfolding length observed by Green et al. and Chen et al. and the simulated unfolding lengths are listed in table 24.

	Estimated	Observed values		Simulated
	$\kappa$	$F_u$	$l_u$	$l_u$
pIBV <sup>a</sup>	110 pN/ $\mu$ m	26.0 pN	16 $\pm$ 1.5 nm	15.5 nm
hpS12 <sup>a</sup>	50 pN/ $\mu$ m	18.8 pN	15 $\pm$ 1.1 nm	14.8 nm
hpS1 <sup>a</sup>	50 pN/ $\mu$ m	18.2 pN	11 $\pm$ 1.1 nm	11.2 nm
$\Delta$ U177 <sup>b</sup>	100 pN/ $\mu$ m	50 pN	19 nm	19.6 nm

Table 24: Experimental and theoretical unfolding lengths.  $F_u$ : Unfolding force,  $l_u$ : unfolding length,  $\kappa$ : spring constant. <sup>a</sup>: Observed values are from Green et al. [Green et al., 2008]. Spring constants were estimated from their figure 3a, figure 4c and figure 4a. Simulated values are results from simulations assuming that the tether consisted of 1200 bp RNA/DNA handles, a pseudoknot or a hairpin, and 5 nt ssRNA outside the structure. <sup>b</sup>: Observed value is from Chen et al. [Chen et al., 2009]. Spring constant estimated from their figure 2A. Parameters for RNA/DNA handles was  $L_p = 45$  nm,  $L_c = 0.28$  nm/bp,  $K = 1$  nN. Parameters for ssRNA was  $L_p = 1$  nm,  $L_c = 0.59$  nm/nt,  $K = 800$  pN.

The results in table 24 indicate that the simulation can be used to accurately model unfolding events of both pseudoknots and hairpins over a wide range of forces. It also indicates that the listed parameters for persistence length, contour length and stretch

modulus are sufficiently accurate to model unfolding of our pseudoknots.

### 13.4.2 Impact of $z$ -level misalignment

We also investigated how misalignment in the  $z$ -level would affect the observed unfolding length. Unfolding of pseudoknot 22/6b was simulated for no  $z$ -misalignment and for  $z$ -misalignments of 100 nm, 200 nm and 400 nm. The result is shown in Figure 63 from which it is clear that misalignments of 400 nm and less will have only a minor impact on the observed unfolding lengths ( $< 3\%$ ).

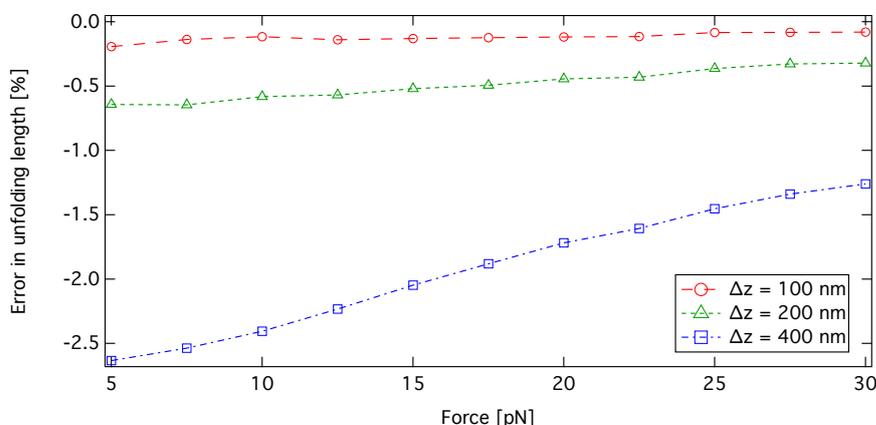


Figure 63: Influence of  $z$ -level misalignment on apparent unfolding length. Unfolding of pseudoknot 22/6b was simulated for no misalignment and for misalignments of 100 nm, 200 nm and 400 nm. The error in unfolding length at the indicated force was calculated as the difference between the observed value at zero misalignment and misalignment of 100 nm (red, circles, dashed), between zero misalignment and 200 nm (green, triangles, dotted) and between zero misalignment and misalignment of 400 nm (blue, squares, dash-dot).

### 13.4.3 Impact of parameter variation

Simulations were performed to estimate how the predicted unfolding length would be affected if the polymer parameters used in the simulations were wrong. Unfolding of pseudoknot 22/6b was simulated using a set of parameters to construct a reference dataset. This dataset was compared to simulations where the parameters were altered. The result is shown in Figure 64.

From Figure 64 it is clear that the persistence length of ssRNA is the only parameter which has a real influence on the predicted unfolding length (Figure 64D). Judging from the predicted unfolding lengths in table 24, a persistence length of 1 nm seems to describe experimental unfolding quite well.

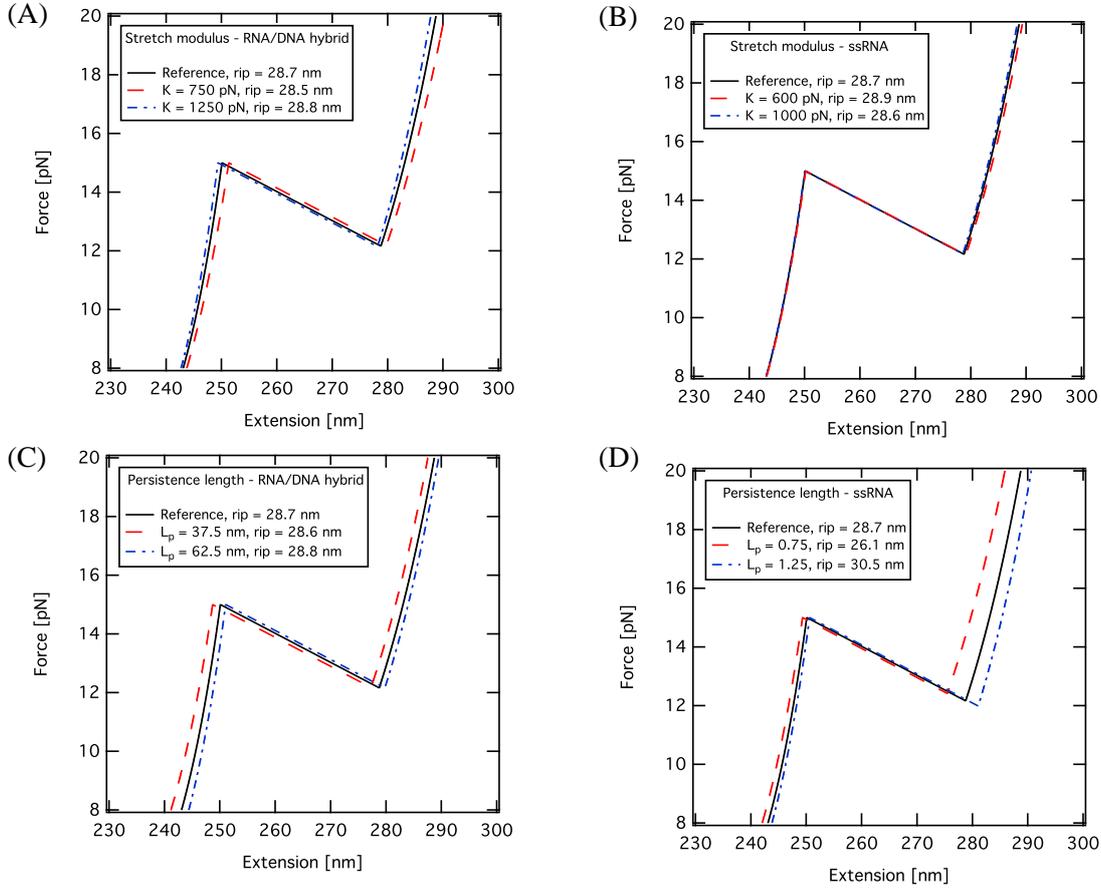


Figure 64: Impact of parameter variation on predicted unfolding length. All simulations were performed assuming that pseudoknot 22/6b was suspended between two RNA/DNA hybrid handles of a total 842 bp and contained 27 single stranded RNA nucleotides outside the folded structure. The following parameters were used for the reference dataset;  $L_p = 50$  nm for RNA/DNA,  $L_p = 1$  nm for ssRNA,  $L_c = 0.28$  nm/bp for RNA/DNA,  $L_c = 0.59$  nm/nt for ssRNA,  $K = 1000$  pN for RNA/DNA and  $K = 800$  pN for ssRNA. Single trap with trap stiffness of  $100$  pN/ $\mu\text{m}$ . **A:** Impact of stretch modulus for RNA/DNA. **B:** Impact of stretch modulus for ssRNA. **C:** Impact of persistence length of RNA/DNA. **D:** Impact of persistence length of ssRNA.

## 14 Discussion - Optical Tweezers Simulations

The described simulation framework seems to be able to accurately model several aspects of the force-ramp experiments conducted in Part III, including structural unfolding,  $z$ -level misalignment and variation in trap stiffness.

The unfolding lengths predicted for several different structures at a wide range of unfolding forces is in excellent agreement with experimentally observed values (see Table 24). This enables us to use the framework to predict unfolding lengths for our pseudoknots

and potential folding intermediates in Part III.

Simulations showed that although  $z$ -level misalignment has a severe impact on the EWLC fitting parameters, the error in estimated force, extension and unfolding length was kept at a acceptable level for misalignments smaller than 400 nm. Experimentally, it should be possible to maintain the  $z$ -level misalignment at a lower level. The persistence length of ssRNA has a great influence on the predicted unfolding length, but it seems that a value of 1 nm gives good results.

The simulations conducted here show that we can accurately predict the unfolding length of both pseudoknots and hairpins, and that both force and unfolding length is relatively unaffected by  $z$ -level misalignments.

Last but not least it should be mentioned, that the simulations rely on accurate knowledge of polymer parameters ( $L_p$ ,  $K$  and contour length per monomer). These parameters are not easy to obtain and vary with experimental conditions. The results obtained by these simulations should therefore only be considered as a estimations.



## Part V

# Final Conclusions and Perspectives

The objective of this thesis was to investigate the functional and physical properties of messenger RNA pseudoknots in relation to -1 PRF. This task was undertaken by approaching the problem from several directions. First, a set of completely artificial pseudoknots were constructed *in silico* and their ability to induce -1 frameshift was evaluated *in vivo* (Part I). Second, attention was turned to the putative ribosomal helicase which could play a role in -1 PRF either before or after the frameshifting event - this element in the frameshift process has received fairly limited attention (Part II). Third, single molecule force spectroscopy was employed to investigate the mechanical properties of the pseudoknots used for *in vivo* frameshift assays (Part III). Finally, a simulation tool was constructed to aid in the interpretation of our force spectroscopy data and to help quantify potential uncertainties (Part IV).

We found that large pseudoknots with a large predicted stability were able to act as roadblocks for translating ribosomes resulting in aberrant translation. This effect was suggested by Chen et al. in 2009, based on single molecule force spectroscopy [Chen et al., 2009]:

*“It is likely that ribosomes would be stalled by a pseudoknot with an unfolding force of 60 pN or higher, resulting in abortive translation.”*

Although 2D SDS-PAGE slightly underestimates the frameshift efficiency due to difficulties in quantifying all the full length frameshift product [Tholstrup et al., 2012], the frameshift efficiency of our pseudoknots were quite low compared to more naturally occurring pseudoknots, even when we estimated the frameshift efficiency from 1D SDS-PAGE (not shown). The reason for the low frameshift efficiency of our pseudoknots is unknown, but the length of loop1 could play a role. Typically, the length of loop1 is around 1-3 nt and interactions between loop1 and stem2 appear to be substantial, e.g. in the BWYV pseudoknot, a loop1 base (C8) forms quadruple interactions with bases in stem2 [Thiel et al., 2003, Su et al., 1999, Naphthine et al., 1999]. It is possible that our 6 nt loop1 is either too long, resulting in a more “loose” structure, or that it fails to make stabilizing interactions with stem2. It would be interesting to investigate the effect of decreasing the length loop1 into something more natural or for changes the loop2 composition.

If the putative ribosomal helicase is involved in -1 PRF it is sufficiently robust to handle the mutations constructed here, as no effect on -1 PRF was observed. Although the mutations did not effect -1 PRF, the helicase remains an interesting target for future studies, as this component of -1 PRF has received little attention. The experiments conducted here show that it is possible to construct these mutations *in vivo* and assert their effect on -1 PRF. A possible next step is to construct double or triple mutations like those used *in vitro* by Takyar et al. or to attempt to make the mutations in *rpsC* used by Takyar

---

et al. [Takyar et al., 2005].

In our single molecule experiments, we were able to assign structural changes to all of the observed one-step unfolding events and for most groups of two-step unfolding events. This fulfills our initial objective of creating completely synthetic pseudoknots - at least with a high degree. All of the pseudoknots from which we obtained unfolding information, were able to fold as expected with the exception of the “inverted” pseudoknot 6/11 investigated by Rebecca Bolt Ettliger [Ettliger, 2012]. If nothing else, this is a testament to the ability of open access tools to accurately predict complex RNA structures.

Unfortunately, we were unable to mechanically unfold two of our pseudoknot 22/6a and 16/6. We are currently trying to repeat the unfolding of pseudoknot 22/6a and 22/6b in the absence of  $Mg^{2+}$  ions; hopefully this lowers the stability sufficiently to allow unfolding. Even though we were unable to show the existence of a structure in the pseudoknot 22/6a construct, we have ample indirect evidence of its existence: GenScript which created the structures was unable to sequence pseudoknot 22/6a construct, Eurofins MWG Operon, whom we used for routine sequencing, had difficulties in sequencing only the pseudoknot 22/6a construct and were able to do so only when using “special chemistry” and optimizing for GC rich sequences, and we were unable to perform mutational PCR with primers annealing to regions designed to be inside the folded structure (not shown). Taken together it is likely that a structure does exist, and in the light of our findings for pseudoknot 22/6b and 22/6c, it also seems likely that the actual structure is identical to the predicted structure.

Although we showed that a construct containing pseudoknot 16/6 and 22/6a is able to stall translating ribosomes, we did not formally show that the folded pseudoknot was responsible. Thus, we cannot exclude the possibility that the folding intermediate hairpin, whose existence was observed for pseudoknot 22/6b and 22/6c, is able to stall ribosomes as well. We did try to construct the 22 bp folding intermediate hairpin of pseudoknot 22/6a, but our inability to anneal primers inside the structure made this impossible. *In vivo* experiments have shown that hairpins of similar predicted stability constitute no obstacle to translating ribosomes [Sørensen et al., 1989]. Nonetheless, a sequence encoding the 22 bp folding intermediate hairpin of pseudoknot 22/6a should be constructed to confirm the inability of this structure to stall ribosomes.

The implications of stalled ribosomes for the results of frameshift assays are difficult to assess. However, if the ribosome continues sufficiently far into the structure, leaving an empty slippery sequence and stop codon, the impact on the frameshift assay measurements could be dramatic. If subsequent ribosomes on the same mRNA do not meet the pseudoknot, they will not change reading frame and consequently terminate translation at the in-frame stop-codon. One could say that an mRNA with such a stalled ribosome is committed to producing in-frame stop products for the rest of its lifetime. This will lead to a dramatic underestimation of the frameshift efficiency, even if all peptide species are taken into account. A quick’n’dirty Monte-Carlo simulation indicates that a pseudoknot with a true frameshift efficiency of 20% will yield a have an apparent frameshift efficiency

of  $\approx 13\%$  in a frameshift assay even if its probability of stalling ribosomes is as low as  $10\%$ <sup>11</sup>.

Interestingly, the pseudoknot 6/11 (investigated by Rebecca Bolt Ettliger) did not fold as expected based on unfolding data from force spectroscopy despite the fact that it had a slightly larger predicted stability than that of pseudoknot 11/6. This was the only investigated structure which did not form the expected structure. The reason is unknown, but it is interesting that no naturally occurring pseudoknots have a stem2 that is significantly longer than stem1 - at least to my knowledge. One could speculate that this topology is incompatible with the tertiary structure of a pseudoknot. In general it appears that as soon as we move too far away from natural pseudoknots, either by making them too strong or by inverting their stems, we observe something unexpected. The strong pseudoknot were able to stall translating ribosomes and the inverted pseudoknot failed to fold as expected. This, to some extent supports, the notion that naturally occurring pseudoknot have evolved to induce the specific frameshift efficiencies needed for viral proliferation [Plant et al., 2010]. It would be interesting to make construct additional pseudoknots which gradually transform pseudoknot 6/11 into something resembling naturally occurring pseudoknots to investigate when the structure is able to form.

It is also interesting that we observed such a complex pattern unfolding and refolding on single RNA molecules, where formation of different structures form interchangeably on the same strand of RNA. For example, we observed that unfolding alternated between the 10hp hairpin, pseudoknot 10/6U and pseudoknot 10/6 on the same tether in successive extension/retraction cycles. This observation most likely reflect a true feature of the dynamic nature of structure formation in solution, where the structural population probably consists of several different species. Using a force-feedback system, it would be interesting to investigate these species in greater detail by looking at alternating states occurring at constant force [Woodside et al., 2006]. Investigating the refolding and unfolding dynamics of RNA in greater detail could revile important aspects of -1 PRF. It would be highly relevant for the understanding of these pathways to reconstruct the entire energy landscape from repeated unfolding experiments although it would require a lot of data [Hummer and Szabo, 2001].

The JPK NanoTracker™ suffered from a significant noise problem which meant that we had to spend valuable time constructing and validating an alternative method of calibration. Apart from this obvious Achilles' heel, the JPK NanoTracker™ was easy to use, and it has great potential as a successful commercial optical tweezers once these vibrations have been eliminated. This could possibly be achieved by altering the response time of the galvanometric mirrors (personal communication from JPK) or by using different components such as Acousto-Optic Deflectors in stead of galvanometric mirrors.

In future unfolding experiments it would be interesting to make extensions in the ssDNA handles such that only stem1 or stem2 is able to form. Such extensions would be easy to make and have been used elsewhere to achive similar effects [Chen et al., 2009]. This could further support the notion that our pseudoknots are able to fold as expected.

---

<sup>11</sup>For a simulation of 10,000 mRNAs with a uniform probability of being translated between 10 and 60 times

It would also be worth while to increase the length of ssDNA/RNA construct in future experiments in an attempt to reduce the potential trap-trap interaction indicated by Figure 22 (page 69). We used ssDNA linkers of approximately 430 nt while others have used ssDNA linkers of approximately 600 nt for similar experiments [Green et al., 2008, Chen et al., 2007].

It would be very interesting to follow a translating ribosome as it translates a pseudoknot containing RNA inside an optical tweezers setup as Wen et al. did for a hairpin structure [Wen et al., 2008]. This would, however, require a reduced noise signal compared to what we are able to obtain with the JPK NanoTracker™ today.

Another interesting aspect of -1 PRF, which has not been discussed here, is the impact of ribosome spacing. Directly related to the refolding kinetics of the pseudoknot, one might expect that the probability of pseudoknot-refolding between translating ribosomes is greater if the ribosomes are spaced far apart, resulting in a higher frameshift efficiency. Lopinski et al. showed that altered initiation frequencies *in vivo* resulted in altered levels of -1 PRF [Lopinski et al., 2000]. Although not necessarily cause by the same underlying mechanism, we observed alterations in frameshift efficiency when the upstream gene10 ORF was substituted by another ORF (not shown). These observations may imply that individual pseudoknots have evolved to function in specific genomic contexts, which makes this effect highly relevant for the understanding of -1 PRF.

Finally, it is worth mentioning that the interplay between theoretical models and practical experiments, like the experiments presented in this thesis, holds a great potential for the future development of RNA folding algorithms such as the *Vfold* model [Cao et al., 2010].

Copenhagen, Denmark, July 2, 2012

---

Jesper Tholstrup Hansen

---

## References

- [Adelman et al., 2002] Adelman, K., La Porta, A., Santangelo, T. J., Lis, J. T., Roberts, J. W., and Wang, M. D. (2002). Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *PNAS*, 99(21):13538–13543.
- [Agarwal et al., 2011] Agarwal, D., Gregory, S. T., and O’Connor, M. (2011). Error-prone and error-restrictive mutations affecting ribosomal protein S12. *Journal of molecular biology*, 410(1):1–9.
- [Andersson et al., 1986] Andersson, D. I., Andersson, S. G., and Kurland, C. G. (1986). Functional interactions between mutated forms of ribosomal proteins S4, S5 and S12. *Biochimie*, 68(5):705–713.
- [Ashkin, 1970] Ashkin, A. (1970). Acceleration and Trapping of Particles by Radiation Pressure. *Physical Review Letters*, 24(4):156–159.
- [Ashkin, 1992] Ashkin, A. (1992). ScienceDirect - Biophysical Journal : Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophysical journal*, 61:569–582.
- [Baranov et al., 2005] Baranov, P. V., Henderson, C. M., Anderson, C. B., Gesteland, R. F., Atkins, J. F., and Howard, M. T. (2005). Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*, 332(2):498–510.
- [Bekaert et al., 2010] Bekaert, M., Firth, A. E., Zhang, Y., Gladyshev, V. N., Atkins, J. F., and Baranov, P. V. (2010). Recode-2: new design, new search tools, and many more genes. *Nucleic acids research*, 38(Database issue):D69–74.
- [Belcourt and Farabaugh, 1990] Belcourt, M. F. and Farabaugh, P. J. (1990). Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell*, 62(2):339–352.
- [Ben-Shem et al., 2011] Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science (New York, NY)*, 334(6062):1524–1529.
- [Berg-Sørensen and Flyvbjerg, 2004] Berg-Sørensen, K. and Flyvbjerg, H. (2004). Power spectrum analysis for optical tweezers. *Review of Scientific Instruments*, 75(3):594–612.
- [Björkman et al., 1999] Björkman, J., Samuelsson, P., Andersson, D. I., and Hughes, D. (1999). Novel ribosomal mutations affecting translational accuracy, antibiotic resistance and virulence of *Salmonella typhimurium*. *Molecular microbiology*, 31(1):53–58.
- [Blanchard et al., 2004] Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S., and Puglisi, J. D. (2004). tRNA selection and kinetic proofreading in translation. *Nature Structural & Molecular Biology*, 11(10):1008–1014.

- [Brierley et al., 1987] Brierley, I., Bournsnel, M. E., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D., and Inglis, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *The EMBO journal*, 6(12):3779–3785.
- [Brierley et al., 1992] Brierley, I., Jenner, A. J., and Inglis, S. C. (1992). Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *Journal of molecular biology*, 227(2):463–479.
- [Brierley et al., 1997] Brierley, I., Meredith, M. R., Bloys, A. J., and Hagervall, T. G. (1997). Expression of a coronavirus ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on frameshifting. *Journal of molecular biology*, 270(3):360–373.
- [Brierley et al., 1991] Brierley, I., Rolley, N. J., Jenner, A. J., and Inglis, S. C. (1991). Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *Journal of molecular biology*, 220(4):889–902.
- [Bustamante et al., 1994] Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994). Entropic elasticity of lambda-phage DNA. *Science (New York, NY)*, 265(5178):1599–1600.
- [Cao et al., 2010] Cao, S., Giedroc, D. P., and Chen, S.-J. (2010). Predicting loop–helix tertiary structural contacts in RNA pseudoknots. *RNA (New York, NY)*, 16:538–552.
- [Caruthers and McKay, 2002] Caruthers, J. and McKay, D. (2002). Helicase structure and mechanism. *Current Opinion in Structural Biology*, 12:123–133.
- [Cecconi et al., 2005] Cecconi, C., Shank, E. A., Bustamante, C., and Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. *Science (New York, NY)*, 309(5743):2057–2060.
- [Chemla et al., 2005] Chemla, Y. R., Aathavan, K., Michaelis, J., Grimes, S., Jardine, P. J., Anderson, D. L., and Bustamante, C. (2005). Mechanism of force generation of a viral DNA packaging motor. *Cell*, 122(5):683–692.
- [Chen et al., 2009] Chen, G., Chang, K.-Y., Chou, M.-Y., Bustamante, C., and Tinoco, I. J. (2009). Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *PNAS*, 106(31):12706–12711.
- [Chen et al., 2007] Chen, G., Wen, J., and Tinoco, I. (2007). Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *RNA (New York, NY)*, 13(12):2175–2188.
- [Cherepanov, 1995] Cherepanov, P. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of F1p-catalyzed excision of the antibiotic-resistance determinant. *Gene*, 158(1):9–14.

- 
- [Collin et al., 2005] Collin, D., Ritort, F., Jarzynski, C., Smith, S. B., Tinoco, I., and Bustamante, C. (2005). Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437(7056):231–234.
- [Crooks, 1999] Crooks, G. (1999). Phys. Rev. E 60, 2721 (1999): Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721–2726.
- [Datsenko and Wanner, 2000] Datsenko, K. A. and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS*, 97(12):6640–6645.
- [Davenport et al., 2000] Davenport, R. J., Wuite, G. J., Landick, R., and Bustamante, C. (2000). Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science (New York, NY)*, 287(5462):2497–2500.
- [Dinman and Wickner, 1994] Dinman, J. and Wickner, R. B. (1994). Translational Maintenance of Frame: Mutants of *Saccharomyces cerevisiae* With Altered -1 Ribosomal Frameshifting Efficiencies. *Genetics*, 136:75–86.
- [Dinman and Wickner, 1992] Dinman, J. D. and Wickner, R. B. (1992). Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *Journal of virology*, 66(6):3669–3676.
- [Doty et al., 1959] Doty, P., Boedtke, H., Fresco, J. R., Haselkorn, R., and Litt, M. (1959). Secondary Structure In Ribonucleic Acids. *PNAS*, 45(4):482.
- [Dudko et al., 2006] Dudko, O. K., Hummer, G., and Szabo, A. (2006). Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical Review Letters*, 96(10):108101.
- [Dulude et al., 2006] Dulude, D., Berchiche, Y., Gendron, K., Brakier-Gingras, L., and Heveker, N. (2006). Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. *Virology*, 345(1):127–136.
- [Ettlinger, 2012] Ettlinger, R. B. (2012). Investigating RNA pseudoknots using a double optical trap (Master thesis). University of Copenhagen, Faculty of Science.
- [Fabuss and Korosi, 1969] Fabuss, B. and Korosi, A. (1969). Viscosities of Aqueous Solutions of Several Electrolytes Present in Sea Water. *Journal of Chemical and Engineering Data*, 14:192–197.
- [Frank and Agrawal, 2000] Frank, J. and Agrawal, R. K. (2000). A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature*, 406(6793):318–322.
- [Gabashvili et al., 2000] Gabashvili, I. S., Agrawal, R. K., Spahn, C. M., Grassucci, R. A., Svergun, D. I., Frank, J., and Penczek, P. (2000). Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell*, 100(5):537–549.

## References

---

- [Gong et al., 2006] Gong, Z., Chen, H., Xu, S., Li, Y., and Lou, L. (2006). Monte-Carlo simulation of optical trap stiffness measurement. *Optics Communications*, 263(2):229–234.
- [Gore et al., 2003] Gore, J., Ritort, F., and Bustamante, C. (2003). Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *PNAS*, 100(22):12564–12569.
- [Green et al., 2008] Green, L., Kim, C.-H., Bustamante, C., and Tinoco, I. J. (2008). Characterization of the mechanical unfolding of RNA pseudoknots. *Journal of molecular biology*, 375(2):511–528.
- [Gross et al., 2011] Gross, P., Laurens, N., Oddershede, L. B., Bockelmann, U., Peterman, E. J. G., and Wuite, G. J. L. (2011). Quantifying how DNA stretches, melts and changes twist under tension. *Nature Physics*, 7(9):731–736.
- [Hanahan, 1983] Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology*, 166:557–580.
- [Hansen et al., 2007] Hansen, T. M., R, S. N. S., Oddershede, L. B., and Sørensen, M. A. (2007). Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *PNAS*, 104(14):5830–5835.
- [Hummer and Szabo, 2001] Hummer, G. and Szabo, A. (2001). Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *PNAS*, 98(7):3658–3661.
- [Jacks et al., 1988a] Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E. (1988a). Signals for Ribosomal Frameshifting in the Rous Sarcoma Virus *gag-pol* Region. *Cell*, 44(3):447–458.
- [Jacks et al., 1988b] Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988b). Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature*, 331(6153):280–283.
- [Jacks et al., 1987] Jacks, T., Townsley, K., Varmus, H. E., and Majors, J. (1987). Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. *PNAS*, 84(12):4298–4302.
- [Jacks and Varmus, 1985] Jacks, T. and Varmus, H. E. (1985). Expression of the Rous-Sarcoma Virus *Pol* Gene by Ribosomal Frameshifting. *Science (New York, NY)*, 230(4731):1237–1242.
- [Jarzynski, 1997] Jarzynski, C. (1997). Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*, 78(14):2690–2693.
- [JPK Instruments, 2011] JPK Instruments (2011). NanoTracker - Force-Sensing Optical Tweezers & 3D Partical Tracking Platform. <http://www.jpk.com/nanotracker-tm.387.html>.

- [Kirthi et al., 2006] Kirthi, N., Roy-Chaudhuri, B., Kelley, T., and Culver, G. M. (2006). A novel single amino acid change in small subunit ribosomal protein S5 has profound effects on translational fidelity. *RNA (New York, NY)*, 12(12):2080–2091.
- [Kontos et al., 2001] Kontos, H., Naphthine, S., and Brierley, I. (2001). Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Molecular and cellular biology*, 21(24):8657–8670.
- [Korosi, 1968] Korosi, A. (1968). Viscosities of Binary Aqueous Solutions of Sodium Chloride, Potassium Chloride, Sodium Sulfate, and Magnesium Sulfate at Concentrations and Temperatures of Interest in Desalination Processes. *Journal of Chemical and Engineering data*, 13(4):548–553.
- [Kramer and Farabaugh, 2007] Kramer, E. B. and Farabaugh, P. J. (2007). The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA*, 13(1):87–96.
- [Kurkcuoglu et al., 2008] Kurkcuoglu, O., Doruker, P., Sen, T. Z., Kloczkowski, A., and Jernigan, R. L. (2008). The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. *Physical biology*, 5(4):046005.
- [Landry et al., 2009] Landry, M. P., McCall, P. M., Qi, Z., and Chemla, Y. R. (2009). Characterization of photoactivated singlet oxygen damage in single-molecule optical trap experiments. *Biophysical journal*, 97(8):2128–2136.
- [Larkin et al., 2007] Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T., and Higgins, D. (2007). Clustalw and clustalx version2. *Bioinformatics*, 23:2947–2948.
- [Léger et al., 2007] Léger, M., Dulude, D., Steinberg, S. V., and Brakier-Gingras, L. (2007). The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed -1 ribosomal frameshift. *Nucleic acids research*, 35(16):5581–5592.
- [Léger et al., 2004] Léger, M., Sidani, S., and Brakier-Gingras, L. (2004). A reassessment of the response of the bacterial ribosome to the frameshift stimulatory signal of the human immunodeficiency virus type 1. *RNA (New York, NY)*, 10(8):1225–1235.
- [Li et al., 2006a] Li, P. T. X., Bustamante, C., and Tinoco, I. (2006a). Unusual mechanical stability of a minimal RNA kissing complex. *PNAS*, 103(43):15847–15852.
- [Li et al., 2006b] Li, P. T. X., Collin, D., Smith, S. B., Bustamante, C., and Tinoco, I. (2006b). Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophysical journal*, 90(1):250–260.
- [Lindahl and Zengel, 1986] Lindahl, L. and Zengel, J. M. (1986). Ribosomal genes in *Escherichia coli*. *Annual review of genetics*, 20:297–326.

- [Link et al., 1997] Link, A. J., Phillips, D., and Church, G. M. (1997). Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *Journal of bacteriology*, 179(20):6228–6237.
- [Liphardt et al., 2002] Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I., and Bustamante, C. (2002). Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski’s equality. *Science (New York, NY)*, 296(5574):1832–1835.
- [Liphardt et al., 2001] Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I., and Bustamante, C. (2001). Reversible unfolding of single RNA molecules by mechanical force. *Science (New York, NY)*, 292(5517):733–737.
- [Liu et al., 2009] Liu, B., Shankar, N., and Turner, D. H. (2009). Fluorescence Competition Assay Measurements of Free Energy Changes for RNA Pseudoknots. *Biochemistry*, 49(3):623–634.
- [Lopinski et al., 2000] Lopinski, J. D., Dinman, J. D., and Bruenn, J. A. (2000). Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Molecular and cellular biology*, 20(4):1095–1103.
- [Mathews et al., 1999] Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology*, 288(5):911–940.
- [Melian et al., 2010] Melian, E. B., Hinzman, E., Nagasaki, T., Firth, A. E., Wills, N. M., Nouwens, A. S., Blitvich, B. J., Leung, J., Funk, A., Atkins, J. F., Hall, R., and Khromykh, A. A. (2010). NS1’ of flaviviruses in the Japanese encephalitis virus serogroup is a product of ribosomal frameshifting and plays a role in viral neuroinvasiveness. *Journal of virology*, 84(3):1641–1647.
- [Michiels et al., 2001] Michiels, P. J., Versleijen, A. A., Verlaan, P. W., Pleij, C. W., Hilbers, C. W., and Heus, H. A. (2001). Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *Journal of molecular biology*, 310(5):1109–1123.
- [Milman and Langridge, 1967] Milman, G. and Langridge, R. (1967). The structure of a DNA-RNA hybrid. *PNAS*, 57(6):1804–1810.
- [Moazed and Noller, 1989] Moazed, D. and Noller, H. F. (1989). Intermediate states in the movement of transfer RNA in the ribosome. *Nature*, 342(6246):142–148.
- [Molloy and Padgett, 2002] Molloy, J. and Padgett, M. J. (2002). Lights, action: Optical tweezers - Contemporary Physics. *Contemporary Physics*, 43(4):241–258.
- [Namy et al., 2006] Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. C., and Brierley, I. (2006). A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, 441(7090):244–247.

- [Naphthine et al., 1999] Naphthine, S., Liphardt, J., Bloys, A., Routledge, S., and Brierley, I. (1999). The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *Journal of molecular biology*, 288(3):305–320.
- [Neidhardt et al., 1974] Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974). Culture medium for enterobacteria. *Journal of bacteriology*, 119(3):736–747.
- [Ogle et al., 2002] Ogle, J. M., Murphy, F. V., Tarry, M. J., and Ramakrishnan, V. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, 111(5):721–732.
- [Olsthoorn et al., 2010] Olsthoorn, R. C. L., Reumerman, R., Hilbers, C. W., Pleij, C. W. A., and Heus, H. A. (2010). Functional analysis of the SRV-1 RNA frameshifting pseudoknot. *Nucleic acids research*, 38(21):7665–7672.
- [Onoa et al., 2003] Onoa, B., Dumont, S., Liphardt, J., Smith, S. B., Tinoco, I., and Bustamante, C. (2003). Identifying kinetic barriers to mechanical unfolding of the T. thermophila ribozyme. *Science (New York, NY)*, 299(5614):1892–1895.
- [Plant and Dinman, 2005] Plant, E. P. and Dinman, J. D. (2005). Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic acids research*, 33(6):1825–1833.
- [Plant et al., 2003] Plant, E. P., Jacobs, K. L. M., Harger, J. W., Meskauskas, A., Jacobs, J. L., Baxter, J. L., Petrov, A. N., and Dinman, J. D. (2003). The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA (New York, NY)*, 9(2):168–174.
- [Plant et al., 2010] Plant, E. P., Rakauskaitė, R., Taylor, D. R., and Dinman, J. D. (2010). Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *Journal of virology*, 84(9):4330–4340.
- [Pleij, 1990] Pleij, C. W. (1990). Pseudoknots: a new motif in the RNA game. *Trends in biochemical sciences*, 15(4):143–147.
- [Python Software Foundation, 2012] Python Software Foundation (2012). Python Programming Language-Official Website. <http://www.python.org>.
- [Reeder et al., 2007] Reeder, J., Steffen, P., and Giegerich, R. (2007). pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic acids research*, 35(Web Server issue):W320–4.
- [Rheinberger et al., 1981] Rheinberger, H., Sternbach, H., and Nierhaus, K. (1981). Three tRNA Binding Sites on Escherichia Coli Ribosomes. *PNAS*, 78(9):5310–5314.
- [Rice et al., 1985] Rice, N. R., Stephens, R. M., Burny, A., and Gilden, R. V. (1985). The gag and pol genes of bovine leukemia virus: Nucleotide sequence and analysis. *Virology*, 142(2):357–377.

- [Ritort and Bustamante, 2002] Ritort, F. and Bustamante, C. (2002). A two-state kinetic model for the unfolding of single molecules by mechanical force. *PNAS*, 99(21):13544–13548.
- [Smith et al., 2001] Smith, D. E., Tans, S. J., Smith, S. B., and Grimes, S. (2001). The bacteriophage  $\phi 29$  portal motor can package DNA against a large internal force. *Nature*, 413:748–752.
- [Smith et al., 1996] Smith, S. B., Cui, Y., and Bustamante, C. (1996). Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science (New York, NY)*, 271(5250):795–799.
- [Sørensen et al., 1989] Sørensen, M. A., Kurland, C. G., and Pedersen, S. (1989). Codon usage determines translation rate in *Escherichia coli*. *Journal of molecular biology*, 207(2):365–377.
- [Su et al., 1999] Su, L., Chen, L., Egli, M., Berger, J. M., and Rich, A. (1999). Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nature structural biology*, 6(3):285–292.
- [Svoboda, 1994] Svoboda, K. (1994). Biological Applications of Optical Forces - Annual Review of Biophysics and Biomolecular Structure, 23(1):247. *Annual Review Of Biophysics And Biomolecular Structure*, 23:247–285.
- [Takyar et al., 2005] Takyar, S., Hickerson, R. P., and Noller, H. F. (2005). mRNA helicase activity of the ribosome. *Cell*, 120(1):49–58.
- [ten Dam et al., 1994] ten Dam, E., Brierley, I., Inglis, S., and Pleij, C. (1994). Identification and analysis of the pseudoknot-containing gag-pro ribosomal frameshift signal of simian retrovirus-1. *Nucleic acids research*, 22(12):2304–2310.
- [Thiel et al., 2003] Thiel, V., Ivanov, K. A., Putics, A., Hertzog, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E., and Ziebuhr, J. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *The Journal of general virology*, 84(Pt 9):2305–2315.
- [Tholstrup et al., 2012] Tholstrup, J., Oddershede, L. B., and Sørensen, M. (2012). mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Research*, 40(1):303–313.
- [Tinoco, 2004] Tinoco, I. (2004). Force as a useful variable in reactions: Unfolding RNA. *Annual Review Of Biophysics And Biomolecular Structure*, 33:363–385.
- [Tinoco et al., 2006] Tinoco, I., Li, P. T. X., and Bustamante, C. (2006). Determination of thermodynamics and kinetics of RNA reactions by force. *Quarterly reviews of biophysics*, 39(4):325–360.
- [Tu et al., 1992] Tu, C., Tzeng, T. H., and Bruenn, J. A. (1992). Ribosomal movement impeded at a pseudoknot required for frameshifting. *PNAS*, 89(18):8636–8640.

- [Valentine et al., 2006] Valentine, M. T., Fordyce, P. M., Krzysiak, T. C., Gilbert, S. P., and Block, S. M. (2006). Individual dimers of the mitotic kinesin motor Eg5 step processively and support substantial loads in vitro. *Nature cell biology*, 8(5):470–476.
- [Varani and McClain, 2000] Varani, G. and McClain, W. H. (2000). The G·U wobble base pair. *EMBO reports*, 1(1):18–23.
- [Vincent and Liebman, 1992] Vincent, A. and Liebman, S. (1992). The yeast omnipotent suppressor SUP46 encodes a ribosomal protein which is a functional and structural homolog of the Escherichia coli S4 ram protein. *Genetics*, 132:375–386.
- [Walter et al., 1994] Walter, A. E., Ming, W., and Turner, D. H. (1994). The Stability and Structure of Tandem G<sub>a</sub> Mismatches in Rna Depend on Closing Base-Pairs. *Biochemistry*, 33(37):11349–11354.
- [Wang et al., 1997] Wang, M. D., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1997). Stretching DNA with optical tweezers. *Biophysical journal*, 72(3):1335–1346.
- [Weiss et al., 1989] Weiss, R., Dunn, D., Shuh, M., Atkins, J. F., and Gesteland, R. F. (1989). *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *The New biologist*, 1(2):159–169.
- [Wen et al., 2008] Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S. H., Noller, H. F., Bustamante, C., and Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. *Nature*, 452(7187):598–603.
- [Wen et al., 2007] Wen, J.-D., Manosas, M., Li, P. T. X., Smith, S. B., Bustamante, C., Ritort, F., and Tinoco, I. (2007). Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophysical journal*, 92(9):2996–3009.
- [White et al., 2011] White, K. H., Orzechowski, M., Fourmy, D., and Visscher, K. (2011). Mechanical unfolding of the beet western yellow virus -1 frameshift signal. *Journal of the American Chemical Society*, 133(25):9775–9782.
- [Woodside et al., 2006] Woodside, M. T., Behnke-Parks, W. M., Larizadeh, K., Travers, K., Herschlag, D., and Block, S. M. (2006). Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *PNAS*, 103(16):6190–6195.
- [Wozniak et al., 2009] Wozniak, A., van Mameren, J., and Ragona, S. (2009). Single-molecule force spectroscopy using the NanoTracker optical tweezers platform: from design to application. *Curr Pharm Biotechnol*, 10(5):467–473.
- [Yu et al., 2011] Yu, C. H., Noteborn, M. H., Pleij, C. W. A., and Olsthoorn, R. C. L. (2011). Stem-loop structures can effectively substitute for an RNA pseudoknot in -1 ribosomal frameshifting. *Nucleic acids research*, 39(20):8952–8959.

## References

---

- [Yusupov et al., 2001] Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science (New York, NY)*, 292(5518):883–896.
- [Yusupova et al., 2001] Yusupova, G. Z., Yusupov, M. M., Cate, J. H., and Noller, H. F. (2001). The path of messenger RNA through the ribosome. *Cell*, 106(2):233–241.
- [Zuker, M., 2003] Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. <http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>.

## A Effective Spring Constant and Damping Function

One way to illustrate the setup created by the JPK NanoTracker™ during a single molecule experiment is illustrated in Figure 65. It consists of three springs linked in sequence, the two traps and the ssDNA/RNA hybrid handles (containing the RNA structure). Our loading rate,  $r = v\kappa_e Q(F, \kappa)$ , defined on page 62, is the amount of force per unit of time exerted on the RNA structure.

From the illustration in Figure 65 it seems intuitive that if the spring constant of the handles (force dependent,  $\kappa_h(F)$ ) is very small (a weak spring) and we move trap2 by  $\Delta x_t$ , then the handle will be extended by something close to  $\Delta x_t$  ( $\Delta x_h \approx \Delta x_t$ ), and the force will increase only slightly as  $\Delta x_1 \approx \Delta x_2 \approx 0$ . This is the situation in the beginning of our pulling experiments where the handles act as a weak entropic spring. Likewise, if  $\kappa_h(F)$  is very large, then  $\Delta x_h \approx 0$  for any movement of trap1, and the force on the handles, and consequently on the RNA structure, will increase by  $\kappa_e(\Delta x_1 + \Delta x_2)$  (defined below).

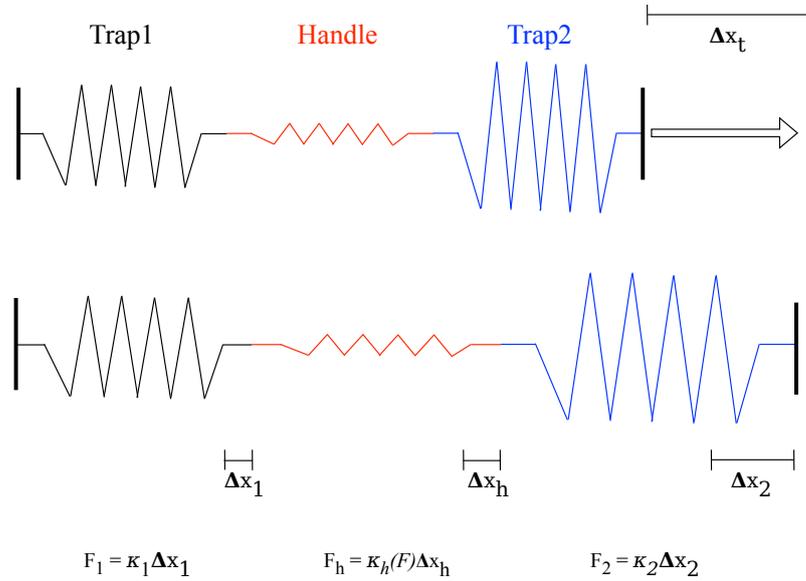


Figure 65: Mechanical illustration of a dual beam Optical Tweezers. We have three linked springs, trap1 (black), the handle (red), and trap2 (blue). If we move trap2 with  $\Delta x_t$  the springs will be elongated with  $\Delta x_1$ ,  $\Delta x_h$ , and  $\Delta x_2$  for trap1, handles, and trap2 respectively. Elongation of Trap1 and trap2 is equivalent to a displacement of the trapped particle away from the beam center.

### A.1 Effective Spring Constant $\kappa_e$

The effective spring constant ( $\kappa_e$ ) is the spring constant of the two traps combined, i.e. the spring constant which fulfills:

$$\Delta F = \kappa_e (\Delta x_1 + \Delta x_2) \quad (47)$$

Thus,  $\kappa_e$  describes how the force changes when we have a combined extension of the springs of both trap1 and trap2. For our force-extension curves this is effectively the slope of the unfolding/refolding events, as  $\kappa_e$  describes how much the force is changed for a given change in total extension (the change in extension is equal to the sum of movement in trap1 and trap2  $[\Delta x]$ ). To find an expression for  $\kappa_e$  we use the balance in force between trap1 and trap2:

$$\kappa_1 \Delta x_1 = \kappa_2 \Delta x_2 \quad (48)$$

We rewrite using  $\Delta x = \Delta x_1 + \Delta x_2$ :

$$\begin{aligned} \kappa_1 (\Delta x - \Delta x_2) &= \kappa_2 \Delta x_2 \\ \Delta x &= \frac{\Delta x_2 (\kappa_2 + \kappa_1)}{\kappa_1} \end{aligned} \quad (49)$$

By insertion into (47):

$$\Delta F = \kappa_e \frac{\Delta x_2 (\kappa_2 + \kappa_1)}{\kappa_1} \Leftrightarrow \kappa_e = \frac{\kappa_1 \kappa_2}{\kappa_1 + \kappa_2} \quad (50)$$

## A.2 Damping Function $Q(F, \kappa)$

During our experiments we increase the force on the tether by moving one trap relative to the other at a defined velocity ( $v$ ). Throughout literature the loading rate ( $r$ ) is defined as a constant for a given velocity:  $r = v\kappa$ , where  $\kappa$  is the spring constant of the trap. This loading rate, however, is not the actual loading rate exerted on the RNA structure, as the ssDNA/RNA handles are elastic with a force dependent spring constant as described above.

The damping function  $Q(F, \kappa)$  depends non-linearly on the force applied as the elasticity of the handles are significant at low force but negligible at high force, at least at forces below 40-50 pN where twisting and force induced melting becomes an issue. The force dependent spring constant of our handles,  $\kappa_h(F)$ , can be determined by numerical differentiation of the EWLC, and consequently depends on the properties of the handles i.e. length and composition. For this example, I have used the handle composition used in our experiments. Once we have  $\kappa_h(F)$  we can, due to the balance of force, write:

$$\Delta x_h \kappa_h(F) = \Delta x_1 \kappa_1 \quad (51)$$

where  $\Delta x_h$  is the change in extension of the handles. We know that if we move e.g. trap1 by  $\Delta x_t$  (illustrated in Figure 65):

$$\Delta x_t = \Delta x_h + \Delta x_1 + \Delta x_2 = \Delta x_h + \Delta x_1 \left( 1 + \frac{\kappa_1}{\kappa_2} \right) \quad (52)$$

where  $\Delta x_2$  is the bead-movement in trap2 and  $\kappa_2$  is the spring constant for trap2. From equation (51) and equation (52) we can write:

$$\begin{aligned}
 \left( \Delta x_t - \Delta x_1 \left( 1 + \frac{\kappa_1}{\kappa_2} \right) \right) \kappa_h(F) &= \Delta x_1 \kappa_1 \\
 \Delta x_t \kappa_h(F) - \Delta x_1 \kappa_h(F) \left( 1 + \frac{\kappa_1}{\kappa_2} \right) &= \Delta x_1 \kappa_1 \\
 \Delta x_t \kappa_h(F) &= \Delta x_1 \left( \kappa_1 + \kappa_h(F) \left( 1 + \frac{\kappa_1}{\kappa_2} \right) \right) \\
 \Delta x_1 &= \frac{\Delta x_t \kappa_h(F)}{\kappa_1 + \kappa_h(F) \left( 1 + \frac{\kappa_1}{\kappa_2} \right)} \tag{53}
 \end{aligned}$$

Our damping function  $Q(F, \kappa)$  is defined as the relative movement in either trap1 or trap2 compared to the movement expect for completely non-elastic handles ( $\kappa_h(F) \rightarrow \infty$ ):

$$\begin{aligned}
 Q(F, \kappa_1) &= \frac{\Delta x_1}{\Delta x_{1, \kappa_h(F) \rightarrow \infty}} \approx \frac{\frac{\Delta x_t \kappa_h(F)}{\kappa_1 + \kappa_h(F) \left( 1 + \frac{\kappa_1}{\kappa_2} \right)}}{\frac{\Delta x_t}{\left( 1 + \frac{\kappa_1}{\kappa_2} \right)}} \\
 Q(F, \kappa_1) &\approx \frac{\kappa_h(F)}{\kappa_1 + \kappa_h(F)} \tag{54}
 \end{aligned}$$

Thus, our damping function reduces to a nice expression which allows us to correct for the flexibility of our handles when calculating the loading rate experienced by the RNA structure during a pulling experiment.

The function  $\kappa_h(F)$  is found by numerical integration of the EWLC and can be described by a Taylor expansion (Figure 66, black solid).

From Figure 66 it is clear that the spring constant of our handles is larger than the typical spring constants of our traps at forces above 5 pN. The insert in Figure 66 shows the damping function  $Q(F, \kappa)$  for the two trap1 spring constants (100 pN/ $\mu\text{m}$  and 200 pN/ $\mu\text{m}$ ). From the estimate of  $Q(F, \kappa_1)$  we see that the actual loading,  $r$ , is 10%-20% smaller than  $r_c$  forces above 10 pN. Fortunately, the asymptotic behavior of  $Q(F, \kappa)$  means that the loading rate  $r$  can be considered constant, even with variation in the unfolding force in experiments with repeated unfolding.

Due to convention, we will consider  $Q(F, \kappa) = 1$  for our kinetic analysis. This is the general approach used throughout literature, and for the spring constants used in our experiments the error introduced is small compared to the variation in spring constants between different experiments [Tinoco, 2004, Hansen et al., 2007, Green et al., 2008, Chen et al., 2007].

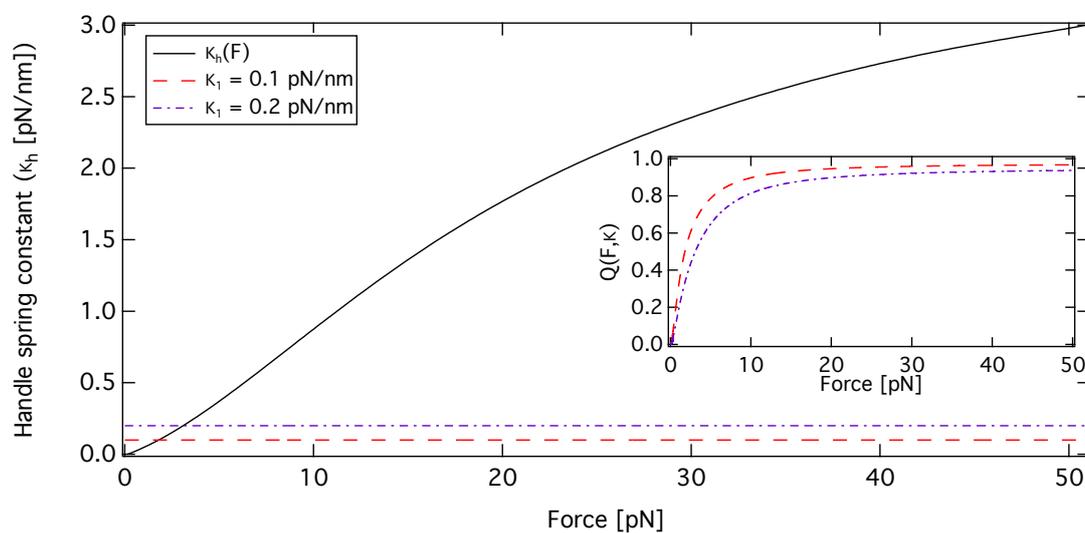


Figure 66: Estimate of  $\kappa_h(F)$  and  $Q(F, \kappa)$ . The estimate of  $\kappa_h(F)$  (black, solid) was obtained by numerical differentiation of the EWLC. The spring constants used in our experiments was in the range of 0.1 (red, long-dash) to 0.2 pN/nm (magenta, dash-dot). **Insert:** Estimate of  $Q(F, \kappa)$  based on  $\kappa_h(F)$  and the indicated spring constants.

## B Data Analysis - iGor Pro

The software required to analyze the output from a force-extension experiment in the JPK NanoTracker™ was written in iGor Pro v.6 and was designed to fulfill three objectives: calculate a force-per-voltage factor ( $\xi$ ) from Stokes calibration, calculate a meter-per-voltage factor ( $\beta$ ) from the PSD (Section 7.3.3), and construct force-extension curves from single molecule experiments.

### B.1 Construction of Force-Extension Curves

The principle behind the computation is outlined in Figure 67 which shows a schematic representation of our single molecule experiments where a single molecule is suspended between two trapped particles.

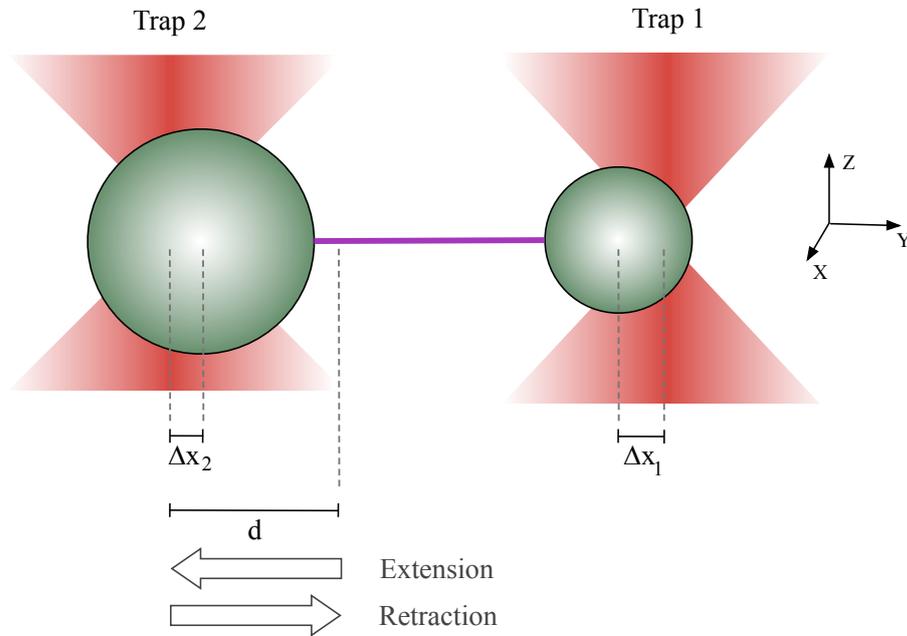


Figure 67: Schematic illustration of experimental setup. A large bead ( $\sim 2 \mu\text{m}$ ) was captured in trap2, a small bead ( $\sim 2 \mu\text{m}$ ) was captured in trap1 and a tether (magenta) was formed between the two beads. During the experiment, trap2 was moved away from trap1 with a velocity of 100 nm/s during the extension, and moved back to its original position with the same velocity during the retraction. At a sampling rate of 10 kHz, the position of trap2 relative to its starting position ( $d$ ), the bead movement in trap1 ( $\Delta x_1$ ), the bead movement in trap2 ( $\Delta x_2$ ) was recorded.

The JPK NanoTracker™ provides the necessary information in the following format where  $d$  is in units of meters and  $V_{QPD,y1}$  and  $V_{QPD,y2}$  are QPD signal for trap1 and trap2 in units of volt for the  $y$ -direction:

$$\begin{array}{ccc}
 d & V_{QPD,y1} & V_{QPD,y2} \\
 d_1 & V_{QPD,y1,1} & V_{QPD,y2,1} \\
 d_2 & V_{QPD,y1,1} & V_{QPD,y2,1} \\
 \vdots & \vdots & \vdots \\
 d_n & V_{QPD,y1,1} & V_{QPD,y2,1}
 \end{array}$$

The distance parameter  $d_n$  describes the distance traveled by trap2 from the initial position to the  $n$ -th recoding. The recorded voltage signal from trap1 and trap2 were smoothed with a boxcar algorithm (bounce method at the ends) with a width of 321 points. The smoothed voltage signal ( $V_{s,QPD,i,n}$ ) can be converted into movement in physical distance through the two conversion factors  $\beta_1$  (trap1) and  $\beta_2$  (trap2):

$$x_{y1,n} [m] = \beta_1 \left[ \frac{m}{V} \right] V_{s,QPD,y1,n} [V]$$

$$x_{y2,n} [m] = \beta_2 \left[ \frac{m}{V} \right] V_{s,QPD,y2,n} [V]$$

From these two quantities and from the value of  $d_n$  it is possible to calculate the *change* in tether-extension ( $e_n$ ):

$$e_n [m] = d_n [m] - (x_{y1,n} [m] + x_{y2,n} [m])$$

It is important to note, that we do not know the exact extension of the tether as we do not know what the trap-trap distance was at the beginning of the extension. As a result of this, our extension on the created force-extension curves ( $e_n$ ) differs from the *true* extension with a constant offset. From the recorded voltage signal from trap1 we can also calculate the force exerted on the tether:

$$F_{1,n} [pN] = \xi_1 \left[ \frac{pN}{V} \right] V_{y1,n} [V]$$

The force-extension curve can now be constructed from  $e$  and  $F$ . The value of  $e$  is not the absolute extension but the change in extension from the initial (unknown) extension of the tether.

## C Viscosity of buffer R

Stokes calibration requires accurate knowledge of the viscosity of the buffer in which the experiment is conducted. Thus, it is relevant to investigate if the viscosity of the experimental buffer is likely to deviate significantly from that of pure water, which is typically used to calculate the drag force. The relative viscosity of the buffer can be estimated from the Othmer rule:

$$\log(\eta_r) = A_0 + B_0 \log(\eta_w) \quad (55)$$

where  $\eta_r$  is the relative viscosity of the solution ( $\eta_r \equiv \eta_{\text{solution}}/\eta_{\text{water}}$ ),  $A_0$  and  $B_0$  are the Othmer constants for the component in solution and,  $\eta_w$  is the dynamic viscosity of pure water (in centipoise<sup>12</sup>). For a ternary solution,  $A_0$  and  $B_0$  can be calculated:

$$A_0 = A_{M,1}I + A_{M,2}I^2 + A_{M,3}I^3 \quad (56)$$

$$B_0 = B_{M,1}I + B_{M,2}I^2 + B_{M,3}I^3 \quad (57)$$

where  $A_{M,i}$  and  $B_{M,i}$  are the sum of the  $i$ 'th viscosity correlation constants for the salts in the buffer:

$$A_{M,i} = A_{i,1} \frac{I_1}{I_T} + A_{i,2} \frac{I_2}{I_T} \quad (58)$$

$$B_{M,i} = B_{i,1} \frac{I_1}{I_T} + B_{i,2} \frac{I_2}{I_T} \quad (59)$$

where  $A_{i,1}$  and  $B_{i,1}$  is the  $i$ 'th correlation constant for component 1 and  $A_{i,2}$  and  $B_{i,2}$  is the  $i$ 'th correlation constant for component 2 respectively.  $I_1$  and  $I_2$  are the ionic strength of component 1 and 2 respectively.  $I_T$  is the total ionic strength of the solution.

The buffer used for the single-molecule experiments in this study, buffer R, contains 250mM NaCl, 10mM MgCl<sub>2</sub> and 10mM Tris-HCl. As no Othmer constants are available for Tris-HCl we will consider only NaCl and the MgCl<sub>2</sub>. The viscosity correlation constants for these salts are listed in table 25 [Korosi, 1968].

Salt	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>
NaCl	0.03550	0.00231	-0.00003	-0.04753	0.01598	-0.00194
MgCl <sub>2</sub>	0.05508	-0.00045	0.00050	-0.00682	0.01614	-0.00217

Table 25: Experimentally determined viscosity correlation constants for salts in buffer R. [Korosi, 1968].

In the following calculations it will be assumed that the density water is 1 kg/L, that the dynamic viscosity of water ( $\eta_w$ ) is 0.891 centipoise and, that MgCl<sub>2</sub> undergoes complete

<sup>12</sup>10 poise =  $kg \cdot m^{-1} \cdot s^{-1}$

dissociation. Furthermore, it is assumed that viscosity correlation constants are additive, which seems to be valid for ionic strengths below 0.71 mol/kg [Fabuss and Korosi, 1969].

The total ionic strength of buffer R can be calculated (omitting Tris-HCl):

$$I_T = \frac{1}{2} \sum_{i=1}^n m_i z_i^2 = \frac{1}{2} (0.250 \cdot 1^2 + 0.250 \cdot (-1)^2 + 0.010 \cdot 2^2 + 0.020 \cdot (-1)^2) = 0.28 \frac{\text{mol}}{\text{kg}}$$

The mixed viscosity correlation constants for the solution can be calculated from equation (58) and equation (59):

$$A_{M,1} = 0.03550 \frac{0.250}{0.28} + 0.05508 \frac{0.030}{0.28} = 0.03760$$

$$A_{M,2} = 0.00231 \frac{0.250}{0.28} - 0.00045 \frac{0.030}{0.28} = 0.002014$$

$$A_{M,3} = -0.00003 \frac{0.250}{0.28} + 0.00050 \frac{0.030}{0.28} = 0.000027$$

$$B_{M,1} = -0.04753 \frac{0.250}{0.28} - 0.00682 \frac{0.030}{0.28} = -0.043168$$

$$B_{M,2} = 0.01598 \frac{0.250}{0.28} - 0.01614 \frac{0.030}{0.28} = 0.015997$$

$$B_{M,3} = -0.00194 \frac{0.250}{0.28} - 0.00217 \frac{0.030}{0.28} = -0.001965$$

From these constants it is possible to calculate the Othmer constants for the solution using equation (56) and equation (57)

$$A_0 = 0.03760 \cdot 0.28 + 0.002014 \cdot 0.28^2 + 0.000027 \cdot 0.28^3 = 0.010686$$

$$B_0 = -0.043168 \cdot 0.28 + 0.015997 \cdot 0.28^2 + -0.001965 \cdot 0.28^3 = -0.010876$$

Finally, from equation (55) it is possible to estimate the relative dynamic viscosity of this solution

$$\log(\eta_r) = A_0 + B_0 \log(\eta_w) \Leftrightarrow \eta_r = 10^{0.010686 - 0.010876 \log(0.891)} = 1.0262$$

Thus, the presence 250mM NaCl and 10mM MgCl<sub>2</sub> in buffer R will result in an increase in the dynamic viscosity of less than 3% compared to pure water. Although tris-HCl has been neglected from these calculations, it seems reasonable to use the dynamic viscosity of water for the Stokes calibrations in this study.

## D Crooks Fluctuation Theorem - Implementation

This appendix contains a short illustration of how Crook Fluctuation Theorem was applied to our data and calculates for the thermodynamic investigation presented in Section 9.

Figure 68 illustrates a force-extension curve, containing both an unfolding and a refolding event.

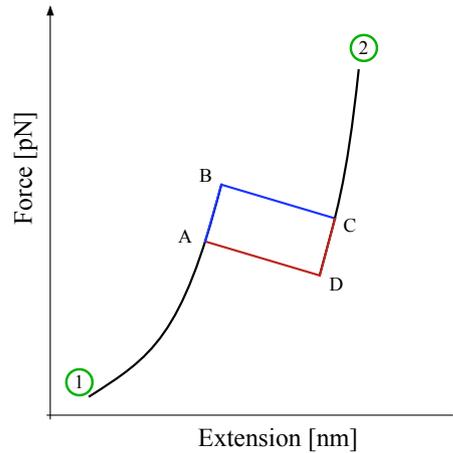


Figure 68: Unfolding and refolding explained

Assuming that we start in point 1 with a tether containing a folded pseudoknot, the flow is as follows: we move trap2 relative to trap1, thus increase the extension and force until point B is reached. At point B the structure unfolds ending in point C where unfolding is complete. The extension is continued to some point, 2. The retraction starts in point 2 and continues to decrease the extension and force until point D is reached. In point D the structure starts to refold which ends in point A where refolding is complete. After point A the retraction curve is once again identical to the extension curve as the tether has regained its original configuration.

When using the CFT we calculate the work between point A and C for the extension and retraction curves as described by Collin et al. [Collin et al., 2005]. Thus, we get two distributions; one for the work used during the extension and one for the work returned during the retraction. The intersection between these two distributions provides the Gibbs free energy difference between point A and C. When we subtract the work required to stretch the DNA/RNA handles and ssRNA outside the structure from point A to C and the work required to stretch the ssRNA of the unfolded structure for 0 pN to the force in point C we get the free energy of the structure.

It should be mentioned that a small amount of work ( $\sim 0.4$  kcal/mol) is also used to stretch the folded structure from point A to B, however, as the fate of this work is elusive when the structure unfolds it has not been included in the correction.

## D.1 Integrating Force-extension Curves

Due to the noise present in the experimental force-extension curves, “integration” of curves with little dissipated work was done using simple geometric shapes as illustrated in Figure 69A. For curves with large dissipated work the integration was performed using the trapezoidal rule on a smoothed version of the curve, as illustrated in Figure 69B.

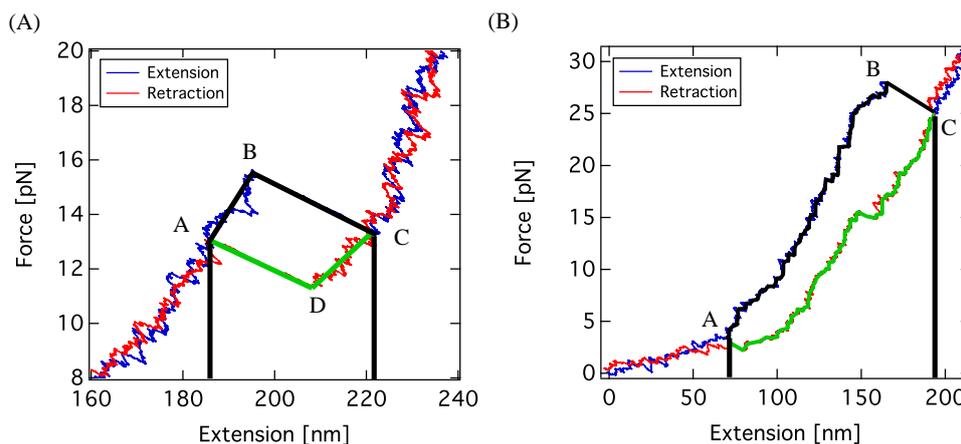


Figure 69: Work calculation on force-extension curves. **A**: For small amounts of dissipated work the work was calculated using simple geometric shapes as indicated by black (unfolding work) and green lines (refolding work). **B**: For curves with large dissipated work the work was calculated using numerical integration as indicated by black and green lines.

## D.2 Calculating the Correction Term

The correction term was calculated using numerical integration of theoretical force extension curves (using the EWLC) for stretching of the components of the RNA/DNA construct. For the curve in Figure 69A the correction term for handles and ssRNA outside the structure was calculated as the work required for stretching the components from the force in point A (13.0 pN) on the graph to point C (13.5 pN). The correction term for stretching ssRNA of the unfolded structure was calculated as the work required to stretch the ssRNA from 0 pN to the force in point C (13.5 pN). Figure 70 show how these work contributions were calculated.

The work is defined as the area under the force-extension curve in the interval defined by the length at the force in point A and the length at the force in point C. For the ssRNA of the unfolded structure the lower limit is 0 nm @ 0 pN because the ssRNA of the structure has zero extension prior to unfolding.

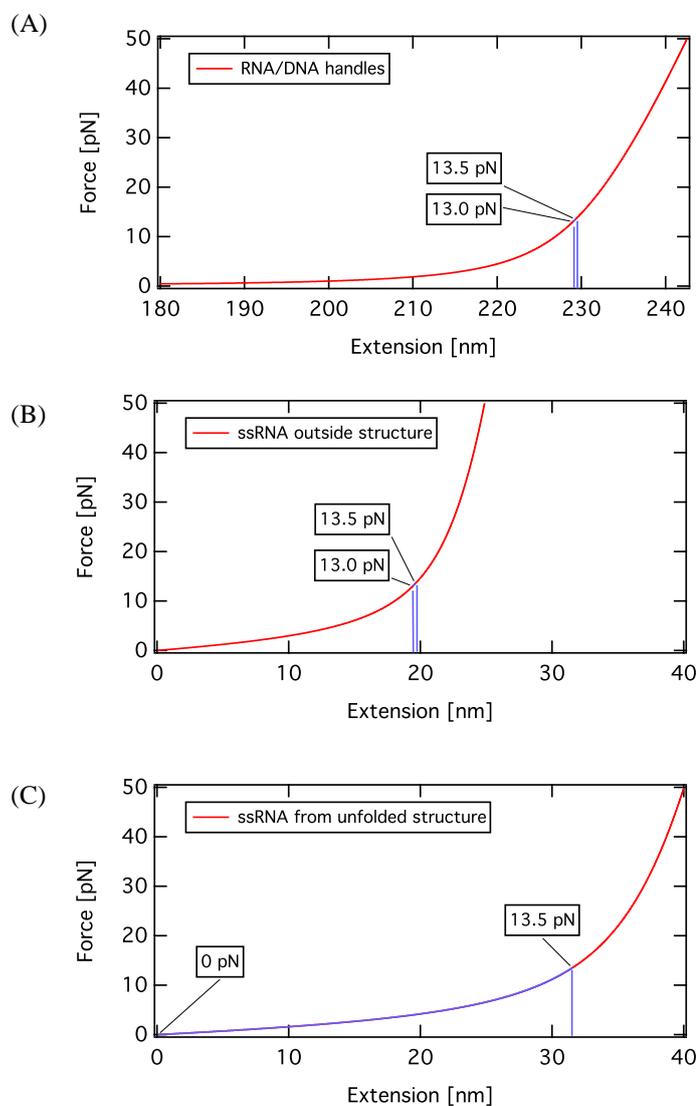


Figure 70: Correction term calculation (for Figure 69A). **A:** Calculation of work required to stretch RNA/DNA handles from 13.0 pN to 13.5 pN (indicated by blue lines). Work required:  $1.0 k_B T$ . **B:** Calculation of work required to stretch ssRNA outside the structure from 13.0 pN to 13.5 pN (indicated by blue lines). Work required:  $0.6 k_B T$ . **C:** Calculation of work required to stretch ssRNA from unfolded structure from 0.0 pN to 13.5 pN (indicated by blue line and blue trace). Work required:  $30.0 k_B T$ . **Total correction term:**  $31.6 k_B T$ . This example assumes that the unfolded structure is pseudoknot 22/6bU.