

Force spectroscopy of nucleic acid folding in the single-molecule limit

by

Daniel André Norman Foster

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Physics
University of Alberta

© Daniel André Norman Foster, 2015

Abstract

Folding of biomolecules is an important problem in structural biology. The physical folding can be projected as a diffusive search over an energy landscape whose dimensions scale by all the internal degrees of freedom which a biomolecule possesses. To explore this idea, folding is studied from the perspective of its biological relevance in RNA, and then the specifics of folding as a physical process in nucleic acids with a particular focus on DNA as a model system.

For studying the relationship between folding and function, I focus on two classes of RNA regulatory molecules whose structural dynamics are integral to their function. The first, a riboswitch, is a sequence within a messenger RNA (mRNA) that regulates gene expression. It does this by ligand binding to an aptamer domain which induces conformational changes in a regulatory expression platform. The second class of molecules is the pseudoknot, a structure also found in mRNA that promotes shifting of the reading frame of the ribosome. These RNA molecules are studied by single-molecule force spectroscopy using optical tweezers, to repeatedly unfold and refold them in order to explore the conformations they form. The study of the *add* riboswitch reveals a thermodynamically-controlled regulatory mechanism and a rare misfolding pathway. The study of a panel of programmed -1 frameshifting pseudoknots reveals that their frameshifting efficiency does not correlate with mechanical resistance to unfolding but instead with the tendency towards the formation of alternate, incompletely folded structures.

To explore the physical process of folding in detail, and to make quantitative tests of diffusive theory, we use DNA hairpins that have been previously well characterised. The first study is to determine the transition times between two states in a folding/unfolding system from energy landscape analysis,

then by direct measurement. These trajectories dominate the dynamics of the folding reaction, encapsulating critical information about how structure forms. New methods proposed for measuring the position-dependence of the diffusion coefficient are tested using the round-trip time with equilibrium data, both experimental and simulated, and by using the average fall time from non-equilibrium force jump experiments. The results deviate from expectation, as well as producing different results when started from the folded state or from the unfolded state. Finally, a novel method proposed for the landscape reconstruction from discontinuous force-jump experiments is explored with both the assumption of a constant diffusion and one that is position dependent. For both of these methods it is found that they do not replicate the results from other techniques. For all these new methods, their disagreement seems most likely because they are sensitive to or do not consider the compliance and dynamics that beads and handles introduce into the force probe.

Preface

The interdisciplinary work presented in chapters 3, 4 and 5 has involved many members of the Woodside research group. The construction, maintenance and calibration of the optical tweezers instruments for the included work has involved myself and the aid of Miro Belov, Krishna Neupane, Michael Woodside, and Hao Yu.

Chapter 4 has been published in another form as Krishna P. Neupane, Hao Yu, Daniel A. N. Foster, Feng Wang and Michael T. Woodside “*Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism*” *Nucleic Acids Res* 39:7677-7687 (2011).

Chapter 5 has been published in another form as: Dustin B. Ritchie, Daniel A. N. Foster, and Michael T. Woodside “*Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding*” *PNAS* 109: 16167-16172 (2012).

Part of chapter 6 (most of 6.1 through 6.6) has been published as: K. P. Neupane, D. B. Ritchie, H Yu, D. A. N. Foster, F. Wang, M. T. Woodside “*Transition path times for nucleic acid folding determined from energy landscape analysis of single-molecule trajectories*” *Phys. Rev. Lett.*, 109: 068102 (2012)

In the aforementioned papers, I participated in the sample preparation, measurements, analysis and writing of the manuscripts.

The remaining chapters detail new work that has not yet been published at the time this thesis was written.

Table of Contents

Abstract.....	ii
Preface	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
List of Abbreviations, Nomenclature and symbols	xiv
Acknowledgements	xvii
Dedication	xix
Chapter 1: Introduction	1
1.1 Biomolecular Folding	1
1.2 Structure in nucleic acids.....	2
1.3 Energy landscapes	5
1.4 The case for single-molecule force spectroscopy.....	7
1.5 Thesis outline	10
Chapter 2: Optical tweezers measurement techniques	11
2.1 Instrumentation	11
2.2 Sample design and measurements	12
2.3 Types of measurements	13
2.3.1 Force ramp experiments	14

2.3.2	Equilibrium measurements	16
2.3.3	Discontinuous force measurements (force-jumps)	18
2.4	Force jump implementations	19
Chapter 3: Single molecule force spectroscopy analysis		23
3.1	Elastic properties of nucleic acids	23
3.2	Force extension curve analysis	25
3.3	Constant force data analysis	28
3.4	Application of Kramers' theory of reaction rates	29
3.5	Energy landscape reconstructions	32
Chapter 4: The <i>add</i> adenine riboswitch's folding and regulatory mechanism.....		37
4.1	Introduction to riboswitches	37
4.2	Folding intermediates of the aptamer	40
4.3	Energy landscape of the aptamer	45
4.4	Misfolded states in the aptamer	49
4.5	The full length riboswitch.....	51
4.6	Comparing the mechanisms of the <i>add</i> and <i>pbuE</i> riboswitches.....	56
4.7	The evidence for thermodynamic control in the <i>add</i> riboswitch	57
Chapter 5: RNA pseudoknot conformational plasticity.....		61
5.1	Ribosomal Frameshifting	61
5.2	Pseudoknot frameshifting across a broad range of efficiencies	66
5.3	Experimental pseudoknot extension changes under tension.....	66

5.4	Pseudoknot alternate structures	72
5.5	Observed contour length changes vs. frameshifting efficiency.....	73
5.6	Possible correlations of frameshifting efficiency with mechanical parameters.....	75
5.7	Possible underlying mechanisms behind the pseudoknot pulling results	76
Chapter 6: Folding transition path times for nucleic acids		80
6.1	Transition paths	80
6.2	Average transition times from energy landscape reconstructions	82
6.2.1	Transition times for DNA hairpins.....	84
6.2.2	Diffusion and transition times from pseudoknots and the <i>add</i> riboswitch	87
6.3	Comparing the transition time and diffusion coefficient across nucleic acids.....	90
6.4	Direct measurement of transition times in DNA hairpins	91
6.5	Testing diffusive theories of folding	94
Chapter 7: Diffusion Coefficient analysis from single molecule trajectories.....		98
7.1	Introduction	98
7.2	Round-trip time analysis of diffusivity in DNA hairpins.....	100
7.3	Testing the round-trip method with simulated hairpin data.....	105
7.4	Diffusivity from average fall times in force-jump measurements	111
Chapter 8: Energy landscape reconstruction from force-jump measurements.....		114
8.1	Motivation	114
8.2	Outline of the method.....	114
8.3	Application of the force jump techniques.....	115

8.4	Analysis of the $G(x)$ calculations.....	121
8.5	Instrumental effects on the analysis.....	126
Chapter 9: Summary and Future work		128
9.1	Summary.....	128
9.2	Future work	129
9.2.1	Folding and function in RNA.....	129
9.2.2	Diffusion and landscape reconstruction.....	131
Bibliography		133
Appendices		157
Appendix A Force Jump technical implementation		157
Appendix B Extension extraction in force jump experiments		158
Appendix C High stiffness calibration.....		160
Appendix D Pseudoknot and riboswitch sequences		164
Appendix E Derivation the distribution of transit times in the limit of long times		165
Appendix F Derivation of free energy as a function of non-equilibrium relaxation profiles		166

List of Tables

Table 4.1 Aptamer unfolding contour length changes.	43
Table 4.2 Aptamer constant force results.	46
Table 5.1 Summary of results.	70
Table 6.1 Diffusion constants and transition path times from energy landscape analysis.	86
Table 6.2 Energy landscape parameters for NAs found from fits to $p(F)$ and $k(F)$	87
appendix table D.1 Sequences of riboswitch RNA molecules measured.....	164
appendix table D.2 Pseudoknot structural properties and sequences	165

List of Figures

Figure 1.1 simplest view of biomolecular folding.	2
Figure 1.2 The chemical structure of DNA.	3
Figure 1.3 Examples of structure in NAs.	4
Figure 1.4 Example crystal structure of a transfer RNA.	4
Figure 1.5 Funnel-shaped free energy landscape with multiple barriers and paths.	6
Figure 1.6 Some examples of single-molecule force spectroscopy (smFS).	9
Figure 2.1 Illustration of the experimental construct.	13
Figure 2.2 Example FECs from hairpin 30R50/T4 (horizontally offset by 45 nm sequentially).	15
Figure 2.3 FEC hysteresis diagram.	15
Figure 2.4 Constant trap separation example traces.	17
Figure 2.5 Constant force measurements.	17
Figure 2.6 Schematic of a force jump (quench) experiment with a force clamp.	19
Figure 2.7 Force profile in an optical trap.	20
Figure 2.8 A schematic of a force jump experiment with constant trap positions.	22
Figure 3.1 Example FEC overlay and WLC fits for hairpin 30R50/T4.	25
Figure 3.2 FEC analysis.	27
Figure 3.3 Energy landscape diagram for a two state system.	30
Figure 3.4 Example of a probability density for HP 30R50/T4 from constant force data.	35
Figure 4.1 Force spectroscopy of add aptamer alone.	41
Figure 4.2 Confirmation of hairpin folding identification using anti-sense oligos.	42

Figure 4.3 Unfolding force distributions and force-dependent rates determined from FECs.	47
Figure 4.5 Misfolded states in the <i>add</i> riboswitch aptamer.....	49
Figure 4.6 Misfolded states without adenine present.	50
Figure 4.7 Folding pathways of the aptamer.	51
Figure 4.8 FECs of the full-length riboswitch.....	52
Figure 4.9 Switching and ligand binding.....	56
Figure 4.10 Folding kinetics and riboswitch mechanism.....	60
Figure 5.1 Schematic of -1 PRF.	63
Figure 5.2 Secondary and tertiary structure of the pseudoknots.	64
Figure 5.3 Force spectroscopy of ScYLV C27A pseudoknot.....	68
Figure 5.4 Representative FECs for the nine pseudoknots.....	69
Figure 5.5 Representative unfolding force distributions.	71
Figure 5.6 Frameshifting efficiency is not correlated with mechanical stability parameters.....	72
Figure 5.7 Frameshifting efficiency correlates with the formation of alternate structures.....	73
Figure 5.8 Frameshifting efficiency is not correlated with mechanical stability parameters.....	75
Figure 5.9 Structural dynamics of SARS pseudoknot at constant force.....	78
Figure 5.10 Structural dynamics of HERV pseudoknot at constant force.	78
Figure 6.1 Schematic free-energy profile for a folding reaction.....	81
Figure 6.2 Extension trajectories of DNA hairpins.....	84
Figure 6.3 Energy landscape analysis of DNA hairpins.....	85
Figure 6.4 Unfolding force distributions for pseudoknots and riboswitches.	88

Figure 6.5 Stem-length dependence of τ_{ip}	90
Figure 6.6 Transition paths in force spectroscopy measurements.....	92
Figure 6.7 Response time measurement of tethered beads.	92
Figure 6.8 Instrumental response time.....	93
Figure 6.9 Transition paths for a DNA hairpin.....	94
Figure 6.10 Distribution of transit times for DNA hairpin.....	95
Figure 7.1 Round trip time as defined in an extension trace.....	101
Figure 7.2 Schematics of DNA hairpins 30R50/T4 (top) and 20TS06/T4 (bottom)	102
Figure 7.3 Constant force of HP 30R50/T4.....	102
Figure 7.4 Average round trip times example.	103
Figure 7.5 The derivatives of the average τ_{RT}	103
Figure 7.6 Round trip time diffusion coefficient for HP 30R50/T4	104
Figure 7.7 Round trip time diffusion coefficient for HP 20TS06/T4.....	104
Figure 7.8 Sample simulated lone HP extension trace with constant D (5 μ s time res.).....	106
Figure 7.9 Average round trip time from simulated HP with constant D at 3×10^5 nm ² /s.....	106
Figure 7.10 Diffusion coefficients for simulations without beads and handles.....	108
Figure 7.11 Diffusion coefficient from simulated HP landscape with varying stiffness.	109
Figure 7.12 Diffusivity from simulation and experiment.....	110
Figure 7.13 Average fall time along collapsing extension traces (τ_c).....	112
Figure 7.14 Example of diffusion coefficient from force jumps for 4.5pN drop.	113
Figure 8.1 example extension pieces.....	117

Figure 8.2 Example non-equilibrium $\rho(x)$, the histogram and KDE.....	117
Figure 8.3 Energy landscape reconstruction for HP 30R50/T4	118
Figure 8.4 Example result for the non-equilibrium $G(x)$ calculation with HP 20TS06/T4.....	119
Figure 8.5 Profile of the free-energy landscape along the molecular extension of HP 30R50/T4.	120
Figure 8.6 Example of diffusion coefficient from force jumps.....	121
Figure 8.7 Comparison of energy landscape with and without position dependence in D	121
Figure 8.8 Manually restricting the extension distribution for $G(x)$ calculation.....	122
Figure 8.9 Varying the multiplier of RHS term in $G(x)$	124
Figure 8.10 The results for the $G(x)$ calculation with HP 20TS06/T4.....	125
Figure 8.11 Varying the multiplier of RHS term in $G(D,x)$	125
Figure 9.1 Structure of the SAM-I riboswitch aptamer	130
Fig A.1 EOD prefilter repair.....	157
Fig B.1 Force-jump finding in Dy1 trace.....	159
Fig B.2 Excision example: (from mol 0_3 nov 7)	160
Fig C.1 AOD power throughput profiles of the OT instrument.....	161
Fig C.2 Laser beam profiles measured before the last telescope before the objective lens.....	162
Fig C.3 Laser power measurements at the PSDs in the form of the sum voltage.	163

List of Abbreviations, Nomenclature and symbols

A	adenine (base) adenosine (nucleoside)
aa	amino acid
AFM	atomic force microscopy
AOD	acousto-optic deflector
AOM	acousto-optic modulator
bp	Base pair
C	cytosine (base) or cytidine (nucleoside)
DFS	dynamic force spectroscopy
DIC	differential interference contrast (microscopy)
DNA	deoxyribonucleic acid
ds	double-stranded
ds NA	double stranded nucleic acid
dsDNA	double stranded deoxyribonucleic acid
dsRNA	double stranded ribonucleic acid
EOD	electro-optic deflector
FEC	Force-extension curve
FRET	Förster resonance energy transfer
FS	force spectroscopy
G	guanine (base) or guanosine (nucleoside)
HERV	human endogenous retrovirus-K10
HP	hairpin (in DNA samples)
kb	kilobase
KDE	kernel density estimation
LHS	left hand side (of an equation)
MD	molecular dynamics
MMTV	mouse mammary tumor virus
mRNA	messenger RNA
NA	numerical aperture
NMR	nuclear magnetic resonance
nt	nucleotide(s)
NTP	nucleotide-triphosphate
ORF	open reading frame
OT	optical tweezers / optical trap
PCR	polymerase chain reaction
PDF	probability density function
PEMV	Luteovirus pea enation mosaic virus-1
PK	pseudoknot (usually in RNA)

PMF	potential of mean force
PSD	duolateral position sensitive diode
PT2G32	the bacteriophage T2 gene 32
RBS	ribosomal binding site
RHS	right hand side (of an equation)
RNA	ribonucleic acid
RNAP	RNA polymerase
SARS	Severe acute respiratory syndrome
SAXS	Small-angle x-ray scattering
ScYLV	sugarcane yellow leaf virus
s.d.	Standard deviation (σ)
s.e.m.	Standard error of the mean(σ/\sqrt{n})
smFRET	single molecule FRET
smFS	single molecule force spectroscopy
SRV1	simian retrovirus-1
ss	single-stranded
ss NA	single stranded nucleic acid
ssDNA	single stranded deoxyribonucleic acid
ssRNA	single stranded ribonucleic acid
T	thymine (base) or thymidine (nucleotide)
tRNA	transfer RNA
U	uracil (base) or uridine (nucleoside)
UTR	untranslated region
VMV	Visna-Maedi retrovirus
WLC	worm-like chain (Marko-Siggia form)
\ddagger	transition state
D	diffusion coefficient / diffusivity
d_h	helix diameter
d_T	distance between termini in PKs
$F_{1/2}$	equilibrium force in a two-state system
F_{unf}	unfolding force
$G(x)$	Gibbs free energy
$G_{stretch\ of\ nt}$	stretching energy of unfolded nt at $F_{1/2}$
K	elastic modulus
k	spring constant /trap stiffness
k_0	Kramers pre-factor

k_B	Boltzmann constant
K_D	dissociation constant (in molar concentration)
$k_{F/U}^0$	Rate at zero force, with subscript specifying F/U
L_c	contour length
L_p	persistence length
n_{nt}	number of nucleotides unfolded
r	bead radius
T	temperature
t	time
β	thermodynamic beta, $(k_B T)^{-1}$
ΔG	free energy difference
ΔG^\ddagger	height of the energy barriers
$\Delta G^{\ddagger,0}$	height of the energy barriers at zero force
ΔG^0	free energy difference at zero force
ΔG_{eqm}	Equilibrium free energy
Δn_h	number of helices removed during unfolding
Δx^\ddagger	distance to the transition state
$\Delta x^{\ddagger,0}$	distance to the transition state at zero force
η	viscosity
κ	Curvature within an energy landscape
λ	wavelength
$\rho(x)$	the time-averaged probability distribution
τ_c	average fall time in a force drop trajectory
τ_p	transition path time over a barrier
$\Psi(x,t)$	probability of a configuration with respect to extension (x) at time t
ω^2	curvatures of the energy profile

Acknowledgements

To begin I must thank my long time supervising professor, Michael Woodside, who has provided guidance to all students in his lab. He has always shown concern for the standard of quality of our publications and thesis writing, recently helping me with the sentence salads of my introductory chapter drafts. Thanks to Profs. Jack Tuszynski and Mark Freeman, who offered helpful advice, a helpful ear and reality checks during the many supervisory meetings over the past years. Thanks is extended to Profs. Mariusz Klobukowski, Sharon Morsink and Hongbin Li (UBC) who served on my defense examination committee. They have taken time out of their stacked September schedules to make insightful contributions to my thesis, especially with the new methods we are studying.

Thanks to many contributors to the optical tweezers backend: Hao Yu, the alignment man, Miro Belov for his additional help even after leaving our group, for Krishna Neupane, the accessible reference for calibration, troubleshooting, and other optical trapping issues. Thanks to help in the biochemistry factory: Feng Wang and Dustin Ritchie, who both have helped with learning, sample design, troubleshooting, as well as helpful discussions with Derek Dee (also Craig Garen and Prasanna Bhomkar!). Thanks to Dave Fortin for his help with labview problem solving, and to Don Mullin, who finds a way to help out all students in condensed matter research. Thanks must also be extended to all the members of the Woodside lab not mentioned here.

Our lab was built with much help from the machine shop: Gilbert Lachat and Paul Zimmermann who helped in the design of our microscope reconstruction, later revisions and the optics boxes for our three OT instruments. The late Boris Tomasevic worked with us to machine our

tediously large optics boxes. Tony Walford and Tony Paget also made contributions in making our OT lab happen.

Specific shout-outs to my friends in Edmonton who helped with the editing of my long-lullaby thesis, namely Allison Solanki & John Hook, Luda Dorosh, Krishna Neupane, Medhi Shaghghi, and others I have forgotten to mention here. Of course, I could also not have gotten here without the friends at the U, most of whom have gone on to post docs at other universities, or to jobs outside of academia – their friendships make it possible to survive the times when grad school becomes a slog.

Thanks to the Kinsellas who always insisted on inviting me over as a family member. Special thanks are owed to my parents and sisters, whose have supported me throughout my years of education.

Dedication

To my parents, Lyse and Michael, who have always been supportive and never forgot about me, and to Kinsella family, who are my family away from home up here in Edmonton.

Dédicace

A mes parents, Lyse and Michael, qui m'ont toujours donné leur appui, et qui ont pensé à moi durant ma thèse. Et la famille Kinsella, car ils sont mes seuls proches ici à Edmonton.

Chapter 1: Introduction

1.1 Biomolecular Folding

The folding of biological macromolecules like proteins and nucleic acids is central to biology at the nanoscale. Whole new fields of research have been spawned by the study of this folding (Dill and MacCallum, 2012). The question of how this folding occurs is central to understanding how living cells function and to the biochemistry driving life itself. It has been well popularised that the protein folding problem is a complex one (Dill and MacCallum, 2012), without an easy solution to predict the structure of a protein by its underlying amino acid sequence. The folding problem applies equally well to nucleic acids (NAs). Although the underlying structure of the deoxyribonucleic acid (DNA) double helix, such as the base-pairing and stacking, was solved a lifetime ago by Watson and Crick with the help of Rosalind Franklin's x-ray photographs of DNA's crystal structure (Watson and Crick, 1953), there is much remaining to be studied (Dolgin, 2015). A central example is the ribosome, a molecular machine with ribonucleic acid (RNA) and protein parts. The active site of the protein synthesis comes from the RNA components (Alberts, 2008; Wimberly et al., 2000). Another example of a large complex encompassing both protein and RNA parts, also known as a nucleoprotein, is the spliceosome (Schellenberg et al., 2013). There are more levels of complexity in DNA as well; for example, at the end of chromosomes. Here we find telomeres, which are long sequences of guanine repeats, forming quadruplexes (Bochman et al., 2012). There is much complexity to be found in NAs, just as there is in proteins.

When discussing the folding problem, proteins are a good place to begin. It was proposed early on that protein folding was spontaneous, and that all the information required for this folding was contained in the amino acid sequence of a protein (Anfinsen, 1973). In opposition to this observation, given the large number of degrees of freedom in an unfolded NA or in a protein, a local search through all the conformations, such as all the possible bond rotations, would take a length of time in a scale similar to that of the lifetime of the cosmos. This paradox (Dill and MacCallum, 2012) was put forward by Levinthal, with an example estimate of 10^{143} conformations for a polypeptide chain (Levinthal, 1969), which would take a time longer than the age of the universe for a protein to settle into its so-called native conformation. Clearly, life driven by evolution has selected for folding to be on a much smaller timescale

(microseconds to minutes). To summarise: an unfolded protein (or NA) goes from an unstructured chain to its native structure generally quite quickly (Figure 1.1), without a simple random conformational search.

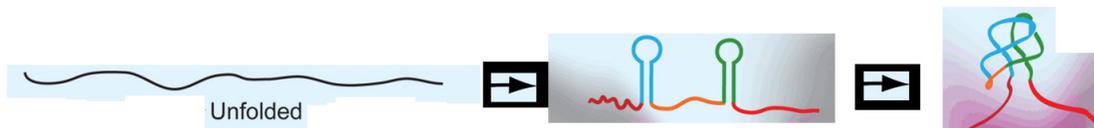


Figure 1.1 simplest view of biomolecular folding.

An unstructured chain of peptides or nucleic acid bases fold up into biologically relevant structures. Here the sequence of folding for the *add* riboswitch aptamer is displayed in condensed fashion.

1.2 Structure in nucleic acids

We can approach nucleic folding much in the way we approach protein folding. The description of NA structure starts with that of the five carbon sugar that forms the backbone and the five principle bases found in RNA and DNA. These bases are of two types, pyrimidines (thymine (T) (in DNA) and uracil (U) (in RNA), and cytosine (C)) and purines (guanine (G) and adenine (A)). The base pairing provides the interaction to produce the primary structures in NAs, with the canonical base pairings between these, $C \equiv G$ and $A=T$ or $A=U$ (in RNA) being dictated by hydrogen bonds (see Figure 1.2). It is the stacking of aromatic rings between the planes of the bases, which result in three-dimensional structures such as the A and B-helices, typically in RNA and DNA, respectively (Neidle, 1999; Saenger, 1984). The ribose sugar of RNA has a 2' hydroxyl (OH), leading to the possibility for more complex hydrogen bonding networks, such that RNAs often takes on more complex structures (such as riboswitches, in chapter 4, pseudoknots in chapter 5, (Neidle, 1999; Saenger, 1984), and non-riboswitch RNA ligand sensing: (Boerneke and Hermann, 2015)).

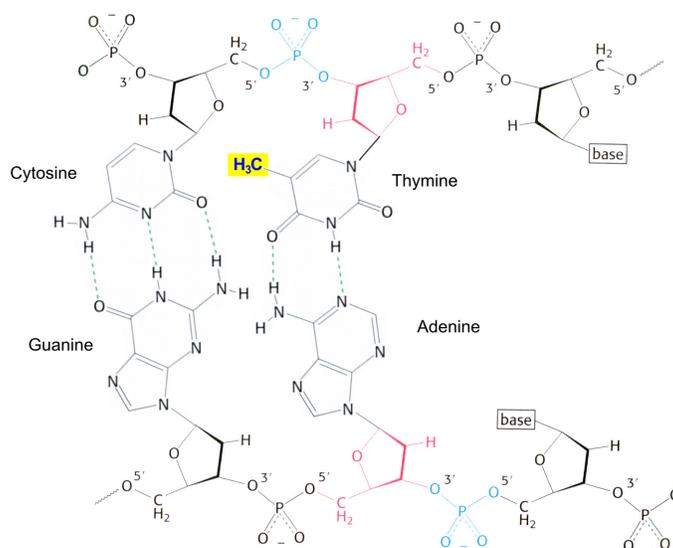


Figure 1.2 The chemical structure of DNA.

RNA differs by having a 2' hydroxyl group (not shown) NAs use 4 bases: adenine (A) pairs with thymine (T), guanine (G) pairs with cytosine (C). Thymine has an extra methyl group that differentiates it from uracil (U), found instead of T in RNA. The phosphate groups linking the (deoxy)ribose sugars are charged. Canonical G:C pairs are more stable than A:T or A:U pairs due to additional hydrogen bonds between the bases (green dashed lines) (Figure adapted from (Berg et al., 2010)).

From the base-pair formations, secondary structure arises in NA: two complementary sequences will form duplexes, hairpins (a ubiquitous feature), and kinks and turns when there is a gap in the complement on one strand, and mismatches will produce bulges and loops. As these structures interact with one another, we find the formation of tertiary structure, such as the interaction of the unpaired bases of two loops (so-called kissing interaction), the looping back of a strand with base triplexes can support pseudoknots, and multiple duplexes or hairpins that can join into multiple helix junctions (Figure 1.3). Structures quantified by techniques such as nuclear magnetic resonance (NMR) or X-ray crystallography illustrate tertiary structure more clearly by providing a 3D image of said structure (Figure 1.4). It should be pointed out that hairpins in particular have found many applications such as molecular beacons (e.g. (Cao et al., 2015)), and recently as molecular tension probes (Blakely et al., 2014).

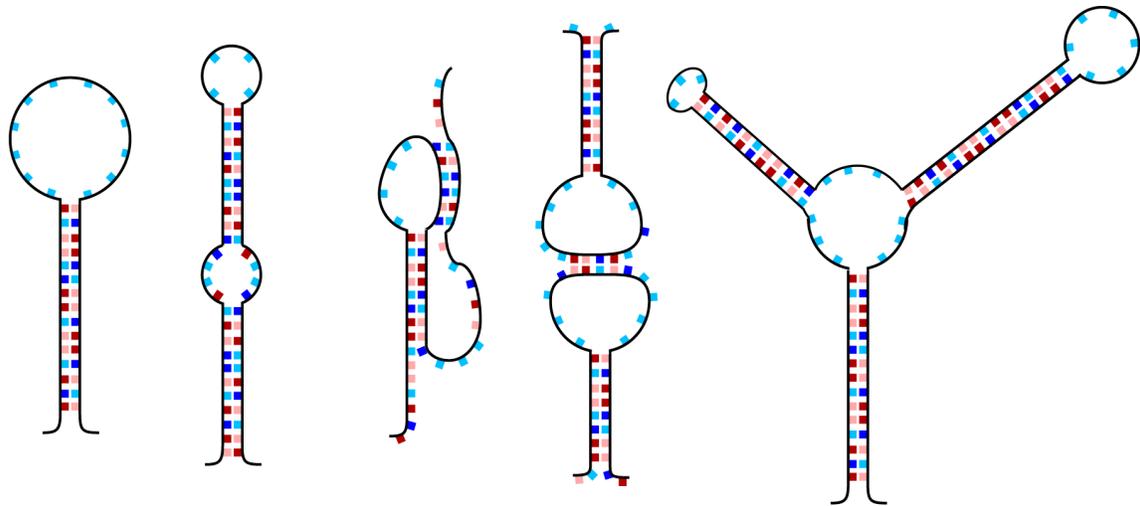


Figure 1.3 Examples of structure in NAs.

From left to right: a hairpin with a large loop, a hairpin with a bulge; both examples of secondary structure. Flat representations of tertiary structures: a pseudoknot geometry, a pair of kissing loops, and a three stem junction. The three stem junction is a feature of the *add* riboswitch aptamer, which serves as a ligand-binding pocket, while the upper helices curl up and form a kissing interaction with their respective loops.

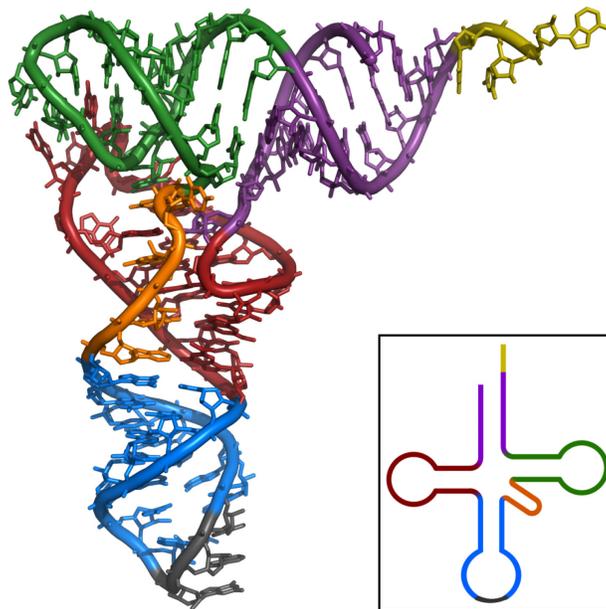


Figure 1.4 Example crystal structure of a transfer RNA

The x-ray crystal structure of the phenylalanine transfer RNA from yeast, rendered in PyMOL, from the PDB #ehz, with the secondary structure inset. This RNA presents a four helix junction, with a loop-loop interaction (red-green). (Figure from wikimedia commons, by Yikrazuul).

The complexity of folding is hard to overstate. We can identify the underlying sequences of DNA and RNA, and predict their structures to some extent (e.g. *mfold*, (Zuker, 2003), and make pseudoknot specific predictions (Andronescu et al., 2003)), but these often produce a wide range of possible structures. The prediction of the structure of a protein from the underlying sequence of its incorporated amino acids has improved by leaps and bounds, though it remains a computationally expensive process that also does not always produce accurate results (Guliaev et al., 2012; Shaw et al., 2008; Wolynes, 2014). Hairpins provide a good model system to study nucleic acid folding dynamics and to test out new experimental methods.

1.3 Energy landscapes

A way to illustrate folding quantitatively is to project it as a diffusive search, in conformational space for the minimum energy structure, over an energy landscape. This landscape's dimensions scale by all the internal degrees of freedom that a biomolecule possesses, which we would describe mathematically as a multidimensional hypersurface. There are two views of protein folding pathways: the classical one of deterministic pathways with distinct intermediate states, and the modern statistical view (e.g. (Wolynes, 2014)), where there exists an ensemble of pathways, competing intermediate states, and so forth. These are part of a bigger picture referred to as the energy landscape. For the simplest case, let us consider a two state system, with an unfolded state in one well, a barrier separating it from a lower well representing the folded, native state. If this molecule were to be measured from when it was initially unfolded, some lifetime for the unfolded state could be quantified as well as a rate over the barrier to the unfolded state. Similarly, depending on the barrier height and the difference in energies between the two states, there would be some rate for the protein or NA to unfold. In experiments, some proteins and NAs are measurable as two state systems. This multi-dimensional landscape, may be projected down to only a few dimensions, limited by the experimenter's probes (by adjusting temperature, pH, pressure, force and/or molecular extension, for example). The projection of the energy landscape onto a single dimension produces a free-energy profile, or potential of mean force (PMF) (Dudko et al., 2011; McQuarrie, 2000). So what might be measured as only two states may in fact reveal other states with a different measurement modality. The underlying folding trajectory through the landscape will be a statistical

combination of all the possible paths, as the diffusive search goes through each degree of conformational freedom of a biomolecule.

The Anfinsen thermodynamic hypothesis posits is that the folded state is at the lowest accessible free-energy minimum under the natural folding conditions. As an extension of this, evolutionary pressures have shaped this energy landscape into a funnel, where proteins are made to fold downhill energetically, and shrink in terms of conformation as they are driven to the bottom (Anfinsen, 1973; Anfinsen et al., 1961; Englander and Mayne, 2014; Lapidus, 2013; Schuler and Hofmann, 2013; Žoldák et al., 2013). To reinforce this funnel idea, Zwanzig et al. calculated that a free energy bias of $2 k_B T$ towards the correct interactions must be present for a protein to fold completely on a timescale of seconds (Zwanzig et al., 1992). Generally speaking, while few microsecond folders are known for proteins, most are known to fold on the timescale of milliseconds to minutes (from hundreds of ns, to few μs , to ms, (Dill et al., 2008), or even longer (Kubelka et al., 2004; Milanesi et al., 2012)). For the case of such funnel folders, minor mutations will shift folding routes, but not the final native state (Onuchic and Wolynes, 2004). The folding funnel idea has had great qualitative success (Wolynes, 2014).

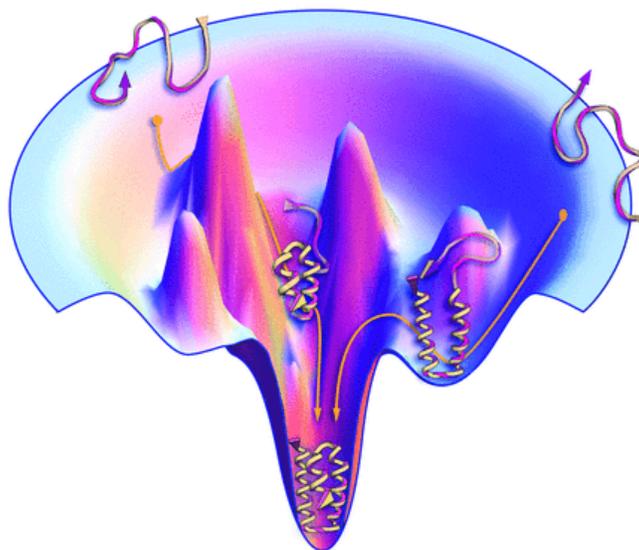


Figure 1.5 Funnel-shaped free energy landscape with multiple barriers and paths

Proteins have a funnel-shaped energy landscape with many high-energy, unfolded structures and only a few low-energy, folded structures. This illustration applies equally well to complicated RNA structures, with their native state at the bottom. (See chapters 4 and 5 for alternate pathway results.) Folding occurs via alternative

microscopic paths. Figure adapted from the review article of Dill and MacCallum (Dill and MacCallum, 2012).

The classical view of protein folding describes it as sequential process with discrete intermediates (Englander and Mayne, 2014; Onuchic and Wolynes, 2004), which is often agreeing with what can be observed in experiments. On the energy landscape, these discrete states will feature as deep wells. Sometimes exponential folding kinetics has been observed, which point to an ensemble of transition structures (states) of which the protein will fold through. Competing structures in the energy landscape (Onuchic and Wolynes, 2004) show that the classical sequential pathway picture is not always true, and experimental measurements may not always be able to detect expected intermediates. Theoretical and computational studies support the modern statistical view of the presence of multiple folding pathways. Simple models have also succeeded despite ignoring stabilising misfolding interactions that could produce kinetic traps (Englander and Mayne, 2014; Onuchic and Wolynes, 2004; Wolynes, 2014).

Given the intrinsic complexity of the energy landscape, it is hard to exaggerate the difficulty in measuring it. The landscape is central to the physical theory of folding, but its application in experimental work has only been qualitative. Furthermore, because it is often projected down to a single dimension by the experiment, it may not capture the underlying dynamics properly. Thus experimental knowledge of a protein-folding landscape and that of NA folding landscapes remain quite limited. The details of what happens between the unfolded and folded states, even if those are the only distinctly measurable states, are important to understand the folding process itself. Many methods have been developed to measure landscapes using single-molecule force spectroscopy (smFS), (as discussed in chapters 6, 7, and more specifically in 8), making it an especially powerful tool for the study of folding. To explore these ideas in this thesis, folding is studied from the perspective of its biological relevance in RNA, and then the specifics of it as a physical process in nucleic acids with a specific focus on DNA as a model system.

1.4 The case for single-molecule force spectroscopy

There are many types of measurements for probing the properties of a biomolecule that pertain to the energy landscape. Numerous bulk experiments (e.g. calorimetry) have gleaned important results in protein and NA folding, and while they measure an ensemble average, they are good for revealing overall

rates (van Oijen, 2008). There are, however, some tradeoffs. It is not always possible to synchronise the folding in an experiment with a large sample size, i.e. molecules often begin and end folding at different points in time (Savinov et al., 2014). In addition, heterogeneous behaviour may be revealed, but its sub-populations may be masked (Hyeon et al., 2014). Take for example a large population of molecules that has been measured and characterised: if we can isolate a single molecule from it, then we can follow that molecule's individual folding trajectory from start to finish, sometimes illuminating previously unseen pathways and states. In other words, intermediate folding states can be seen directly without otherwise being averaged-out. Generally speaking, there has been an ongoing concern about the reproducibility of experiments in biology, e.g. (Freedman et al., 2015), so another advantage presents itself to single molecule measurements: due to the purity of solution components and the simplicity of the experimental setup, the reproducibility of these experiments is also good.

Many approaches have now been established for single molecule experimentation. These include particle tracking, single molecule approaches using fluorescence as well as Förster resonance energy transfer (FRET). These techniques can also be parallelised to record multiple trajectories at once, and after some initial external perturbation, the behaviour of the molecule can be observed passively (Greenleaf et al., 2007; Zhuang, 2005). Force spectroscopy at the single-molecule level offers a different and complementary modality to these other techniques. A key advantage of smFS is that it provides one of the few good ways to measure energy landscapes experimentally (Woodside and Block, 2014). Different methods for landscape reconstruction, based on different physical principles are available. The first use equilibrium thermodynamic properties (Woodside et al., 2006a; Woodside et al., 2006b), others make use of non-equilibrium thermodynamic properties (Engel et al., 2014; Gupta et al., 2011; Hummer and Szabo, 2001), and some on the kinetic properties (Dudko et al., 2006, 2008), while others work on the statistical properties of diffusion (Lannon et al., 2013; Manuel et al., 2015). This ensures that if force spectroscopy is appropriate to the biomolecule under study, then the variety of methods ensures that there is usually some way to characterise the energy landscape properties for that molecule (despite particular properties that might make some forms of measurement difficult to implement, such as sample purity constraints or aggregation tendencies). Moreover, it is often possible to apply multiple

measurement techniques, providing an increase in the confidence of a result (Woodside and Block, 2014).

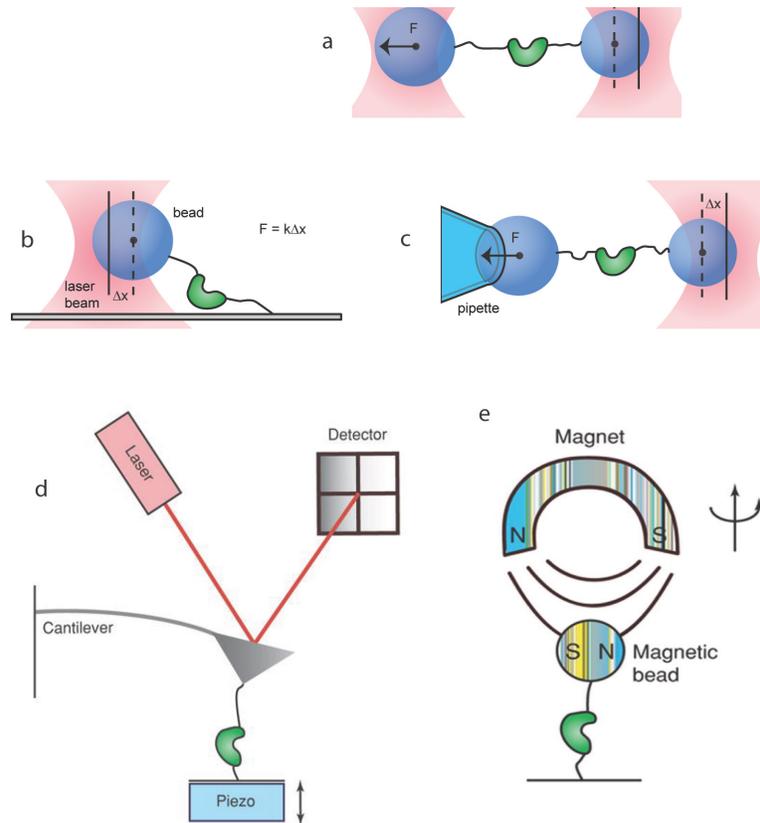


Figure 1.6 Some examples of single-molecule force spectroscopy (smFS)

a to c are optical trap (OT) configurations, a: the dual beam OT used in our set up, b: a single trap surface assay, c: using a pipette as the second anchor point. d: a schematic of in solution AFM, where force is adjusted by moving a piezo-electric stage, e: magnetic tweezers can be used to apply both force and torsion on a molecule. (Greenleaf et al., 2007)

There are a number of approaches available to do smFS (see Figure 1.6). These include atomic force microscopy (AFM) in solution (aqueous), viscous fluid flow, biomembrane force probes (Evans et al., 1995), magnetic tweezers and optical tweezers (OTs). AFM provides forces within the nanoNewton range (nN) and recently down to the tens of picoNewtons (pN) (Sullan et al., 2013), but they typically work with a very stiff probe. Magnetic and optical tweezers function in with the limit of smaller forces (pN), and with lower stiffnesses (few pN/nm to 0.1pN/nm or lower) (Greenleaf et al., 2007). In the case of

OTs, the low stiffness is accompanied with higher spatial resolution. Both AFM and OTs may be constructed to have high stability in their applied force (Neuman and Block, 2004; Perkins, 2014). The measurement of extension in smFS is one of the few types of measurements for which the quality of the reaction coordinate (extension) has been tested and found to be an appropriate projection (at least for DNA hairpins) (Morrison et al., 2011; Neupane et al., 2015): the measured length changes can be converted straightforwardly to the number of residues or nucleotides involved. (Reaction coordinates can be complicated because they are nonlinear functions of the atomic positions (Best and Hummer, 2010)). Our goal herein is to use force spectroscopy to understand the folding thermodynamics and kinetics of NAs at the nanoscale.

1.5 Thesis outline

The ambition of this thesis is twofold: first off, smFS and energy landscape analysis are applied to understand how RNA folding affects function. Secondly, we seek to extend smFS techniques to study the physics of folding in terms of the diffusion and transition through energy landscapes. The measurement techniques of OTs are discussed in chapter 2, and then the analysis methods for smFS are introduced and explained in chapter 3. We have applied these methods to study two specific types of RNAs: riboswitches and pseudoknots. In chapter 4, competing structures in riboswitches are studied, showing how ligand binding regulates gene expression, whereas chapter 5 describes work probing how the resistance to mechanical unfolding of pseudoknots that stimulate programmed -1 frameshifting is related to the amount of frameshifting stimulated. In chapter 6, we investigate the physics of folding in more detail, looking at the specific time required to undergo structural transitions as a probe of conformational diffusion, determining this transition time from energy landscapes and then measuring it directly. From then on, using our well-studied DNA hairpins, in chapter 7, we test a theoretical method to determine the position-dependence of the diffusion over the landscape, and then in chapter 8, we explore a newly proposed method for landscape reconstruction from force-jump measurements. The last chapter, 9, includes a discussion of the future directions for the concepts within the thesis and the more subtle effects of the beads and handles on experimental measurements.

Chapter 2: Optical tweezers measurement techniques

2.1 Instrumentation

An optical trap (OT) is a device that allows one to capture a small dielectric object in a focused laser beam. (Also referred to interchangeably as optical tweezers (OTs)) Once an object is trapped, it can be moved and manipulated in a controlled manner. Our OTs are built around an inverted microscope, as the microscope optics allow for the tight focussing of a laser beam that is required, at the same time as providing a convenient way to observe the sample visually. In this case, the trapping takes place on a microscope slide placed between the microscope condenser and objective lenses. The objective lens must be of a high numerical aperture (NA, defined in microscopy as the half angle of the maximum cone of light that can enter or exit the lens), in order to focus beams at a high angle to the normal. This focusing of the laser creates the conditions for the attractive force that acts on a dielectric object, such as a polystyrene bead. From the electromagnetic interaction between the light and the dielectric, there exists a small region where the bead is held by a harmonic force, akin to a Hookean spring in three dimensions (Foster, 2010; Neuman and Block, 2004).

OTs have become well established as a biophysical tool ever since the pioneering efforts of Arthur Ashkin, Steven Chu and their colleagues at Bell Labs who first applied optical trapping for bio-compatible measurements (Ashkin and Dziedzic, 1971; Ashkin et al., 1986), and the technology continues to be further refined by the contributions of many groups, with numerous applications (Perkins, 2014). Some special considerations must be taken into account when manipulating single biological molecules. They themselves are too small to trap directly, so instead the molecules must be handled by proxy: they are attached to dielectric beads via molecular “handles.” These beads are then held in the traps. The dielectric beads are functionalised chemically, allowing them to be connected to the molecule being studied with geometric specificity. This specificity ensures that the design of the experiment (bead—handle—molecule) is well controlled (see Figure 2.1).

The construction of the OTs we use for single-molecule force spectroscopy (smFS) in our lab has been described previously (Foster, 2010). In brief, our OT uses a 1064-nm diode-pumped solid-state infrared laser to create two traps with orthogonal polarizations. The bead position in each trap is monitored by measuring the light of a 633 nm He-Ne laser, similarly split into two beams by polarization

and scattered off the two beads, using position sensitive detectors (formerly Pacific Silicon Sensor, now First Sensor). The trap stiffnesses are controlled independently by adjusting the power applied to acousto-optic deflectors (AODs) in the paths of the trapping beams. The stiffness of each trap is calibrated from an average of three different methods as described previously (Foster, 2010; Svoboda and Block, 1994). The position of each trap is controlled in the pulling axis using an electro-optic deflector, and in the orthogonal axis using an AOD. The optical trapping instruments are located in a low-vibration environment where the temperature each room is maintained at $20.0 \pm 0.1^\circ\text{C}$ during the measurements using a closed water-cooling system. All measurements are performed using the Labview platform (2012 SP1). Data are sampled digitally at twice their low-pass filtered rates with an 8-pole low-pass filter (Krohn-Hite models 3384 and 3988, 48dB/octave attenuation). (The data is sampled at twice the Nyquist frequency to avoid aliasing effects.)

2.2 Sample design and measurements

In general, RNA constructs, and a few DNA hairpins were created by inserting the sequence for the molecule of interest into the pMLuc-1 plasmid between the SpeI and BamHI restriction sites. The resulting transcription template containing the sequence flanked on each side by kb-long “handle” sequences is transcribed *in vitro* using T7 RNA polymerase. They are then annealed to single-stranded (ss) DNA complementary to the handles, and attached to beads held in optical traps (Figure 2.1), as described previously (Neupane et al., 2011). The 3' end of the transcript and labeled with biotin, and the 5' end of the transcript and labeled with digoxigenin. Otherwise, DNA hairpins have been prepared as described previously (Woodside et al., 2006b), but with three abasic sites. The handles are annealed with the construct transcript then incubated with 600 nm and 820 nm diameter polystyrene beads labelled with avidin DN (Vector Labs) and anti-digoxigenin (Roche), respectively, to create dumbbells. (Any other specific alterations are described in later chapters.)

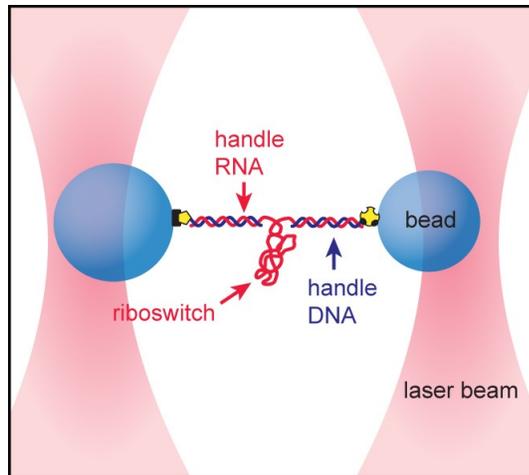


Figure 2.1 Illustration of the experimental construct

A ssRNA or ssDNA annealed to two complementary ssDNA handles, the latter of which have the chemical functionalisations to link them to polystyrene beads which are trapped optically (left: dig-antidig, right: biotin-avidin).

2.3 Types of measurements

smFS involves measuring the change in extension in a molecule as it folds or unfolds under the effect of a force applied to the molecule. The extension of the molecule is determined very precisely by measuring the position of the beads attached to the handles holding the molecule, and this extension is then recorded—along with the force on the molecule as determined from the position of the bead within the trap—as a function of time during the measurement. This approach offers three different measurement paradigms. First is force-ramp measurement, yielding the force-extension curve (FEC), in which the molecular extension is monitored while changing the force applied to the molecule. (This measurement is typically out of equilibrium because of the changing force.) The second, equilibrium extension measurements, includes the force clamp measurement, in which the extension is measured while maintaining a constant force on the molecule and the constant trap separation configuration, which provides more folding transitions than the force clamp. The third type is the force-jump, where the force is changed discontinuously (Woodside and Block, 2014). These measurements formally represent different ensembles, providing both similar and complementary information (Dudko et al., 2006, 2008; Kreuzer et al., 2001; Manosas and Ritort, 2005).

2.3.1 Force ramp experiments

The dynamic pulling experiment typically involves pulling on the molecule with a constant extension velocity, and provides a plot of the force as a function of this extension, (force-extension curve or FEC). As the force increases, structures in the molecule will be pulled apart. Whenever a structural component of the molecule unfolds, the part of the molecule that had been folded up is stretched out under the applied force, causing an abrupt increase in the molecular extension. Since the extension increase lets the bead move back towards the centre of the trap, there is a simultaneous decrease in the force resulting in a “sawtooth” pattern, or a “rip”, whenever structural features unfold. When structural features remain folded, the force rises monotonically with extension as the molecule and attached handles are stretched. FECs thus have a very characteristic shape (e.g. Figure 2.2, Figure 4.1, Figure 5.3). The force can also be ramped down at constant velocity to observe refolding, and if unfolding and refolding FECs are overlaid, hysteresis can be observed (Figure 2.3). Note that FECs reflect all elastically compliant components in the experimental setup, including not just the molecule under study but also the trap stiffness and the stiffness of the dsNA handles. The analysis of FECs must therefore take this into account, as well as the fact that the measurement is not necessarily in equilibrium due to the changing force (Foster, 2010).

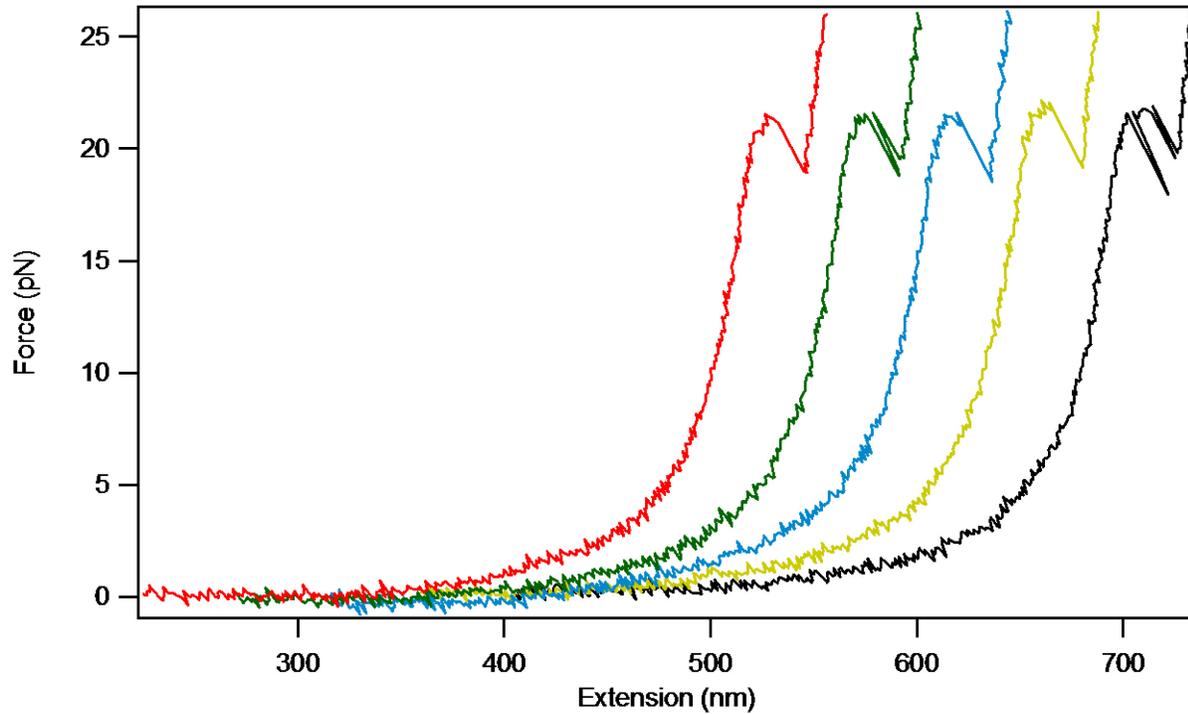


Figure 2.2 Example FECs from hairpin 30R50/T4 (horizontally offset by 45 nm sequentially)
 In repeated unfolding FECs, the hairpin unfolds over a distribution of forces. With 30R50/T4, unfolded/folded switching can be seen at the unfolding force (green, black). Data sampled at 20 kHz.

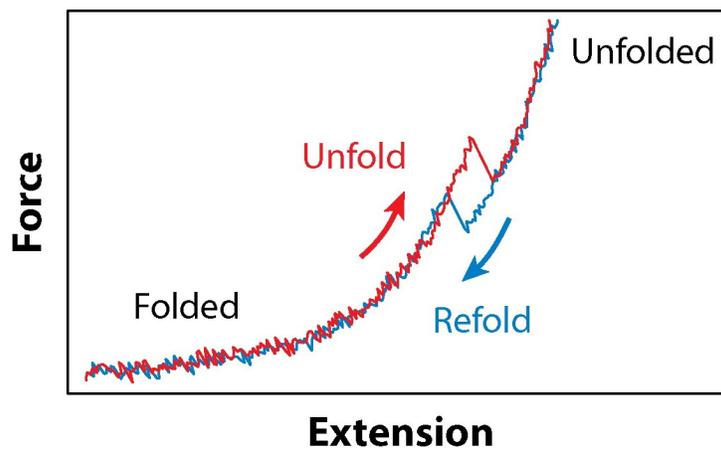


Figure 2.3 FEC hysteresis diagram
 Example unfolding FEC in red, and refolding FEC in blue. Hysteresis between unfolding and refolding curves indicates an underlying nonequilibrium process. Figure adapted from (Woodside and Block, 2014)

2.3.2 Equilibrium measurements

Constant trap separation (position clamp or constant trap position) is the simplest to implement: the molecule is pulled apart to the region of the FEC where the molecule can unfold and refold back and forth (hopping) and held fixed there. These are easier to interpret than FECs because the system is in equilibrium (no energy dissipated), but we still have to account for the changes in force (Figure 2.4). This force varies linearly with the molecular extension owing to the finite probe stiffness (Neuman and Block, 2004; Woodside and Block, 2014). (Both beads are in the harmonic regions of their respective traps.) The use of a constant position instead of a force clamp, means that when molecular extension shortens, force *increases*, and when the molecular extension lengthens, force similarly *decreases*. This has the effect of quickening the rates in both folding and unfolding directions, allowing the recording of more transitions than could be seen with a constant force configuration. A good example application of the constant trap separation can be seen in the folding of calmodulin, where the molecule displays long dwell times between folding events (Stigler and Rief, 2012).

Measurements at constant force are more technically demanding, but easier to interpret, since there is no concern about the force being ramped with time: the applied tension is kept constant, and extension is measured simply as a function of time (50 – 250 kHz sampling rate) (e.g. Figure 2.5, Figure 4.1D, Figure 4.5A, Figure 5.9, Figure 5.10). In practice there are two ways to implement a constant force. An “active” force clamp uses a negative feedback loop to maintain the trap position in such a way that the force will remain constant. This is done by monitoring the displacement of the bead from the centre of the trap; whenever the bead moves within the trap, the trap is moved by precisely the same amount in order to keep a constant bead displacement and hence applied tension. However, this method has a limited bandwidth because of a finite feedback loop closure time, which can introduce measurement artefacts when the feedback loop attempts to catch up to events that are too fast (Elms et al., 2012; Manosas and Ritort, 2005; Seol and Perkins, 2009). This can result in spurious transitions and changes in the apparent dynamics.

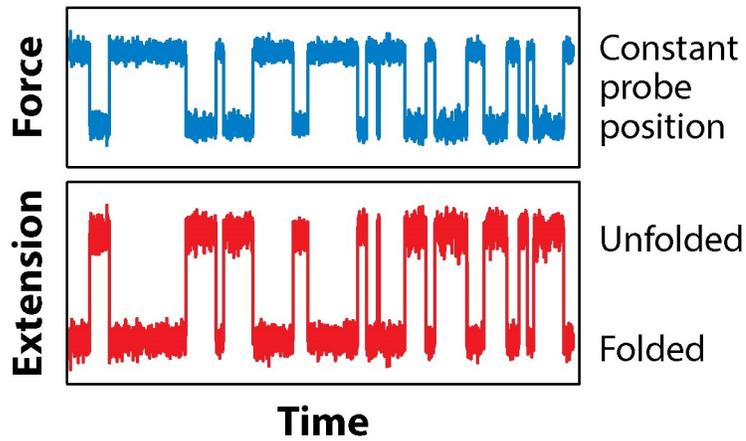


Figure 2.4 Constant trap separation example traces

Both the extension and the force fluctuate as the structure changes.

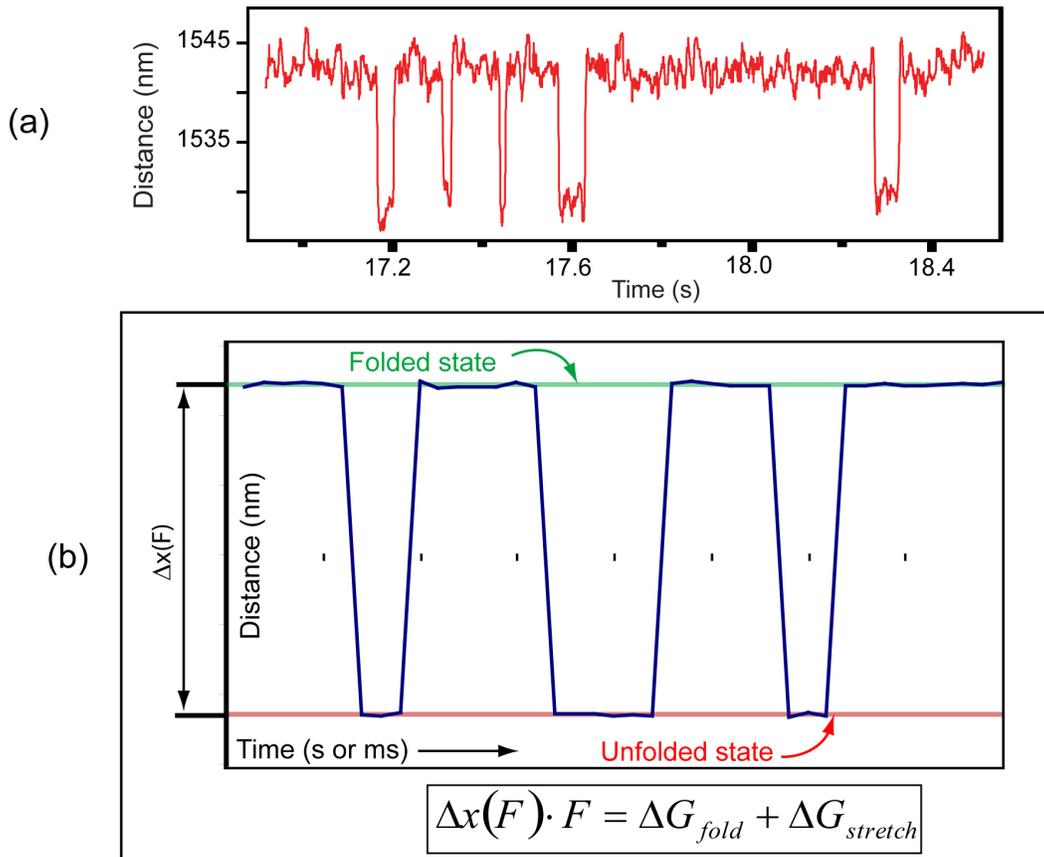


Figure 2.5 Constant force measurements.

(a) Extension vs time at constant force for the *pbuE* aptamer. (Foster, 2010) (b) Extension changes and lifetimes can be measured directly from the recorded traces.

To avoid the complications of a feedback loop, constant force can be maintained by using a “passive” force clamp. This method exploits the anharmonic region of the trapping potential: near the maximum of the force displacement curve of the trap, the force is approximately constant (*i.e.* the trap has zero stiffness) for small displacements of the bead. The tension applied to the molecule in this approach is set by adjusting the intensity of the trapping beam (Greenleaf et al., 2005). In a dual optical trap setup as used in the measurements here, one trap is stiffer than the other. The bead is held in the anharmonic region of the weaker trap, and measurements of position are made in the weak trap while measurements of force are made in the strong trap (in which the bead remains in the harmonic region at all times). The bead in the strong trap is also stationary on average if the force is not changing, so all motion from folding/unfolding occurs in the weak trap. The size of this constant-force region (within a few percent) is limited to about 40 to 50nm, which is sufficiently large for most measurements (Foster, 2010). The observation of larger unfolding distances will require some active feedback from the stage (in a surface assay), or of the trap position, to stay within the desired trapping region.

2.3.3 Discontinuous force measurements (force-jumps)

A third measurement modality involves perturbing the system as far from equilibrium as possible, by jumping the force discontinuously between two values. (To quickly jump the molecule to a force where it must unfold, or to rapidly drop or quench the force down to a small value where refolding occurs.) The molecule will take some time to respond to the jump, after which it will unfold/refold. Ideally, after the jump, the force is kept at a constant value, so that the change in structure occurs at a constant force. When a molecule folds too slowly under equilibrium conditions for it be practical to make such measurements, force-jumps allow the structural changes to be observed under constant-force. This has become popular for molecules with slow kinetics (e.g. pseudoknots (Chen et al., 2007), and especially for protein folding with AFM (Jagannathan and Marqusee, 2013). Force-jumps have been used most commonly in AFM measurements, but they also have been implemented with tweezers previously, using feedback loops to maintain constant force (Green et al., 2008).

2.4 Force jump implementations

As was discussed previously for the force-clamp measurements, when implementing a force jump with OTs, we would like to avoid artefacts in the measurement, such as those that can be contributed by feedback loops. If we use a force-clamp at the end of the jump, which has not yet been done, then the folding transition can take place over a constant applied force. In order to accomplish this, careful positioning of the beads is required so that the bead movements are restricted to lie within the anharmonic region of the trap as folding/unfolding occurs. Then the subsequent folding events' discrete steps can be observed at constant force. (See Error! Reference source not found.)

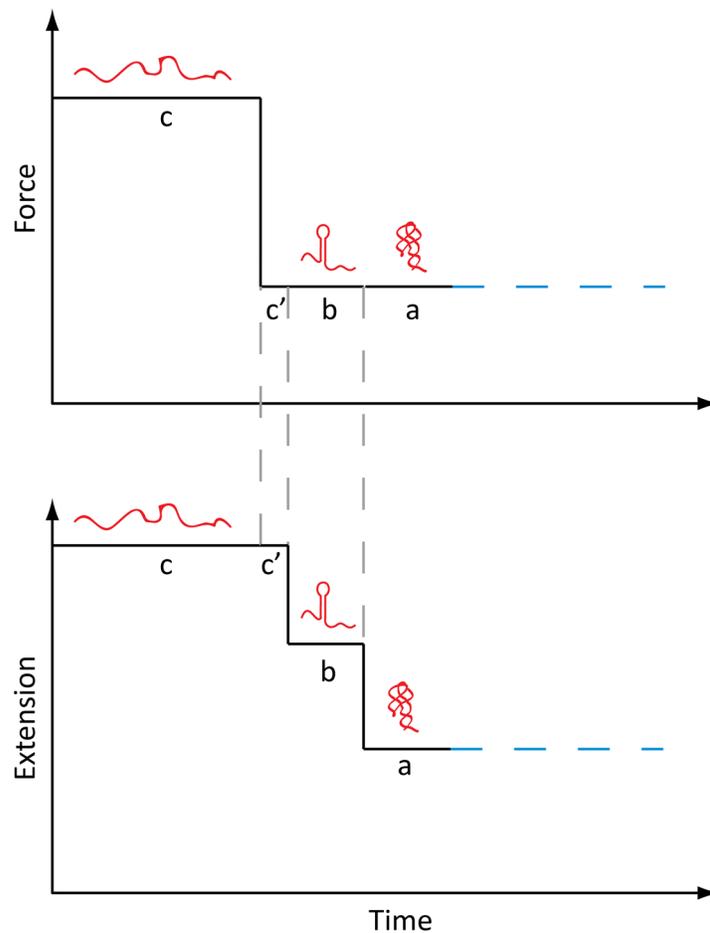


Figure 2.6 Schematic of a force jump (quench) experiment with a force clamp

Configuration for a force quench to probe refolding: c) Initially a waiting period at a high force, to assure enough time for the molecule to be unfolded. c') The molecule is immediately dropped to a low force with a force clamp, b) the molecule folds into a possible intermediate state and extension decreases accordingly while force remains approximately constant with each step (discrete state). A two state hairpin would potentially

fold from c' to b. a) A folding event to complete the folding if the first folding event was not a single step, as illustrated with the *add* riboswitch aptamer cartoon. (i.e. a folding could go c' to b to a, or skip b) At the end of the designated waiting period at low force, the sequence is repeated. (See also Error! Reference source not found.)

On our instrument we abruptly change the force in the trap by changing the laser intensity in only one of the traps, the one in which the bead will move into the zero-stiffness region. We change the beam power going into the trap by modifying the signal to the AOD. Because of the design of AODs, a diffraction grating created by a travelling sound wave (Foster, 2010), we can switch between two settings without going through the whole range separating them: an AOD provides random access for beam position (by changing the applied wavelength of the sound wave), as well as for the intensity of the deflected beam in our instrument (by changing the applied amplitude of that sound wave). The implementation is hardware specific, and is described in more detail in appendix A. The lengths of the linear region and zero stiffness regions require some estimation that is done by calibration (Foster, 2010). In the case of our OT instrument, the middle of the zero stiffness region occurs at ~ 170 nm from the centre of the lower trap (T0), with an overall length of about 50 nm. Beyond that distance, the derivative of the Gaussian intensity of the beam decreases, producing a region of negative stiffness, which we want to be mindful to avoid moving into because the extension changes will be measured as being longer – an overestimation. In our traps the linear region extends ~ 100 nm, see Figure 2.7.

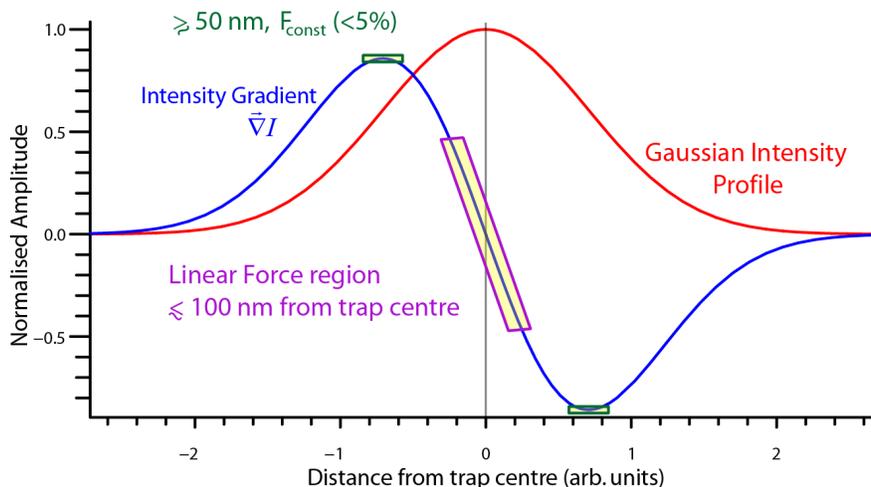


Figure 2.7 Force profile in an optical trap.

In the Rayleigh approximation, the force varies as the derivative of the intensity as a function of the displacement from the trap centre (Foster, 2010). For a Gaussian beam, a roughly linear region is observed near the trap centre (within ~50–100 nm). Farther out (~200 nm), there is a constant force region ~ 50 nm wide (Greenleaf et al., 2005). Further from the trap centre is a negative stiffness region, and then this simple presentation also breaks down, due to the truncation of the trapping beam at the objective lens.

There are some conditions under which maintaining correct positioning using constant force may be difficult, such as cases where the length of the zero-stiffness region would be a limiting factor, for example. In that case, it is then easier to operate the traps in a constant position mode, where we restrict bead movements to the linear stiffness regions of both traps. In this case, when we observe discrete steps in the change in extension after a force jump, they will be accompanied by changes in force as well (see Figure 2.8). On our instruments, force jumps in the linear trap regions can be implemented in two ways: by jumping the voltage to the EOD or jumping the frequency signal to the AOD. This type of jump is qualitatively similar to the implementation of a force-ramp, except the step is as abrupt as the instrument hardware allows (more detail in Appendix A).

In chapter 8, we use a force jump that goes from a high force then quickly down to a low force, in order to observe refolding in a sudden collapse situation (Kubelka et al., 2004), where we modulate the driving power to the AOD in a stepwise fashion as described above. The distances are pre-set in a way such that the molecule would unfold with the bead in the anharmonic region of the trap where the force is constant (more detail in appendix A).

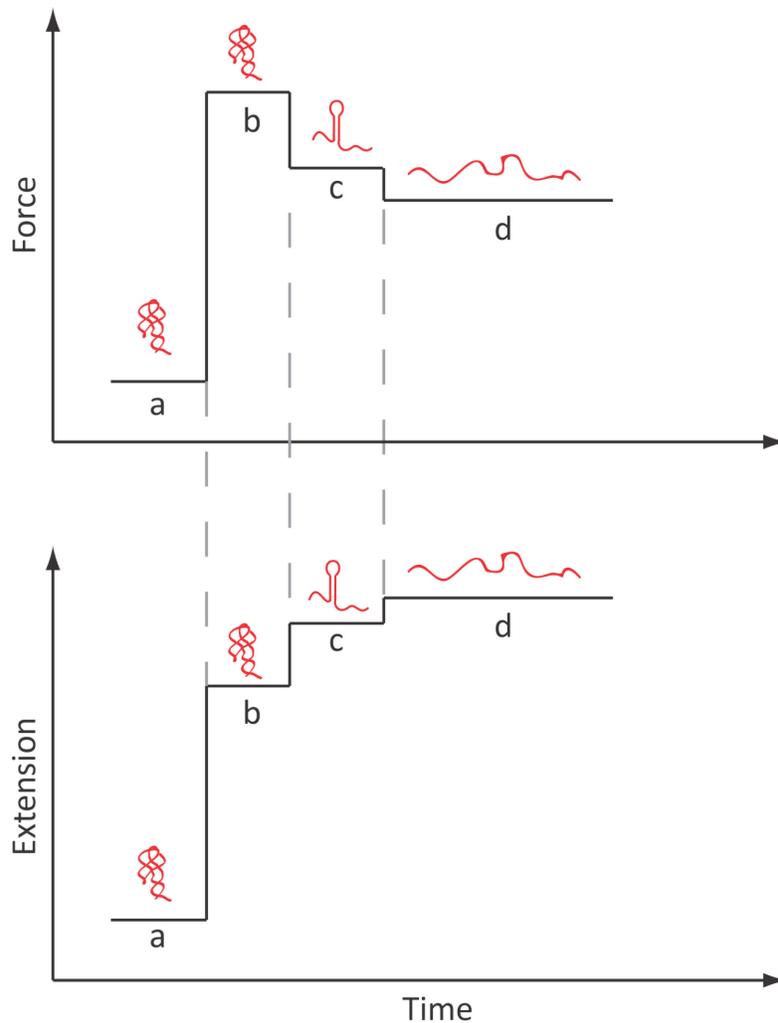


Figure 2.8 A schematic of a force jump experiment with constant trap positions

Configuration for a jump to high force to probe unfolding: a) Initially a waiting period at zero or low force, to assure enough time for the molecule to be folded, or a restricted time to explore the first folding steps, such as with the kinetic competition in the *add* riboswitch between the expression platform hairpin and the aptamer. b) The molecule is immediately jumped to a high force at a constant trap separation, c) the molecule unfolds and extension increases while force decreases as a result with each step (discrete state). d) A subsequent unfolding event to complete the unfolding if the first unfolding event was not a single step. (i.e. an unfolding could go b to c to d, or skip c) At the end of the designated waiting period at high force, the sequence is repeated.

Chapter 3: Single molecule force spectroscopy analysis

A wealth of biophysical detail can be extracted from single-molecule force spectroscopy (smFS) measurements, as was presented in the previous chapter. With it we can characterise features such as the number of intermediate states which are present, and the pathways followed between these states. We can also measure the sizes of the folded structures, the microscopic transition rates between states, and the energies of these states, as well as the positions and heights of the barriers between these states, culminating with the profile of the energy landscape. These features help us to elucidate molecular folding in detail. In this chapter the analysis of the measurement modalities presented in the previous chapter is outlined; the analysis methods are described in greater detail elsewhere (Dudko et al., 2008; Foster, 2010; Gupta et al., 2011; Hoffmann and Woodside, 2011; Hummer and Szabo, 2001; Stigler et al., 2011; Woodside and Valentine, 2009; Yu et al., 2012a).

3.1 Elastic properties of nucleic acids

When we record an FEC, we get a measure of extension as a function of force. To then relate this extension to intrinsic properties of the molecule under study, such as the contour length, we need models of the elastic behaviour. Fortunately, the characteristic, non-linear rise with extension as the dsDNA handles are stretched due to their elasticity has been studied already (Liphardt et al., 2001; Smith et al., 1996). This rise in force in the FEC continues until there is an abrupt extension increase and concomitant force decrease, the characteristic “sawtooth” pattern(s), when the structure, or substructures, between the two handles unfolds (Woodside et al., 2008). We can calculate the contributions of both the handles and the unfolded molecule. The elasticity of dsNAs, for instance, can be modeled by a variant of the worm-like chain (WLC) (Bustamante et al., 1994; Marko and Siggia, 1995; Smith et al., 1996). WLC models treat the polymer chain as a flexible rod with a given contour length (effectively, the length of the chain along the chemical backbone) and a characteristic persistence length (the distance over which the orientation of the chain is correlated). Previous work on nucleic acids has shown that large applied forces may stretch the chemical bonds, however, requiring the inclusion of an “enthalpic elasticity” to account for this stretching (Wang et al., 1997). ssRNA has also been found to be well-described by such a modified WLC model, though with quite different persistence length: ~ 1 nm for ssRNA vs. ~ 50 nm for dsDNA

(Bustamante et al., 1994; Seol et al., 2004). Combinations of dsDNA and ssDNA may thus be modeled as two WLCs in series. Fitting FECs with such a model then provides the change in contour length as structural elements of the molecule unfold (e.g. Figure 3.1, Figure 3.2a, Figure 4.1, Figure 5.3). As an additional benefit, because the persistence length of a single molecule of dsDNA is well known, and is different from the effective persistence length when more than one DNA molecule is present, fitting FECs to a WLC model also allows us to determine when we truly have a single molecule attachment, and discard samples with multiple handle attachments.

The exact solution of the WLC model is non-trivial, but a very effective interpolation formula has been developed by Marko and Siggia for use in fitting FECs (Marko 1995). Including the enthalpic elasticity of the polymer, this takes the form:

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

where L_p is the persistence length, L_c the contour length, and K the elastic modulus of the dsDNA or DNA/RNA duplex (describing the stretching of bonds under tension). This interpolation deviates by 10% around $F \approx 0.1$ pN, but becomes asymptotically exact in large and small force limits (Marko and Siggia, 1995; Wang et al., 1997).

The WLC parameters for different nucleic acids have been measured in a number of experiments. For dsDNA, $L_p = 40-50$ nm (depending on the ionic environment and on strand annealing) and $K \sim 1000-1200$ pN (Baumann et al., 2000; Bouchiat et al., 1999; Smith et al., 1996; Wang et al., 1997), whereas for ssRNA, $L_p \sim 1$ nm and $K \sim 1500-1600$ pN (Seol et al., 2004; Seol et al., 2007). For dsRNA, which generally takes on an A-form helix in contrast to the B-form helix of dsDNA (Neidle, 1999; Saenger, 1984), $L_p \sim 60$ nm (Abels et al., 2005); DNA-RNA hybrids, which also take on A-form helices, are expected to have a similar L_p . (In practice, we may measure a lower L_p , likely owing to incomplete or imperfect annealing between the DNA and RNA.) The contour lengths are also well known: for dsDNA, it is 0.34 nm/bp (the rise of the B-form helix), for dsRNA it is 0.29 nm/bp (the rise of the A-form helix), for ssRNA it is 0.59 nm/nt, and for ssDNA it is 0.6-0.7 nm/nt (Saenger, 1984).

3.2 Force extension curve analysis

As it happens, FECs provide information about many aspects of the folding. The changes in contour length during unfolding, ΔL_c , are found by fitting the folded and unfolded branches of the FECs (e.g. in Figure 3.1) one for the dsDNA handles, and a second for the unfolded ss RNA or DNA (Seol et al., 2004). Multiple FECs from a single molecule can be overlaid and then partitioned into the separate states of the molecule having different contour lengths, so that each branch of the FEC can be fit sequentially or globally. Because of instrumental drift that may occur over the course of measurements (up to several hours), individual FECs measured on the same molecule under the same pulling conditions are aligned vertically using the low-force ($F \sim 2\text{--}3$ pN) part of the data, where the FECs have very low slope, and horizontally using the high-force ($F \sim 20$ pN) part of the data, where the FECs have very high slope. Typical drift corrections are less than a few nm.

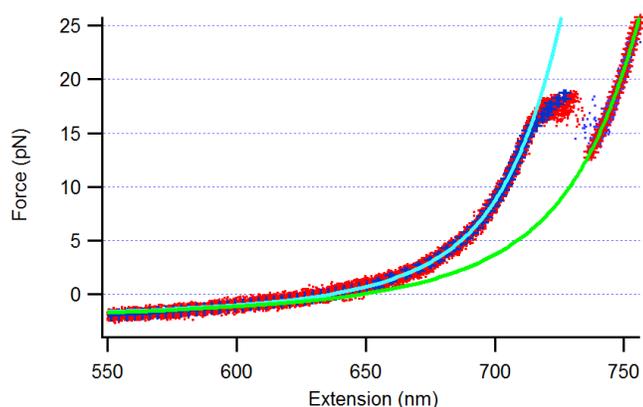


Figure 3.1 Example FEC overlay and WLC fits for hairpin 30R50/T4

In this example, 60 FECs are overlaid and binned, then the folded and unfolded portions (before and after the ‘rip’) are fit with the WLC (bright blue and green smooth lines) with the data longer than ~ 650 nm (to avoid curvature artefacts from the instrument contributing to the fits). Here a contour length difference of 36.7 nm is measured. With the addition of the helix width (2.0 nm for a B-form helix), and dividing by the contour length per nucleotide, the contour length from the fit equates to 65 nt. (The true value of the hairpin length is 64 nt).

To convert ΔL_c into the number of nucleotides unfolded, n_{nt} , we included the effects of changing the number of folded helices: $n_{nt} = (\Delta L_c + \Delta n_h d_h) / L_c^{nt}$, where Δn_h is the number of helices unfolded, $d_h = 2.2$ nm is the diameter in the case of A-form dsRNA helix (Saenger, 1984) or, when one DNA hairpin is unfolded, 2.0 nm (the width of one B-helix) is added to the fitted contour length change. For pseudoknots (PKs), the distance between the termini (d_t) takes the place of d_h . When possible, this distance is estimated from NMR or crystal structures or from the structure of a similar PK: d_t varies from ~ 2 –6 nm for those reported in chapters 5 and 6.

FECs also yield the work required to unfold the molecule, by integrating the curve (Figure 3.2b). The force is being changed rapidly, so the system is not in equilibrium, and therefore energy may be dissipated. Hence the work done will differ from the free energy change in the molecule. Moreover, the total work done on the handle-molecule construct includes the work done to stretch out the handles and unfolded NA, which must also be accounted for (Woodside et al., 2008). In order to account for the dissipated energy, the Jarzynski equality (Jarzynski, 1997) may be used to determine the equilibrium free energy from the distribution of non-equilibrium work measurements:

$$\Delta G_{eqm} = -k_B T \ln \left\langle \exp \left(-\frac{W}{k_B T} \right) \right\rangle \quad (3.2)$$

Here the equilibrium free energy, ΔG_{eqm} is found from an exponentially-weighted average of the measured non-equilibrium works, W (Figure 3.2c). A closely-related formulation is found in Crooks' theorem (Crooks, 1999), which relates the work distributions for the unfolding and refolding reactions, $P_U(W)$ and $P_R(-W)$ respectively, to the energy dissipated (Figure 3.2c):

$$\frac{P_U(W)}{P_R(-W)} = \exp \left(\frac{W - \Delta G_{eqm}}{k_B T} \right) \quad (3.3)$$

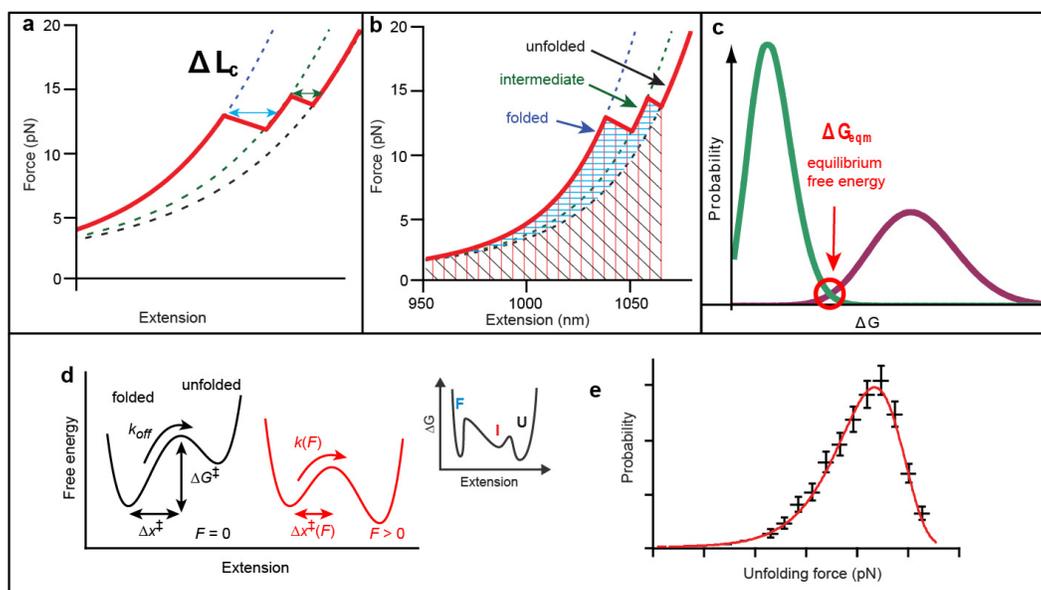


Figure 3.2 FEC analysis.

(a) FEC cartoon showing changes in the contour length ΔL_c . (b) To determine the work done to unfold the molecule, the FEC is integrated numerically out to the final unfolding event (red area), and the integral of the unfolded construct (black area) is subtracted (blue area). (c) Equilibrium free energy can be determined from the distribution of non-equilibrium work. Here, Crooks' theorem is used to determine the equilibrium free energy from the intersection of the unfolding (purple) and refolding (green) distributions. (d) Folding can be modeled as motion through an energy landscape. A landscape with a single barrier is characterised by Δx^\ddagger , the location of the barrier along the reaction coordinate, and ΔG^\ddagger , the height of the barrier. Force tilts the landscape as shown (red). A possible three-state landscape is shown in the inset. (e) The distribution of unfolding forces follows a characteristic distribution from which kinetics and landscape parameters may be extracted.

FEC measurements also contain kinetic information on the unfolding transitions, but again this is not straightforward to extract, due to the fact that the force is constantly changing. To illustrate this, the context of an energy landscape diagram is helpful (Figure 3.2d). A molecule with two states (folded and unfolded) has a folding landscape consisting of two potential wells separated by a barrier at a distance Δx_f^\ddagger from the folded state (minimum of the first well) and Δx_u^\ddagger from the unfolded state (minimum of the second well). As the tension on the molecule increases, the work done on the molecule by the optical trap tilts the landscape, lowering the relative height of the barrier and moving it with respect to the folded and unfolded states. During FEC measurements, the barrier height is continuously lowered as the force is raised, thereby increasing the probability of unfolding relative to folding (and

hence increasing unfolding rate and decreasing the folding rate), until at some point the molecule unfolds.

When unfolding intermediates exist, this simple two-state picture cannot generally be directly modified to include additional barriers and potential wells (Figure 3.2d inset), as the dynamics become more complex. The distribution of unfolding forces, which reflects the cumulative survival probability in the folded state, does contain information about the unfolding rate as a function of force (Garai et al., 2014; Zhang and Dudko, 2013). In a 2008 publication, Dudko *et al.* showed that the lifetimes in terms of the rupture-force distribution and loading rate from FEC experiments: $\tau(F) = \int_F^\infty p(F') dF' / \dot{F}(F) p(F)$, where $\dot{F} \equiv dF/dt$ is the loading rate. Now the lifetimes from the constant force and force-ramp measurements can be unified (Dudko et al., 2008). Dudko *et al.* go on to show that for the force dependent lifetimes it can be written: $\tau(F) = \tau_0 \exp - \beta \int_0^F \langle x^\ddagger(F') \rangle dF'$, where $\langle x^\ddagger(F) \rangle$ denotes a difference in the average positions of the transition state and the bound state, so that a plot of the natural logarithm of the lifetimes as a function of force will give a direct measure of that quantity (Dudko et al., 2008). This approach also works well for relaxation curves (refolding), and it has been generalised for a multi-state system (Pierse and Dudko, 2013; Zhang and Dudko, 2013).

3.3 Constant force data analysis

Constant force measurements yield the same quantities that can be obtained from FEC measurements, but with simpler analysis because the system is in equilibrium throughout (Woodside and Valentine, 2009). Extension changes can be read directly from the data (for example, using histograms of the trajectory to identify the peaks corresponding to different states (such as Figure 4.2, Figure 4.3, Figure 4.8)) and converted to contour length changes using a WLC model as before (equation 3.1) (Foster, 2010). We can read the free energy from an extension change multiplied by the force (but with a subtraction of the energy required to stretch the unfolded NAs, for comparison to experiments at zero force). Moreover, the lifetime in each state can be measured directly from the trajectory (see for example Figure 4.1D), allowing the detailed distribution of lifetimes to be measured. By comparison, the histogram transformation used with FECs provides an average rate, as opposed to a lifetime distribution.

To be brief, to measure the rates, we apply thresholds of different extensions (such as Figure 4.1D, Figure 4.2C) to identify the states in the constant force extension traces. This can also be done using hidden Markov models, e.g. (Stigler et al., 2011), or with correlation analysis (Hoffmann and Woodside, 2011).

3.4 Application of Kramers' theory of reaction rates

As has been discussed previously, folding is described in terms of a diffusive search over the configurational free energy surface for the minimum-energy structure and is often reduced to a 1D profile along an appropriate reaction coordinate (Oliveberg and Wolynes, 2005). Theories of diffusion from chemical physics can therefore be applied here as a framework to relate the kinetics of folding to the thermodynamics of the energy landscape. In particular, the theory of diffusive barrier crossing due to Kramers (Hänggi et al., 1990; Kramers, 1940) is commonly applied to biomolecular folding. If the temperature dependence of the folding rates follows the Arrhenius form $k = k_0 \cdot \exp(-\Delta G^\ddagger / k_B T)$, then we can apply Kramers' theory, since it encompasses this rate description. In his celebrated reaction rate theory paper of 1940 (Kramers, 1940)(and following the work of others), Kramers described the barrier crossing mechanism in solution as governed by Brownian motion (Hänggi et al., 1990). The theory describes diffusion across a barrier of height ΔG^\ddagger , where the barrier represents the bottleneck formed by the ensemble of transition states. A transition state corresponds to the highest energy along the reaction coordinate, or in other words, the position at the top of a barrier.

The folding/unfolding rate can be determined from the free-energy profile by Kramers theory (Hänggi et al., 1990):

$$k = k_0 \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right), \text{ where } k_0 = \frac{\sqrt{\kappa_w \kappa_b}}{2\pi k_B T} D, \quad (3.4)$$

ΔG^\ddagger dominates the kinetics, D is the diffusion constant for barrier crossing, κ_w is the stiffness (curvature) of the potential well, and κ_b is the stiffness of the barrier. Here Kramers has assumed that we can approximate both D as a constant over the barrier region, and the potentials as being harmonic. Specifically, we can write out two forms for the rate in each direction:

$$k_{folding} = \frac{\beta D}{2\pi} \sqrt{\kappa_u \kappa_b} \exp(-\beta \Delta G_u^\ddagger) \quad \text{and} \quad k_{unfolding} = \frac{\beta D}{2\pi} \sqrt{\kappa_{well(folded)} \kappa_b} \exp(-\beta \Delta G_f^\ddagger),$$

and $\beta = 1/k_B T$ is the inverse thermal energy (see Figure 3.3). In the physical picture, the exponential term in the above equations relates to the fact that most of the time is spent waiting (apart from the folding/unfolding transitions ($\tau = 1/k$) for a thermal fluctuation that has enough energy to kick the folding reaction over the barrier. During this waiting time, the molecule diffuses within the potential well. The prefactor, k_0 , describes the “attempt frequency” for crossing, which is higher for stiffer potentials (proportional to the square root of the curvature, and analogous to their natural resonant frequencies), and gets larger as the diffusion gets faster. Determining the barrier height and curvatures require knowing the shape of the energy landscape, while D is generally not known. Given that the diffusion is a more fundamental physical feature of the folding, this means that the problem is typically reversed, and measured rates must then be used deduce D .

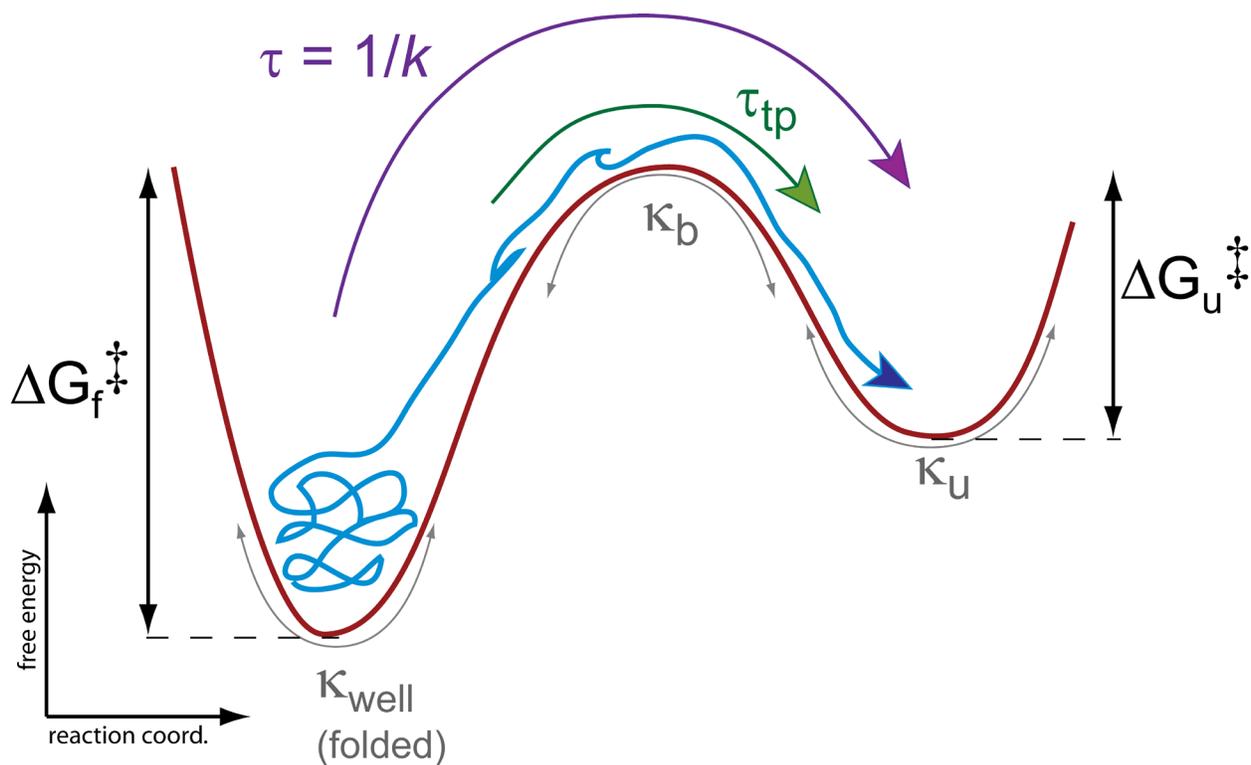


Figure 3.3 Energy landscape diagram for a two state system.

κ represents the curvature of a well or barrier, τ represents times (inverse of rates, k) and ΔG^\ddagger are barrier heights. The meandering folding path is in blue, and the actual transition time τ_{tp} over the barrier is in green.

Although the curvature of the wells and barriers of a given energy landscape is difficult to measure and generally not known, we can estimate it from the typical shift in the position of the transition state due to the applied force for a molecule whose landscape has been measured (Woodside et al., 2006a; Woodside et al., 2006b): the DNA hairpin 30R50/T4 has a barrier with curvature of ~ 2 pN/nm. The actual transition time over the energy barrier, τ_{tp} is generally much faster than the lifetime in the states on either side of the barrier. The curvature of the wells and barriers are found from a quadratic fit to the landscape profiles, and barrier heights are measured from the profiles for both folding and unfolding. D and τ_{tp} are then calculated for both folding and unfolding. For all molecules, when the values for the two directions are equal within error, then the average values are used (see chapter 6). When quantitatively measuring the transit times, assuming again, 1D diffusive motion over a harmonic barrier, τ_{tp} is related to, D , in terms of the properties of the barrier by:

$$\tau_{tp} \approx \frac{\ln(2e^\gamma \beta \Delta G^\ddagger)}{\beta D \kappa_b} = \frac{\ln(2e^\gamma \beta \Delta G^\ddagger)}{2\pi k_0 \sqrt{\kappa_b / \kappa_w}}, \quad (3.5)$$

where $\gamma \approx 0.577$ is the Euler–Mascheroni constant (Chaudhury and Makarov, 2010; Chung et al., 2009). The RHS of the equation is produced by rearranging κ_b from equation 3.4. We can note that for the transition time, it varies inversely with the diffusion coefficient (specifically the diffusion coefficient in the barrier region), but only logarithmically with the barrier height; (Chaudhury and Makarov, 2010; Chung et al., 2009; Chung et al., 2012). Intuitively, this is what we would expect (for $\tau_{tp} \sim \alpha^{-1} D$), we can make an analogy to parabolic trajectories, where the time required to return to ground state does not depend so much on the height of the trajectory, since to reach higher you have to start and end with higher speeds. (Especially where D relates to the relative end-to-end distance (Schuler et al., 2002)). The implementation of the barrier crossing theories is discussed further in chapter 6.

As was stated above, we have made two approximations: a constant diffusion, and harmonic potentials. There is much evidence that shows that D is not actually constant (Best and Hummer, 2010; Chahine et al., 2007; Oliveira et al., 2010a; Oliveira et al., 2010b; Xu et al., 2012) but depends on the nature of the reaction coordinate. Furthermore, to what degree that D is position dependent is contested; some computational studies have found relatively small changes, on the order of factors of 2, e.g. (Best

and Hummer, 2006), whereas others have found larger effects, closer to 5 to 10-fold or more (e.g. (Chahine et al., 2007; Xu et al., 2012; Yang et al., 2006). Recent work has also demonstrated that it is possible to transform coordinates such that D is position-independent (Best and Hummer, 2011). Empirically, much work has demonstrated Kramers' theory to be consistent with experimental data over a wide range of conditions, for example with the hamster prion protein with the predictions of Kramers' theory of 1D diffusion with constant D (Yu et al., 2012a). Given the success of Kramers' theory, most experimental data is analysed in the context his framework (Borgia et al., 2012; Cellmer et al., 2008; Chung and Eaton, 2013; Gebhardt et al., 2010; Heidarsson et al., 2014; Jagannathan et al., 2012; Popa et al., 2011; Stigler et al., 2011; Wensley et al., 2010).

3.5 Energy landscape reconstructions

There are multiple ways to reconstruct the properties of the energy landscape from the different smFS measurement modalities, as reviewed elsewhere: (Woodside and Block, 2014). The primary methods applied in this thesis are based on considerations of how the force applied to the molecule changes the lifetime (and hence rate) of the structure. The simplest approach is the Bell-Zhurkov model, which makes no assumptions about the shape of the landscape, except that the transition state position (Δx^\ddagger) is independent of force, and offers the force dependence of the lifetimes: $\tau_{Bell}(F) = \tau_0 \exp(-\beta F \Delta x^\ddagger)$. (Here τ_0 is the intrinsic lifetime in the absence of applied force). The Bell-Zhurkov relation can equivalently written as: $k_{Bell}(F) = k_0 \exp(+\beta F \Delta x^\ddagger)$, with the rate at zero force, k_0 , and rate under load, $k(F)$. Here k_0 is just the Kramers' rate expression that was shown previously in equation 3.4. The effect of the force as it tilts the landscape is simply to change the height of the barrier, leading to an exponential rate change.

For a simple two-state transition, following Bell's interpretation (Bell, 1978), the logarithm of the rate of unfolding $k_u(F)$ (or folding $k_f(F)$) is linearly related to the force, with a slope proportional to Δx^\ddagger , the distance to the transition state (subscripts u and f representing unfolding and folding, respectively, while k^0 is the rate at zero force):

$$\ln \frac{k_u(F)}{k_u^0} = \beta \Delta x_u^\ddagger F \quad \text{and} \quad \ln \frac{k_f(F)}{k_f^0} = -\beta \Delta x_f^\ddagger F \quad (3.6)$$

The logarithms of the rates vary linearly with force, indicating that the transition state position is effectively constant over the force range probed here. Although the transition state location is generally force-dependent, it can often be approximated as force-independent for constant force measurements in which the force is not varied much. In such measurements we would expect the position of the transition state to shift by an amount of order $\sim \Delta F / \kappa$, the ratio of the range of forces probed to the curvature of the potential barrier (Foster, 2010).

Although the Bell-Zhurkov model does reasonably well as an approximation to fit for the rates over the narrow range of forces that are typically probed in constant-force measurements, it retains the unphysical assumption that the barrier distance, Δx^\ddagger , does not change with force. As a consequence it systematically overestimated the unfolding rate (Dudko et al., 2006). A more sophisticated treatment of the problem has been offered by Dudko *et al.* (Dudko et al., 2006, 2008), who derived an analytic expression for the force-dependent rate using Kramers' theory under the assumption of particular shapes for the energy profile:

$$k(F) = k_{\text{off}} \left(1 - \frac{\Delta x^\ddagger F}{\Delta G^\ddagger} \nu \right)^{1/\nu - 1} \exp \left\{ \beta \Delta G^\ddagger \left[1 - \left(1 - \frac{\Delta x^\ddagger F}{\Delta G^\ddagger} \nu \right)^{1/\nu} \right] \right\} \quad (3.7)$$

k_{off} is the unfolding rate at zero force, r is the loading rate, k_B is Boltzmann's constant, and ν parameterizes the shape of the energy barrier ($\nu = 1/2$ for a cusp-like barrier, $2/3$ for a linear-cubic potential). This expression extends a previous treatment (Evans and Ritchie, 1997) that effectively had $\nu = 1$, which recovers Bell's formula from the above equation ($k = k_0 \cdot \exp(\beta F \Delta x^\ddagger)$). This new form also depends on the barrier height, ΔG^\ddagger , which did not contribute in the previous model. The Dudko theory also relates the force-dependent rates to the shape of the unfolding force probability distribution (Figure 3.2d), $p(F)$:

$$p(F) \propto \frac{k(F)}{r} \exp \left\{ \frac{k_{\text{off}}}{\Delta x^\ddagger r} - \frac{k(F)}{\Delta x^\ddagger r} \left(1 - \frac{\Delta x^\ddagger F}{\Delta G^\ddagger} \nu \right)^{1-\frac{1}{\nu}} \right\} \quad (3.8)$$

where $k(F)$ is given by the previous equation (3.7). This form for $p(F)$ can therefore be used to extract landscape parameters by fitting histograms of unfolding forces (Dudko et al., 2006). We average the results obtained under the two limiting cases for the shape of the energy barrier since this shape is unknown. For a given set of FECs all at the same loading rate, these equations are used to fit $p(F)$, (as in Figure 3.2e), and extract the parameters describing the kinetics and the shape of the energy landscape. Of particular interest in this analysis is the parameter Δx^\ddagger , the distance to the transition state. This distance can be converted into the number of nucleotides associated with the transition state structure, giving a clue as to how the reaction proceeds. Such information is very difficult to obtain by other means.

The two approaches (rates or unfolding force distributions) are actually complementary, considering histograms can be converted into rates via the cumulative survival probability or rupture times, and since these unfolding states are stochastic and irreversible processes out of equilibrium (Dudko et al., 2008). A similar approach can be used to determine landscape parameters from refolding force distributions and rates (Pierse and Dudko, 2013). Even though these methods are approximate, they provide us with determinations of parameters we are most interested in, namely the height and position of the barrier. The barrier height is an energy threshold that the folding molecule must have, while its position provides clues as to the nature of the transition state, such as the number of nt pertinent to its structure.

The methods described above are very useful and powerful, but they depend on certain assumptions about the shape of the landscape, or else only provide incomplete information about the landscape. The full shape of the profile can indeed be reconstructed from several methods. Some use theories from non-equilibrium statistical mechanics (Hummer and Szabo, 2001, 2010) to recover the equilibrium free-energy profile from non-equilibrium FECs (Engel et al., 2014; Gupta et al., 2011). In this

thesis I use a simpler approach, the inverse Boltzmann transform, based on equilibrium thermodynamics: given the equilibrium extension distribution $P(x)$ from a constant-force measurement, the free-energy profile is simply: $G(x) = -\ln(\rho(x))/\beta$, where, $\rho(x)$ represents the probability distribution, typically from a normalised extension histogram.

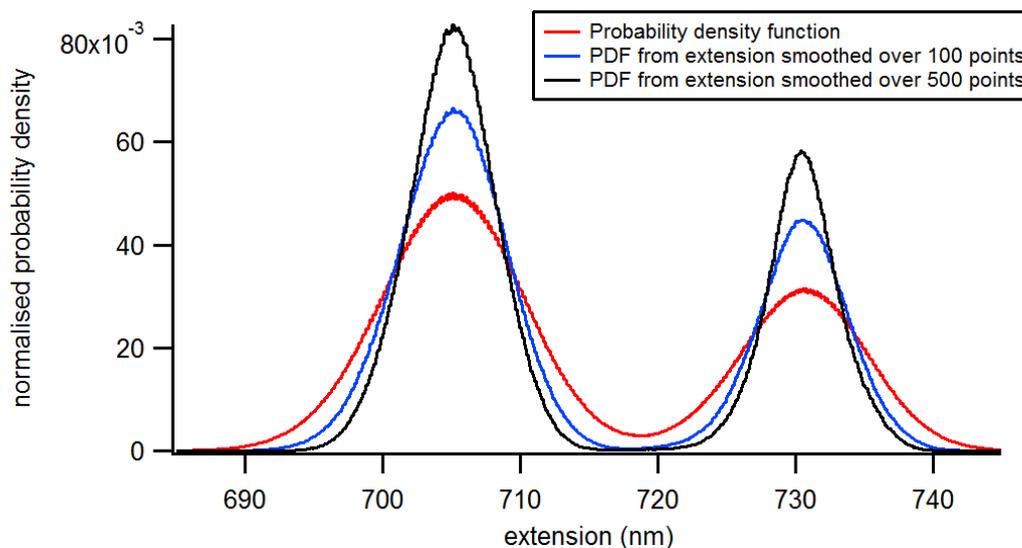


Figure 3.4 Example of a probability density for HP 30R50/T4 from constant force data.

Red: unsmoothed data, black, blue: median smoothed at 500 and 100 points (1953 and 390 μ s, data sampled at 256 kHz). Constant trap separation would narrow the distance between the unfolded and folded peaks compared to equivalent constant force data.

The compliance of the traps and handles broadens the distribution, (see above Figure 3.4). Moreover, with constant trap separation traces, the two state peaks are also pushed closer together in addition to the broadening, owing to compliance effects (Greenleaf et al., 2005), but if we were to pass a median smoothing filter over the extension trace before producing a probability density plot, we can see more closely where the individual states lie. To be done correctly, this effect can be removed using deconvolution methods as described elsewhere (Woodside et al., 2006a; Woodside et al., 2006b).

The results of free-energy profile reconstructions using methods such as the inverse Boltzmann-transform can be checked for consistency by testing specific points on the landscape, such as the position

and height of the barrier (e.g. obtained from the theories of Dudko et al.), or the free-energy change between states. The latter can be determined in several ways. From FECs, ΔG can be found by applying Jarzynski's equality to the distribution of non-equilibrium work done (Jarzynski, 1997; Liphardt et al., 2002). More straightforwardly, from constant force measurements, which record equilibrium fluctuations, the folding free energy can be found without the use of fluctuation theorems, simply from basic thermodynamics: ΔG can be found from the product of the extension change between states (Δx) and the force at which the two states are equally occupied ($F_{1/2}$). Because the free-energy difference under tension includes the energy required to stretch the unfolded nucleotides, the free energy at zero force is given by: $\Delta G^0 = F_{1/2} \cdot \Delta x - G_{stretch\ of\ nt}$ (where $G_{stretch\ of\ nt}$ can be found for example by integrating a WLC model for the unfolded nucleotides).

Lastly, some additional novel methods to reconstruct and characterise energy landscapes are explored in greater detail in later chapters (7 and 8), with a focus on force jumps in chapter 8.

Chapter 4: The *add* adenine riboswitch's folding and regulatory mechanism

The previous chapters outlined the experimental and analytical tools available for studying folding at the single-molecule level with force spectroscopy. In this chapter, I describe their application to a type of RNA for which the conformational dynamics plays an essential role in its function: a riboswitch, specifically the *add* riboswitch from *V. vulnificus*. We unfolded and refolded the entire riboswitch, as well as each of its structural domains, allowing us to relate folding to function. Its study revealed multiple partially-folded states, including several misfolded states not part of the native folding landscape. This chapter is based on the work published as: Krishna P. Neupane, Hao Yu, Daniel A. N. Foster, Feng Wang and Michael T. Woodside “Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism” *Nucleic Acids Res* 39:7677-7687 (2011). In this work, FW and DANF prepared the samples; KN, HY, and DANF performed the measurements; KN, HY, DANF, and MTW analysed the data.

4.1 Introduction to riboswitches

Riboswitches are structured and untranslated parts of messenger RNA (mRNA) (in either its 5' or 3' end) with the ability to sense a small metabolite and modify the activity of its cognate gene frequently as a feedback loop (Coppins et al., 2007b; Roth and Breaker, 2009a; Winkler and Breaker, 2005). Found in a wide range of organisms, including archaea, plants, fungi, and algae (Breaker, 2012), but most notably in prokaryotes, and as such they offer new potential targets for antibiotics, e.g. (Blount et al., 2015; Lunse et al., 2014; Matzner and Mayer, 2015; Yu and Olsthoorn, 2015). (Not to be confused with the other medical RNA application riboSNitch (Solem et al., 2015).) Riboswitches are often bipartite in structure, with a ligand-sensing aptamer and an expression platform to modulate gene activity. Ligand binding, with very sensitive target discrimination (with a dissociation constant (K_D) down to the nM range), stabilizes the aptamer domain through changes in secondary and/or tertiary interactions, typically switching the structure of the expression platform between two alternative conformations and thereby altering transcription, translation, self-cleavage, splicing, or other processes during expression (Baird and Ferré-D'Amaré, 2010; Coppins et al., 2007a; Grigg and Ke, 2015; Roth and Breaker, 2009b).

Riboswitches typically fold into two functional conformations, each associated with one of the two states for the gene, “on” or “off”. As such, their folding landscapes feature two lowest energy states for either the “on” or the “off” (especially for those regulated thermodynamically, but not true for kinetically controlled riboswitches whose lowest energy state is not changed by ligand binding). Generally speaking, the energetic end-point for a folding trajectory is modified by the binding of a metabolite in the aptamer (Savinov et al., 2014; Woodside and Block, 2014) (this does not necessarily apply to temperature sensing riboswitches, for example (Reining et al., 2013)). We can reconstruct an energy landscape for the riboswitch, on which is projected its hierarchical folding, intermediate states, and in the case of the *add* aptamer, a misfolding pathway (see page 49). The purine riboswitch aptamer structures often feature several hairpins, whose foldings are in cooperation and competition. Studying riboswitches allows us to further understand folding in the context of regulatory RNAs and how riboswitch function depends on its folding.

The *add* adenine-binding riboswitch, from *vibrio vulnificus*, is an example of the purine riboswitch class, which is among the smallest and structurally simplest riboswitches. As they are amongst the structurally simpler examples, are a good starting point for understanding riboswitch function and yet collectively they still demonstrate much of the variety of mechanisms used by those more complex RNAs (Kim and Breaker, 2008). For example, the *xpt* guanine riboswitch controls expression via transcription termination (Mandal et al., 2003), the *pbuE* adenine riboswitch through transcription anti-termination (Mandal and Breaker, 2004), and the *add* adenine riboswitch through translation activation (Serganov et al., 2004). Despite such different mechanisms, the aptamers of these riboswitches all have very similar structures: a “tuning-fork” architecture wherein the purine binding site is located at a specific residue in a pocket formed at the junction of three helices, two of which are hairpins interacting via kissing loops and aligned on top of the third helix (Batey et al., 2004; Serganov et al., 2004). Comparative studies of these riboswitches thus provide an opportunity to investigate the molecular features involved in their different regulatory mechanisms.

Previous studies have investigated various different features of the purine riboswitch aptamers, such as ligand specificity (Mandal and Breaker, 2004; Noeske et al., 2005) and its structural basis (Gilbert et al., 2006b; Noeske et al., 2005; Rieder et al., 2007), the rates and energies for ligand binding and

dissociation (Rieder et al., 2007; Wickiser et al., 2005a), the kinetics of the loop-loop interaction (Lemay et al., 2006), the Mg^{2+} -dependence of the folding (Buck et al., 2010), the folding energy landscape (Greenleaf et al., 2008), and the structural changes and induced fit upon ligand binding (Gilbert et al., 2006b; Noeske et al., 2007; Ottink et al., 2007; Stoddard et al., 2008){Stoddard, 2008 #207}. The relation of aptamer folding to regulatory mechanism has been less well studied, however, whether in purine riboswitches or more generally. The *pbuE* riboswitch has been shown, along with the FMN riboswitch, to function via kinetically-controlled folding of the aptamer in competition with ligand binding and transcription (Greenleaf et al., 2008; Lemay et al., 2006; Wickiser et al., 2005b), while the folding kinetics of the preQ₁ riboswitch aptamer suggest that transcription regulation is achieved instead by ligand-induced thermodynamic changes (Rieder et al., 2010). All of these riboswitches operate via transcriptional regulation; in contrast, there was little research at the time of our publication (2007 through 2011) on the folding of riboswitches that function through translation, such as the *add* riboswitch (Lemay et al., 2011; Rieder et al., 2007). (Recently, for example, the folding of the lysine translational riboswitch has been studied (Caron et al., 2012).) Additionally, little had been done more generally to characterize the interaction between an aptamer and its corresponding expression platform, a key question for understanding and manipulating riboswitch function (Zhang et al., 2010). (Single-molecule techniques have since been applied to TPP, various SAM, various PreQ₁, cyclic-di-GMP and lysine riboswitches (Savinov et al., 2014).)

Using smFS, we first built an integrated picture of folding in the aptamer alone, similar to previous force spectroscopy work on the closely-related *pbuE* adenine riboswitch aptamer (Foster, 2010; Greenleaf et al., 2008). Since these riboswitches have very similar aptamer structures but work by completely different regulatory mechanisms—translation activation (*add*) rather than transcription anti-termination (*pbuE*)—we can begin to discern which aspects of the folding arise from shared structural features and which may be important for the different regulatory mechanisms. We then extended these measurements to the complete riboswitch, probing the interaction between aptamer and expression platform and thereby demonstrating that the regulation is controlled by the folding thermodynamics.

4.2 Folding intermediates of the aptamer

RNA constructs were created by inserting the sequence for the *add* aptamer alone, expression platform alone, or full-length riboswitch into the pMLuc-1 plasmid between the SpeI and BamHI restriction sites. (Sequences are listed in Appendix table D.1) The resulting transcription template containing the riboswitch sequence flanked on each side by kb-long “handle” sequences was transcribed *in vitro* using T7 RNA polymerase. They were then annealed to single-stranded (ss) DNA complementary to the handles, and attached to beads held in optical traps (Figure 2.1), as described previously (Neupane et al., 2011). The handles were 842nt at the 3' end and 1289nt 5' at the end. The handles were produced by asymmetric PCR from double-stranded DNA PCR products corresponding to the flanking handle sequences (Saiki et al., 1986). Dumbbells were placed in measuring buffer (50 mM MOPS, pH 7.0, 130 mM KCl, 4 mM MgCl₂, ≥50 U/mL Superase-In RNase inhibitor (Ambion) and oxygen scavenging system: ≥40 U/ml glucose oxidase, ≥185 U/ml catalase, and 8.3 mg/ml D-glucose) and inserted into a sample chamber on a clean microscope slide in the optical trap. Force-extension curves (FECs) were sampled at 20 kHz, otherwise for constant force measurements 50 kHz. Multiple FECs were aligned to correct for slow drift of a few nm over several hours. Constant force data were median filtered offline in a 1–2.5 ms window, and states with different extensions were separated and identified using a modified thresholding algorithm (Woodside et al., 2006b).

First the aptamer alone was measured, using force-extension curves and trajectories at constant force to identify states in the folding pathway. The optical trapping assay is illustrated schematically in Figure 4.1A. Three FECs measured in the absence of adenine (Figure 4.1B: black, blue, and red lines) displayed typical features, and four states were seen with different contour lengths. The change in the contour length associated with each unfolding transition, ΔL_c , was found by fitting the aggregated FEC data from hundreds of pulls on the same molecule (Figure 4.1B: grey dots). The average values of ΔL_c determined from more than 5,200 FECs measured on 4 different molecules were: $\Delta L_c = 39.6 \pm 0.5$ nm from the fully folded state (labeled “F”, purple line) to the fully unfolded state (labeled “U”, yellow line), $\Delta L_c = 19 \pm 1$ nm from F to the shorter intermediate (labeled “P2P3”, cyan line), and $\Delta L_c = 30.3 \pm 0.5$ nm from F to the longer intermediate (labeled “P3”, green line). All uncertainties were reported as the standard error.

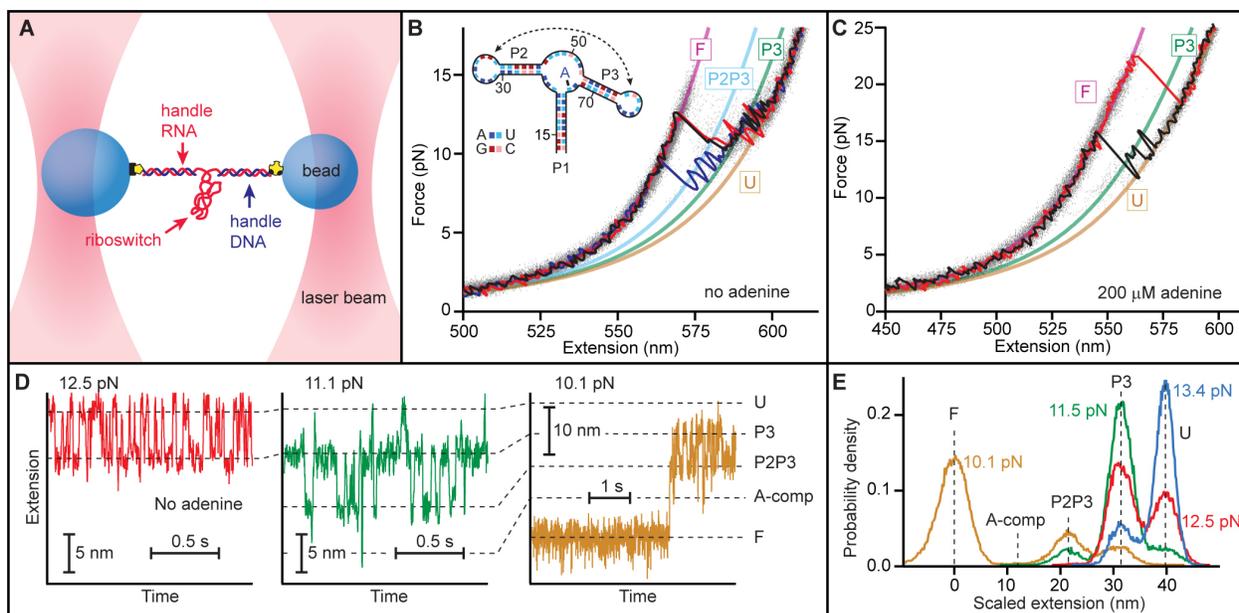


Figure 4.1 Force spectroscopy of add aptamer alone.

(A) RNA containing the riboswitch aptamer flanked by two kilobase-long handle sequences was annealed to DNA strands complementary to the handles and attached to beads held in optical traps. (B) Force-extension curves (FECs) in the absence of adenine revealed 2 intermediate states corresponding to the unfolding of everything except hairpins P2 and P3, followed by hairpin P2. Three FECs (black, blue, red) are plotted above the aggregated data from 700 FECs (grey dots). Worm-like chain fits are shown for the four states: F—fully-folded (purple); U—fully-unfolded (yellow); P3—P3 folded (green); and P2P3—both P2 and P3 folded (cyan). (C) Adenine binding resulted in similar behaviour but higher unfolding forces. State P2P3 was observed less frequently if at all. (D) The extension as a function of time at different levels of constant force in the absence of adenine revealed 5 distinct states corresponding to the major structural features: fully-unfolded at the largest extension (“U”), then P3 folded (“P3”), then both P2 and P3 folded but no loop-loop interaction (“P2P3”), then the P2-P3 kissing loop complex (labeled “A-comp:” competent to bind adenine), and finally fully-folded (“F”). Only short segments of the full records are displayed. (E) Histograms of the full extension records show force-dependent occupancies of the states. The A-comp state always had low occupancy.

We related these states to the unfolding of distinct structural elements in the aptamer (Figure 4.1B, inset) via the ΔL_c values expected from these structures: $\Delta L_c = N_{nt} \cdot L_c^{nt} - \Delta n_h \cdot d_h$, where N_{nt} is the number of nucleotides unfolded, $L_c^{nt} = 0.59$ nm is the contour length/nt for ssRNA (Saenger, 1984), Δn_h is the number of RNA helices removed during unfolding, and $d_h = 2.2$ nm is the diameter of the A-form dsRNA helix (Saenger, 1984). The ΔL_c values expected for different structures are listed in Table 4.1, alongside the measured values. We found excellent agreement with the expected value for full unfolding

(F to U), confirming that the aptamer was being completely unfolded. P1 unfolding by itself was not observed, but the intermediate labeled P2P3 matched the result expected for unfolding P1 and the loop-loop interaction (*i.e.* everything but P2 and P3). The last intermediate (labeled P3) was in good agreement with the expectation for unfolding everything except P3. However, since hairpins P2 and P3 differ by only 2 nt in length, additional confirmation for this state was obtained using an anti-sense DNA oligomer complementary to the 5' stem and loop of hairpin P2 to block P2 folding and loop and 3' stem of hairpin P3 to block P3 folding (Figure 4.2A). Notably, whereas state F never reformed once it was unfolded in a given FEC, multiple transitions between states P2P3, P3, and U were often observed (Figure 4.1B), indicating much faster kinetics for hairpins P2 and P3 than for helix P1 and the junction.

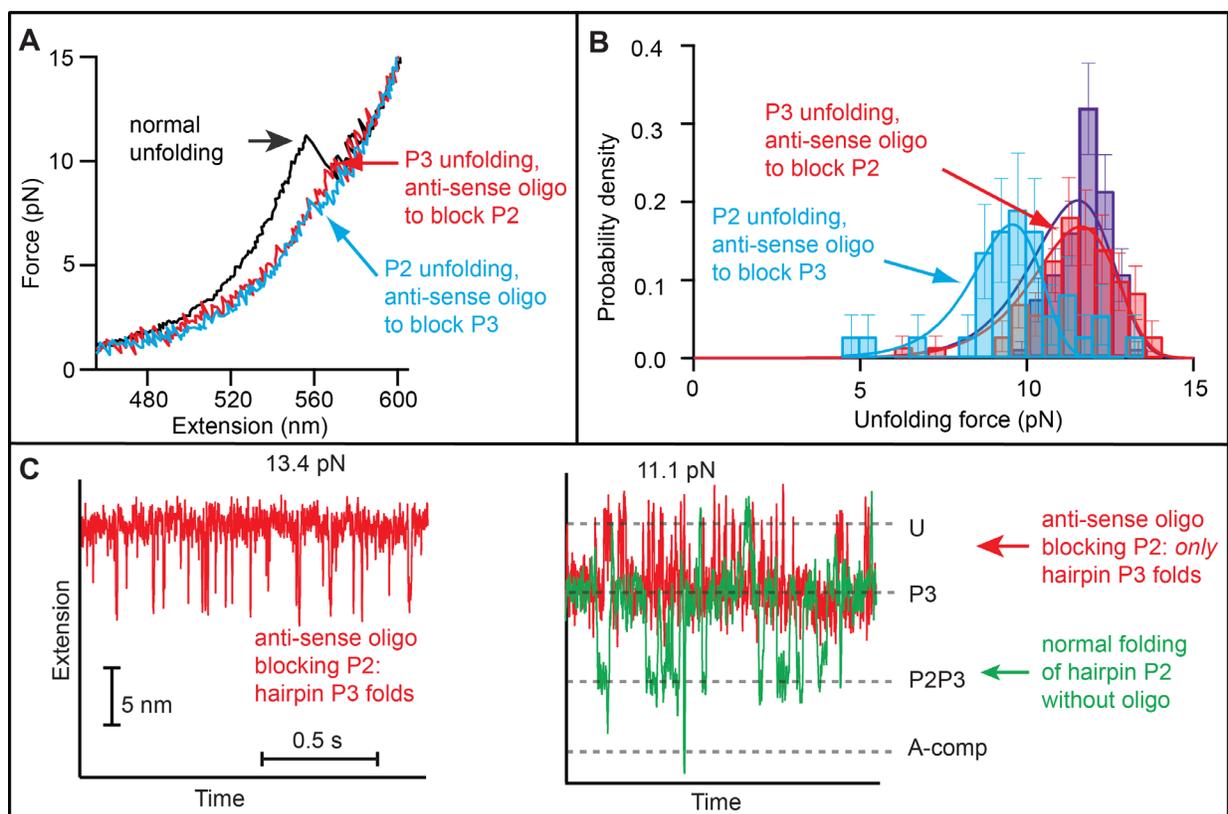


Figure 4.2 Confirmation of hairpin folding identification using anti-sense oligos.

(A) FECs under normal conditions (black) and with a DNA oligomer bound to the loop and 5' stem of hairpin P2 (red and with a DNA oligomer preventing formation of hairpin P3 (blue). The high-force unfolding transition under normal conditions matched the unfolding when P2 is blocked, when the only part of the aptamer that can fold is P3. (B) The unfolding force distribution for P2 (when P3 was blocked, blue) differed from the distribution for P3 (when P2 was blocked, red). The distribution of unfolding forces for the last

transition seen without oligomers (violet) matched the distribution for P3 unfolding. (C) Under constant force, again only the high force transition was seen when A DNA oligomer blocks the formation of P2 (red), confirming that this transition involved P3 folding.

Similar behaviour was observed when measurements were repeated in the presence of saturating adenine concentration ($200 \mu\text{M} \gg K_D \sim 600 \text{ nM}$ (Rieder et al., 2007)), as shown in Figure 4.1C. In this case, the aptamer was held at $F \sim 0$ for 5 s between FECs to ensure that adenine binding occurred (Greenleaf et al., 2008); with an adenine concentration of $200 \mu\text{M}$, and using an average k_{on} of $3.10 \cdot 10^4 \text{ M}^{-1} \text{ s}^{-1}$ (Rieder et al., 2007), ligand binding should take $\sim 0.16 \text{ s}$, much shorter than the waiting time. The principal change due to adenine binding was an increase in the typical unfolding force for the fully-folded state, and a concomitant reduction in the population of the intermediate states; P2P3 was in particular often not observed. The average contour length changes between states, measured from over 3,000 FECs on 8 different molecules, were found to be identical to the results above within experimental uncertainty, indicating the same intermediates as when adenine is absent (Table 4.1).

Transition	F to U	F to A-comp	F to P2P3	F to P2	F to P3
ΔL_c expected	39.7 nm (71 nt)	10.6 nm (18 nt)	20.5 nm (31 nt)	29.5 nm (50 nt)	30.7 nm (52 nt)
ΔL_c measured (no adenine)	$39.6 \pm 0.5 \text{ nm}$		$19 \pm 1 \text{ nm}$		$30.3 \pm 0.5 \text{ nm}$
ΔL_c measured (with adenine)	$40.1 \pm 0.8 \text{ nm}$		$20.6 \pm 0.8 \text{ nm}$		$30.4 \pm 0.4 \text{ nm}$

Table 4.1 Aptamer unfolding contour length changes.

The expected contour length change from unfolding different structural components of the aptamer are listed, along with the values obtained from analysis of force-extension curves in the presence and absence of adenine. Uncertainties represent the standard error.

To characterise the intermediate states in greater detail, folding trajectories were measured at constant force using a passive force clamp (Greenleaf et al., 2005). The aptamer was first fully unfolded at

high force then the force was reduced in discrete steps, measuring the molecular extension while maintaining a constant force after each step for up to 1 minute to observe equilibrium behaviour. The force was stepped down until the aptamer was fully folded. In the absence of adenine, transitions between 5 different extensions were observed (Figure 4.1D), indicating 5 separate structural elements forming in apparently sequential order. The populations of these states changed as the force was reduced, as seen from histograms of the extension (Figure 4.1E), with the more-folded states (at shorter extensions) becoming more populated at lower forces. All five states could, however, be observed coexisting in equilibrium at $F \sim 10$ pN.

Analogously to the FEC analysis, the states were identified by converting the extension changes between them into the number of nucleotides folded, using $N_{nt} = [\Delta x(F) + \Delta n_h \cdot d_h] / \Delta x_{nt}(F)$, where $\Delta x_{nt}(F)$ is the extension/nt at a given force. The first folding transition from the unfolded state, at $F \sim 11$ –14 pN (Figure 4.1D, red), involved 19.0 ± 0.4 nt, precisely the number expected for hairpin P3. The second transition (at $F \sim 10$ –12 pN) involved 21.3 ± 0.4 nt, matching the expectation for P2. These are the same partially-folded intermediates seen in the FECs (P3 and P2P3); once again, the identifications were confirmed using an anti-sense oligo to block P2 folding (Figure 4.2C). The last two transitions displayed very different behaviour. At $F < 9$ –10 pN, the aptamer folded into a long-lived state 30.3 ± 0.7 nt shorter than P2P3 (Figure 1D, yellow). Since only 31 nucleotides remain unfolded once P2 and P3 have formed, this must be the fully-folded state F. Before F formed, however, a rarely- and transiently-occupied state ~ 5 nm shorter than P2P3 was typically seen (Figure 4.1D, green), at the extension expected when the loop-loop interaction has formed to pre-organize the binding pocket and only P1 is unfolded (8.6 ± 0.7 nt of folding measured from P2P3, 8 nt expected). We identified this as the state competent to bind adenine (labeled “A-comp” in Figure 4.1D), by analogy to a similar state in the *pbuE* aptamer (Greenleaf et al., 2008). This state was the only one not observed in the FECs.

When 200 μ M adenine was added, the qualitative behaviour of the aptamer was unchanged, with the same progression of states seen at similar forces. The principal differences involved helix P1: it folded at a slightly higher force, and most prominently, its lifetime was greatly increased. Indeed, once P1 folded

in the presence of adenine, it was never observed to unfold again (on the typical timescale of ~1 min) without a large force increase. Adenine thus clearly shifts the equilibrium for P1 folding, moving the mid-point between folded and unfolded P1, $F_{1/2}$, to a higher force that is above the equilibrium for formation of the kissing loops and possibly even hairpins P2 and/or P3. As a result, with adenine bound it was not possible to observe equilibrium folding of P1 under constant force. The results of the constant force analysis are summarised in Table 4.2.

4.3 Energy landscape of the aptamer

Since folding of the aptamer was apparently sequential, each transition could be analysed as a two-state system. For each transition, the force at equipose (where the molecule spends equal time in folded and unfolded states), $F_{1/2}$, was determined by fitting the probability distribution for the unfolded state, $P_u(F)$, to the Boltzmann relation (Woodside et al., 2006b): $P_u(F) = \{1 + \exp[(F_{1/2} - F) \Delta x / k_B T]\}^{-1}$. The position of the transition state along the reaction coordinate, Δx^\ddagger , was determined from the force-dependence of the kinetics (Woodside et al., 2006b): $k(F) = k_0 \cdot \exp(-F \Delta x^\ddagger / k_B T)$. This method was used for the transition state distances from both the folded and unfolded state, based on the respective folded and unfolded state kinetics. We assumed that the position of the barrier is force-independent (borne out fairly well over the relatively narrow range of forces measured in each transition by the linear force dependence of the logarithm of the rates). As a consistency check, $F_{1/2}$ was also determined from the force at which folding and unfolding rates were equal; this value was found to agree (within experimental uncertainty) with the value obtained from the probability distribution.

Folding transition	Δx (nm)	Δx (nt)	$F_{1/2}$ (pN)	$\ln(k_{1/2})$ (s^{-1})	$\Delta G_{1/2}^\ddagger$ (kcal/mol)	Δx_f^\ddagger (nm)	Δx_u^\ddagger (nm)
P3	5.2 ± 0.2	19.0 ± 0.5	12.9 ± 0.2	4.2 ± 0.1	4.3 ± 0.1	2.7 ± 0.2	4.1 ± 0.3
P2	6.3 ± 0.2	21.3 ± 0.5	10.8 ± 0.2	4.1 ± 0.1	4.4 ± 0.1	2.5 ± 0.3	4.6 ± 0.4
AC	5.1 ± 0.2	8.6 ± 0.6	10.1 ± 0.2	4.3 ± 0.2	4.3 ± 0.1	2.1 ± 0.2	3.9 ± 0.2
NAF	7.7 ± 0.4	22 ± 1	10 ± 2	2 ± 1	6 ± 1	N/A	N/A
AF	7.7 ± 0.4	22 ± 1	N/A	N/A	N/A	N/A	N/A

Table 4.2 Aptamer constant force results.

“P3” represents folding of hairpin P3, “P2” represents folding of hairpin P2, “AC” represents folding of the adenine-competent state, “NAF” represents complete folding in the absence of adenine, and “AF” represents complete folding in the presence of adenine.

Using the measurements above, we reconstructed the 5-state folding landscape of the aptamer. Because P1 folding was never in equilibrium in the presence of adenine, we first used Jarzynski’s equality (Jarzynski, 1997) to determine from the FECs the free energy change between F and P2P3. The free energy for unfolding P1, the binding pocket junction, and the kissing loop complex was found to be 10 ± 1 kcal/mol without adenine and 18 ± 2 kcal/mol with 200 μ M adenine. Compared to the free energy predicted for the secondary structure of P1 and the junction loops, estimated as ~ -8.6 kcal/mol, our results indicated a stabilization energy of ~ -1.5 kcal/mol from tertiary interactions that form in the absence of adenine (principally the loop-loop complex), and a more substantial -8 kcal/mol of additional stabilization upon adenine binding. This is consistent with NMR results showing some tertiary interactions without adenine but considerable stabilization of tertiary interactions upon adenine binding (Buck et al., 2007; Noeske et al., 2007), with a binding energy of ~ -8 kcal/mol as calculated from dissociation constants obtained by fluorescence studies (Rieder et al., 2007). We next found the height and location of the barrier for unfolding P1 by analysing the distributions of unfolding forces from the FECs. Representative unfolding force distributions measured at similar loading rates (~ 40 pN/s) with and without adenine (Figure 4.3A) showed a typical increase of several pN in unfolding force upon adenine binding. Values of $k(F)$ from datasets at different loading rates collapse to a single curve (Dudko et al., 2008), well fit by the same type of landscape model using Equation 3.7 (Figure 4.3B).

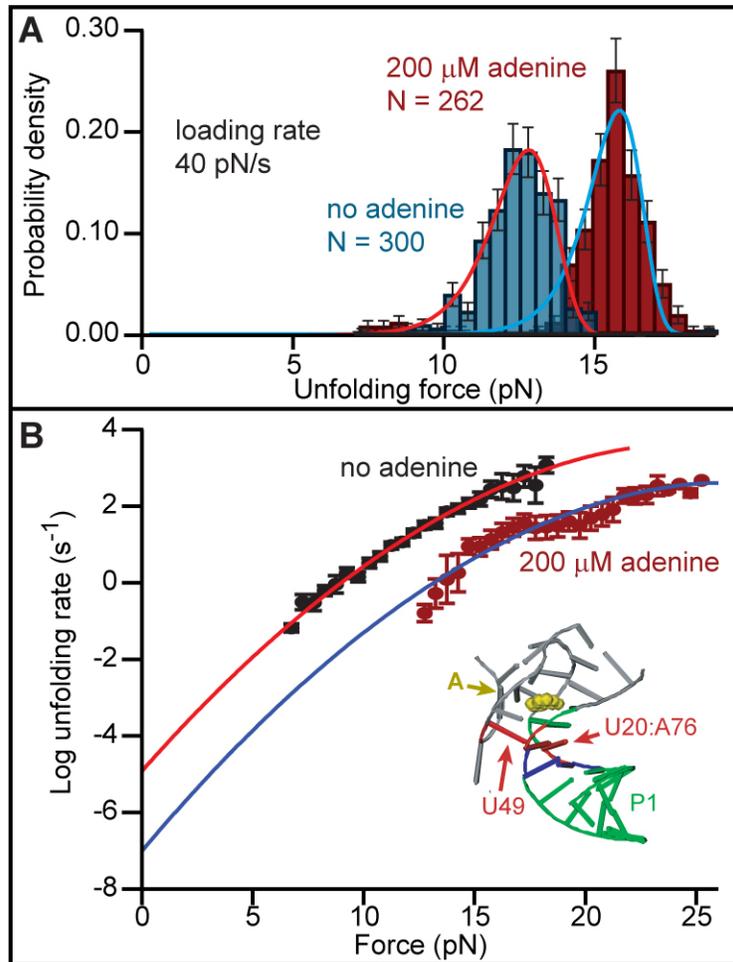


Figure 4.3 Unfolding force distributions and force-dependent rates determined from FECs.

(A) The unfolding force is higher with adenine bound (red) than without adenine (blue). Solid lines: fits to Equation 3.8. Error bars show s.d. (B) Unfolding rate as a function of force, fit to Equation 3.7. Error bars show s.e.m. The barrier to unfolding is located at the first base triple between P1 and the junction loop (inset, red, generated with VMD (University of Illinois at Urbana-Champaign)).

Distributions measured at loading rates ranging from 5–400 pN/s, with and without adenine, were analysed by both methods. The unfolding rate at zero force was found to be $1.5 \times 10^{-5} \text{ s}^{-1}$ ($\log k_{\text{off}} = -4.8 \pm 0.1$) without adenine and $3 \times 10^{-7} \text{ s}^{-1}$ ($\log k_{\text{off}} = -6.5 \pm 0.4$) with adenine bound. The barrier height was $12.3 \pm 0.4 \text{ kcal/mol}$ without adenine, and $16 \pm 2 \text{ kcal/mol}$ with adenine bound. The barrier position did not depend on adenine binding: $\Delta x^\ddagger = 6.2 \pm 0.6 \text{ nm}$ with adenine and $6.1 \pm 0.5 \text{ nm}$ without. Since ssRNA has an extension of 0.40–0.45 nm/nt in the force range 12–20 pN matching the most

probable unfolding forces, this result places the transition state for unfolding at 7–8 bp along helix P1. A comparison to the crystal structure of the ligand-bound aptamer (Serganov et al., 2004) reveals that the transition state is located next to the first basepair in P1 involved in triplex interactions with the junction loops: nucleotides U20:A76 in P1 and U49 in the loop, shown in red in Figure 4.3B (inset). These interactions therefore appear to act as structural keystones preventing unfolding.

Finally, these FEC results for P1 unfolding were combined with the constant force results for all the other transitions to reconstruct piecewise the energy landscape for the native folding pathway of the aptamer at a constant force of $F = 10.8$ pN, with and without the binding of an adenine molecule (Figure 4.4). The transitions other than P1 unfolding were analysed as sequential two-state processes. The relative positions and energies of the states were determined from constant-force extension histograms, whereas the positions and heights of the energy barriers between states were determined from the force-dependent kinetics (Greenleaf et al., 2008; Woodside et al., 2006a). All of the positions and energies were expressed relative to the unfolded state, which is the same regardless of the presence or absence of adenine.

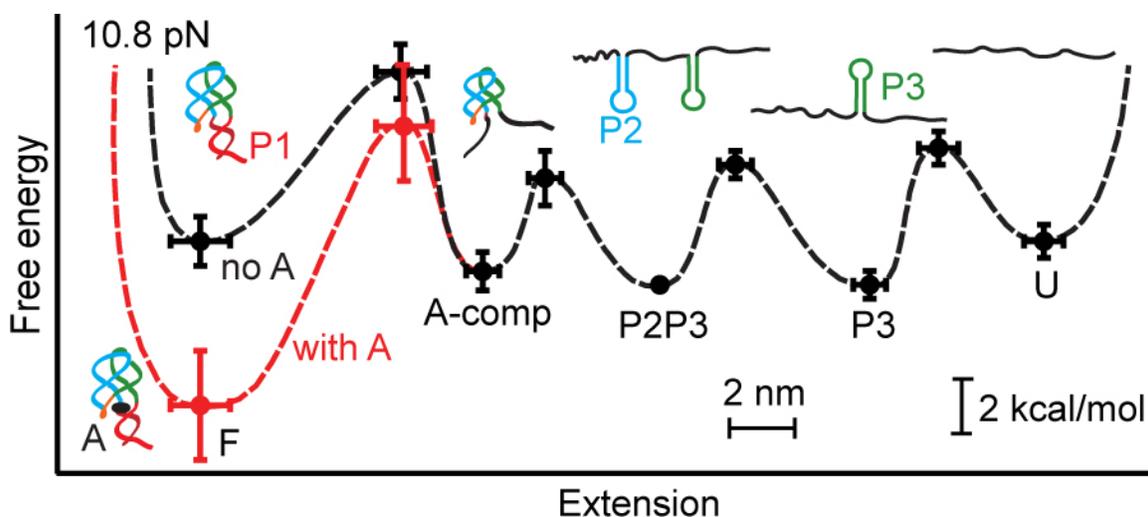


Figure 4.4 Folding energy landscapes for the *add* riboswitch aptamer.

The key features of the energy landscapes for the 5-state native folding pathway were reconstructed from piecewise two-state analyses of each transition. Energies and positions are plotted with reference to the P2P3 state. Error bars show s.e.m. Dotted lines indicate notional landscape shapes in the presence (red) and

absence (black) of adenine. Cartoons indicate the structural features associated with each state, deduced from the contour length changes relative to the fully unfolded state.

4.4 Misfolded states in the aptamer

Intriguingly, an additional set of states with different extensions and distinct features (Figure 4.5A) was observed occasionally at medium forces (~ 10 pN). Most prominent was a long-lived state labeled “M” in Figure 4.5A. This state was observed for all molecules, independent of adenine concentration (Figure 4.6A), occurring at a rate of $\sim 1 \text{ min}^{-1}$ with a lifetime of $\sim 1\text{--}10$ s. Within state M, additional transient folding events were also seen, such as the spike labeled “M3” in Figure 4A, with lifetimes of ~ 10 ms. The extension change from U to M, determined from extension histograms (Figure 4.6B), corresponds to 31 ± 1 nt of folded RNA. This does not match the length of any structural feature in the regular folding pathway of the aptamer. As the aptamer almost always ($\sim 95\%$ of the time) entered and exited this state via the fully unfolded state, rather than any of the partially folded intermediates, we therefore attribute it to an off-pathway, “misfolded” structure.

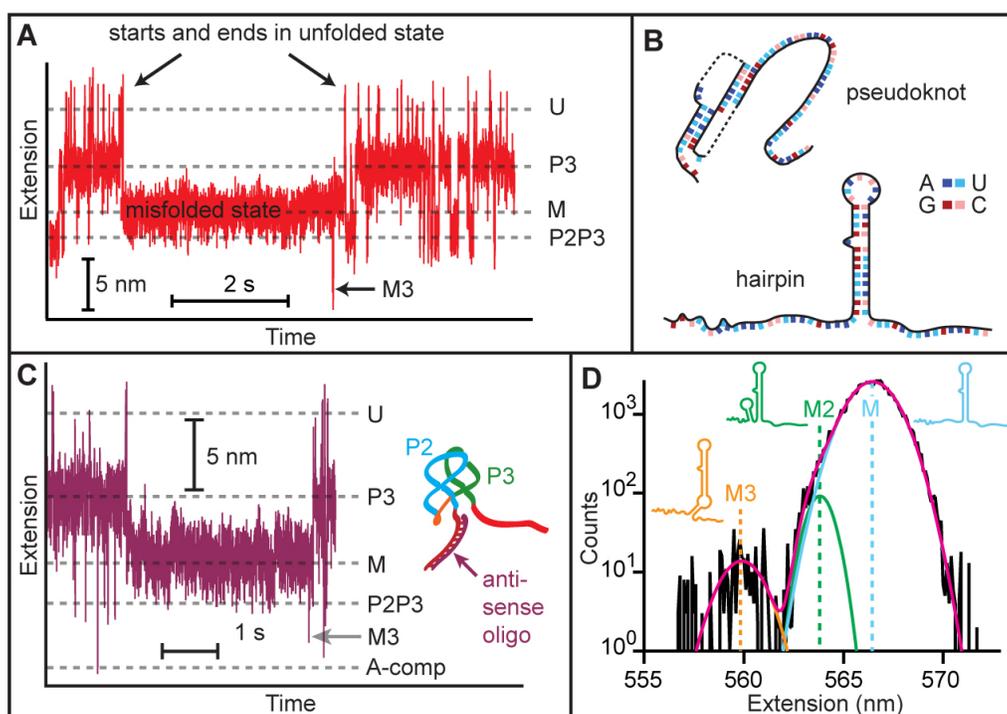


Figure 4.5 Misfolded states in the *add* riboswitch aptamer.

(A) Occasionally a state with extension between P3 and P2P3 (labelled “M”) forms from the unfolded state. It does not lead to any of the other states on the pathway to the natively folded state F—it begins and ends at U—but it does lead occasionally to other misfolded states with different extensions, such as the state M3. (B)

Two potential misfolded structures include a pseudoknot (nucleotides 17-45) and a hairpin (37-68). (C) With a DNA oligomer bound to the 5' strand of P1, the misfolded state still occurs, indicating that misfolding does not involve a pseudoknot. The A-comp state is still observed even when P1 is prevented from forming. (D) Extension histograms of the off-pathway, misfolded states reveal 3 distinct misfolded states which are well fit by Gaussian distributions. Insets: structures associated with each state.

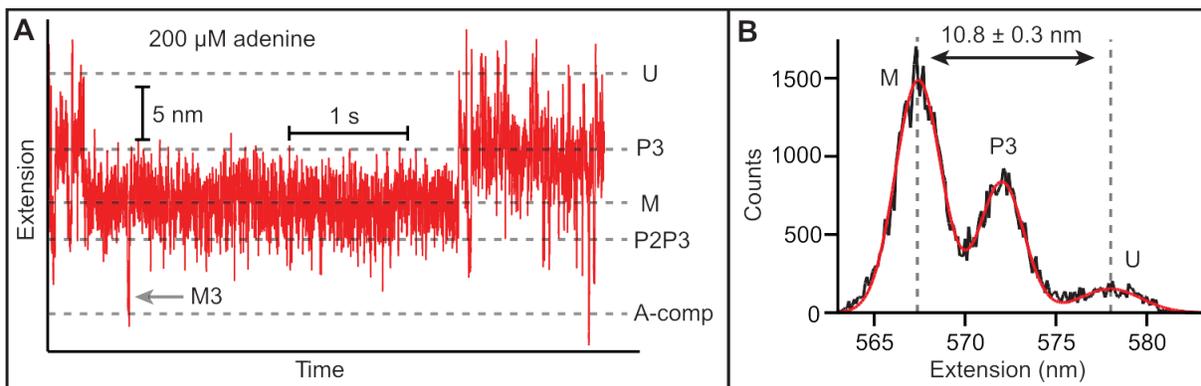


Figure 4.6 Misfolded states without adenine present.

(A) The same misfolded states were observed with adenine present as without adenine. (B) The extension change from the unfolded state (U) to the misfolded state (M) was determined from extension histograms.

A search of suboptimal structures using *mfold* (Zuker, 2003) and Hotknots (Andronescu et al., 2003; Ren et al., 2005) revealed two possible candidates for state M: a potential pseudoknot involving 29 nucleotides that would normally form hairpin P2 and the 5' end of P1, or a hairpin made from 32 nucleotides which would normally form the 3' stem of P2 through to the 5' stem of P3 (Figure 4.5B). By using anti-sense oligos to block in turn the 5' stem of P1 and the 8-nt junction loop between P2 and P3 (J2/3), we found that M still forms when P1 is blocked (Figure 4C) but not when J2/3 is blocked, indicating that M is the hairpin, not the pseudoknot. Evidence for at least two additional, transient folding transitions out of state M was found from the extension histograms (Figure 4.5D): one to a state that is 2.6 ± 0.5 nm shorter than M (labeled "M2") and another to a state that is 7.1 ± 0.5 nm shorter ("M3"). These distances are consistent with two of the extra misfolded states that may form along in concert with M (inset, Figure 4.5D): an 11-nt hairpin causing an extension decrease of 2.5 nm (M2: nucleotides 24–34) and a 16-nt helix-bulge decreasing the extension by 6.5 nm (M3: nucleotides 30–36 and 69–77).

The misfolded state clearly prevents the adenine binding site from forming, rendering the riboswitch non-functional. However, given that M forms rarely (only $\sim 0.3\%$ of the folding transitions initiated from U go to M), it seems unlikely to exert much influence on biological function. Indeed, numerous RNAs are known to misfold into non-native structures simply due to the rugged energy landscape generated by alternative patterns of basepairing (Russell, 2008), although few have been studied with single-molecule methods allowing individual trajectories to be followed (Li et al., 2007; Russell et al., 2002). Combining all the results above, we can build an integrated picture of the alternative pathways, each with multiple states (Figure 4.7).

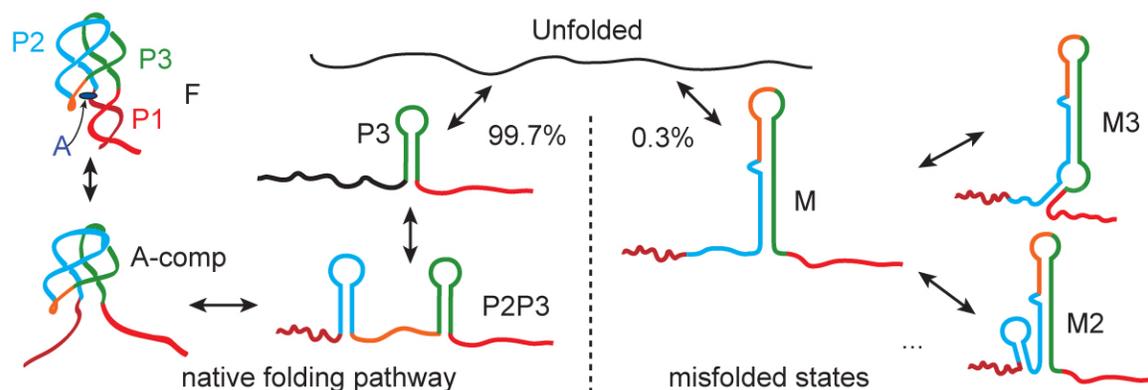


Figure 4.7 Folding pathways of the aptamer.

The native pathway involves sequential formation of hairpins P3 and P2, then the P2-P3 kissing loop complex (A-comp), and finally the fully-folded aptamer. 0.3% of the time the unfolded state folds into state M, preventing formation of the adenine binding pocket and leading to other off-pathway, misfolded states.

4.5 The full length riboswitch

In order to probe the interaction between the aptamer and the expression platform, we also made constructs containing the full riboswitch sequence, consisting of the aptamer plus the expression platform. FECs were measured by pulling the traps apart until the riboswitch unfolded, then rapidly bringing the traps together, waiting 5 s for folding and ligand binding, and repeating. FECs measured in the presence of 200 μM adenine (Figure 4.7A) look very similar to FECs of the adenine-bound aptamer

alone: the riboswitch unfolds between two principal states at relatively high force, with an unfolding force distribution (Figure 4.7A, inset) that matches the result found for the adenine-bound aptamer (Figure 4.3A, red).

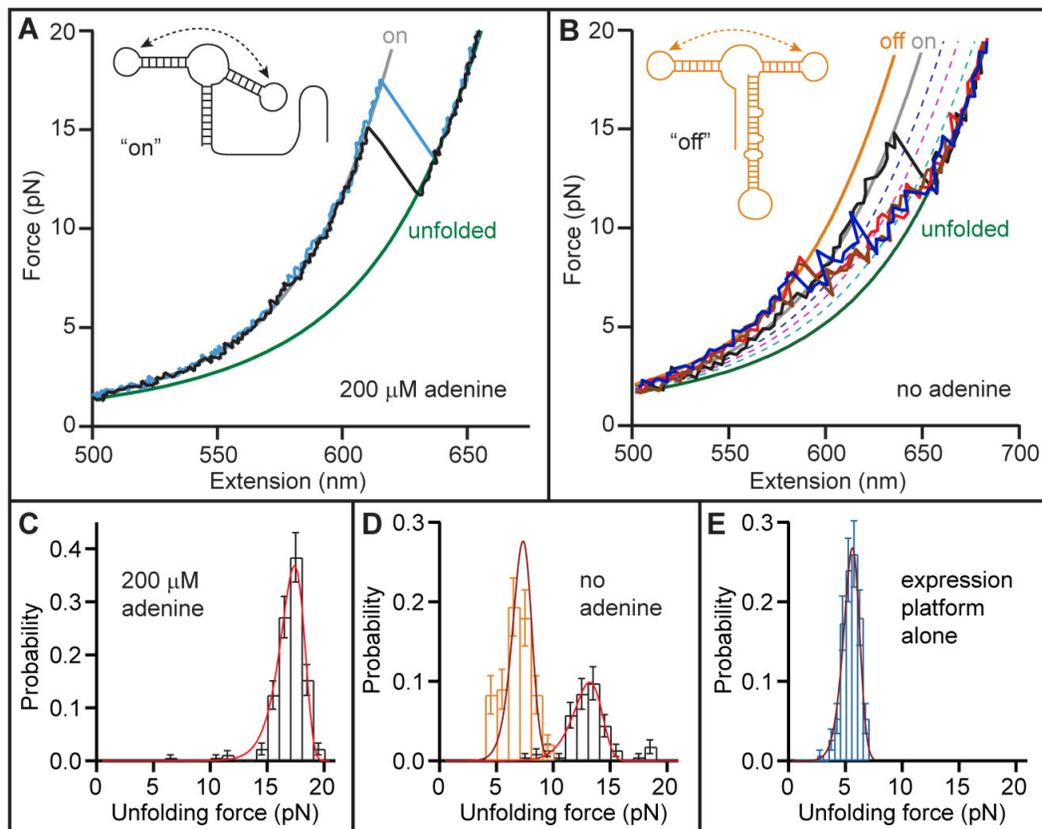


Figure 4.8 FECs of the full-length riboswitch.

(A) With adenine bound, the aptamer is folded (“on” state) and the contour length change was the same as for the aptamer alone. Solid lines: fits to “on” (grey) and unfolded (green) states. Inset: The unfolding force distribution of FECs measured with adenine present matches that expected for the adenine-bound aptamer. Red line: fit to Equation 3.8. (B) Without adenine, the expression platform was folded (red, blue, brown) more often than was the aptamer (black). Solid lines indicate fits to “on” (grey), “off” (orange), and unfolded (green) states. Dotted lines indicate partially-folded intermediates of the aptamer alone: A-comp (blue), P2P3 (pink), and P3 (cyan). Inset: secondary structure in the “off” state. Inset: The unfolding force distribution without adenine revealed two sub-populations. The higher-force peak (black) comes from curves initially in the “on” state and matches the distribution expected for the adenine-free aptamer (red line: fit to equation 3.8). The lower peak (orange) comes from curves initially in the “off” state and matched the distribution expected for the expression platform alone (brown line: fit to equation 3.8). (C) Unfolding force distribution of the expression platform alone (brown line: fit to equation 3.8). The most likely force was lower than for the

“off” state in (upper inset) due to a 15-fold lower loading rate. (D) Secondary structure in the “on” state. (E) Secondary structure in the “off” state.

The “on” and “off” states of the riboswitch were identified from the total contour length change upon complete unfolding of the full-length riboswitch. In the case of the “on” state, when only the aptamer is folded this should be $\Delta L_c = 39.7$ nm (71 nt and 1 helix). In the case of the “off” state, this should be 53.4 nm (98 nt and 2 helices, assuming that the loop-loop interaction between P2 and P3 is still formed). The two initial states were also characterized by different unfolding forces: the “on” state unfolded at high force (as for the adenine-bound aptamer), whereas the “off” state unfolded at lower force and moreover displayed additional partially-unfolded intermediates. The contour length change between the two states, determined from WLC fits to 990 FECs from 2 molecules, was $\Delta L_c = 40.7 \pm 0.5$ nm. This is very close to ΔL_c for complete unfolding of the aptamer alone (39.7 nm), indicating that the aptamer is almost always folded, the expression platform is unstructured, and the riboswitch is in the “on” state for gene expression (Figure 4.7D).

The relative thermodynamic stabilities of the translationally-active (“on”) and repressed (“off”) states of the riboswitch were estimated from the predicted secondary structure energy of the full-length riboswitch in the “on” and “off” states, adding in the tertiary structure stabilities measured from force spectroscopy. The secondary structure alone was predicted to be more stable by 1.4 kcal/mol in the “off” state (−17.9 kcal/mol) than in the “on” state (−16.5 kcal/mol). In the absence of adenine, tertiary structure in the “on” state (mostly the loop-loop interaction) brings another −1.5 kcal/mol of stability. However, since the loop-loop interaction can still form when P1 is prevented from folding (as seen by using the P1 blocking oligo, Figure 4.5C), the difference between “on” and “off” states remains unchanged. Since adenine binding provides an additional stability of −8 kcal/mol, with adenine bound the “on” state is more stable by 6.6 kcal/mol, whereas without adenine the “off” state is more stable by 1.4 kcal/mol.

When no adenine was present, the signature of aptamer unfolding was still observed some of the time (Figure 4.7B, black), but a majority of the FECs displayed qualitatively different behaviour (Figure 4.7B: red, blue, and brown). These two sub-populations of unfolding behaviour are reflected very clearly in the unfolding force distribution, which revealed two distinct peaks for the force of the first unfolding event in each curve (Figure 4.7B, inset). The higher-force peak (Figure 4.7B, inset black) results from the curves where the contour length change indicates the aptamer is initially folded and the riboswitch is “on” (Figure 5B, black). Fits to this part of the distribution using equation 3.8 (Figure 4.7B, inset red) returned the parameters expected for the adenine-free aptamer (compare to Figure 4.3A, blue), confirming the identification of the state as “on”. The riboswitch was observed to be “on” approximately 35% of the time, even though no adenine was present. Similarly, a small fraction (~ 10-20%) of the FECs measured with 200 μ M adenine revealed that the riboswitch was in the “off” state, much higher than the < 0.01% expected from the energy difference between “on” and “off” states predicted above. While these results show a clear shift in the minimum energy state, they also indicate that the alternative structures were not equilibrated properly during refolding before each FEC: likely, insufficient time was allowed to bind adenine to the aptamer or permit “on”/“off” equilibration, as seen in Figure 4.8E.

The lower-force peak was due to the other, majority class of FECs (Figure 4.8B: red, blue, and brown), which lacked the single prominent sawtooth feature seen for the aptamer alone and displayed instead a larger number of smaller, repeated unfolding/refolding transitions. In these curves, the first unfolding event occurred at considerably lower force (~ 5–8 pN) than for the “on” state, as seen from the unfolding force distribution (Figure 4.8B inset orange). The contour length of these curves also starts off shorter than that of the “on” state: fitting the low-force region of these FECs to a WLC model (Figure 4.7B, orange) revealed $\Delta L_c = 52.2 \pm 0.9$ nm to the fully-unfolded state (Figure 4.8B, green), measured from 1989 FECs on 4 molecules. This agrees well with the value of 53.4 nm expected for unfolding the full riboswitch in the “off” state, assuming that the expression platform is folded as well as P2, P3, and the loop-loop complex (Figure 4.8E). The low unfolding force observed is also consistent with unfolding the expression platform, since its three bulges and low G:C content reduce its mechanical stability (Woodside et al., 2006a; Woodside et al., 2006b). Indeed, the low-force peak of the distribution in Figure

4.7B, inset matches the unfolding force distribution measured using constructs that contain the expression platform alone (1400 FECs on 2 molecules), shown in Figure 4.7C. With no adenine present, we observed the riboswitch to be in the “off” state ~ 65% of the time.

In the FECs measured without adenine, the riboswitch was seen to fluctuate between 6 different states: the 5 states of the aptamer alone and the “off” state with expression platform folded. Indeed, what looks like noise in the FECs between the “off” state and the unfolded state is consistent with frequent inter-conversion between the different partially-folded states of the aptamer, shown in Figure 4.8B as dotted lines. However, the unfolding of the full riboswitch from the “off” state is not simply sequential, as for the aptamer alone, even though the unfolding force for the expression platform is several pN lower than that of any other structural element of the riboswitch. Instead, the presence of two alternative, fully-folded states was observed directly, through switching between the “on” and “off” states.

Examples of this behaviour are seen in Figure 4.8B (brown curve at ~ 6 pN, blue curve at ~ 8–9 pN). It is even seen very occasionally with adenine present, as in Figure 4.9. Here, the riboswitch was initially folded in the “off” state (without adenine bound), but at $F \sim 5$ pN the contour length twice increased briefly to the “on” state value before returning to the “off” state, *i.e.* the folded expression platform was transiently replaced by a folded P1. A last switching event was seen at ~ 7 pN, when the “off” state unfolds to the contour length expected for the adenine-competent state (*i.e.*, with repressor hairpin completely unfolded) before P1 folded to produce the “on” state one last time. Adenine clearly binds to the aptamer soon thereafter, since the “on” state was stabilised and does not unfold until a high force characteristic of the adenine-bound state.

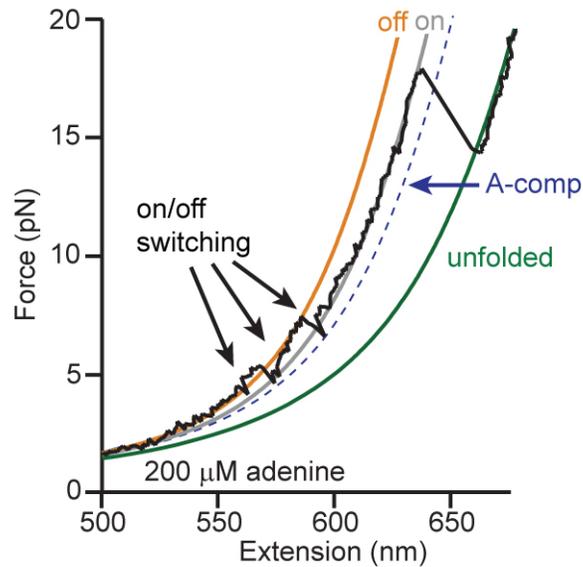


Figure 4.9 Switching and ligand binding.

A riboswitch, initially “off” when in the presence of adenine, switches several times between “off” and “on”, eventually binding adenine and remaining in the thermodynamically more stable adenine-bound “on” state before unfolding completely. For comparison, with 200 μM adenine, binding should take ~ 0.16 s. Here the pulling rate is 200 nm/s

4.6 Comparing the mechanisms of the *add* and *pbuE* riboswitches

These results provide new insight into the relation between folding and functional mechanism in the *add* riboswitch, both at the level of the aptamer alone as well as the full-length riboswitch, especially through comparison to previous results for the *pbuE* riboswitch (Foster, 2010; Greenleaf et al., 2008). While the two riboswitches have aptamers with very similar structures, they operate by quite different mechanisms: translation activation for *add* as opposed to transcription anti-termination for *pbuE*. Comparison of the behaviour of the aptamers reveals many commonalities due to the similar structures, but also some important differences. In both cases, a total of 5 states were observed, and the aptamers were found to fold sequentially: first the two hairpins (P2 and P3) forms, then the loop-loop interaction, and finally P1. Crucially, the helix P1—which acts as the mechanical switch—is the last part of the molecule to fold. In both cases, this helix is strongly stabilised by adenine binding, whereas the energetic and kinetic properties of the other states are not affected (Figure 4.4). The reconstructed energy

landscapes are in fact quantitatively very similar, due to the extensive structural and sequence similarities between the two aptamers.

The landscapes in Figure 4.4 also highlight some of the important differences between the aptamers. First, the order of folding of the two hairpin loops is reversed: P3 folds at higher force than P2 in the *add* aptamer, vice versa for the *pbuE* aptamer. This effect was predicted by simulations (Lin and Thirumalai, 2008) and is due to the different relative stabilities of the helices P2 and P3 in the two different aptamers. More fundamental is the difference in stability of helix P1. In the *add* aptamer, P1 is similar in stability to P2 and P3 even in the absence of adenine ($F_{1/2} \sim 10$ pN) and it has a low unfolding rate at $F=0$ ($k_{\text{off}} \sim 10^{-5} \text{ s}^{-1}$). In contrast, P1 in the *pbuE* aptamer is quite unstable without adenine bound ($F_{1/2} = 3$ pN) and has an unloaded unfolding rate k_{off} that is 10,000 times higher. As a result, the *pbuE* aptamer is much more susceptible to strand invasion by the expression platform, the process which is required to form the terminator hairpin and switch the gene off.

4.7 The evidence for thermodynamic control in the *add* riboswitch

The different stabilities of P1 appear to be related directly to the different functional mechanisms of the riboswitches. For the *pbuE* riboswitch, the conformation with terminator hairpin folded and aptamer unfolded is the most stable thermodynamically, regardless of whether or not adenine is bound to the aptamer (Lemay et al., 2006); all indications point to a mechanism whereby the outcome of riboswitch folding (and hence level of gene expression) is governed by a competition between the kinetics of aptamer folding, ligand binding, and transcription elongation (Greenleaf et al., 2008; Lemay et al., 2006; Wickiser et al., 2005a). In such a situation, an unstable P1 is what is needed to permit strand invasion on a timescale competitive with transcription rates. Kinetic control of riboswitch folding has also been demonstrated for the FMN riboswitch (Wickiser et al., 2005b). The measurements on the full-length *add* riboswitch, in contrast, show quite clearly that the thermodynamically stable conformation switches upon adenine binding: the state with the aptamer folded and expression platform unfolded (“on”) is dominant, and hence most stable, when adenine is bound, but the state with the expression platform folded and aptamer partially-unfolded is dominant when adenine is not bound. Results such as

those in Figure 4.8 encapsulate the essence of the mechanism of this riboswitch: the RNA inter-converts from “off” to “on”, with the outcome ultimately determined by thermodynamic changes due to ligand binding.

The fact that this riboswitch is controlled by the thermodynamics of the competing states can also be deduced from the measurements on the aptamer alone. Adding the measured tertiary-interaction and ligand-binding energies to the stabilities of the secondary structures in the “on” and “off” states calculated using *mfold*, we estimated that the “off” state should be more stable than the “on” state by ~ 1.4 kcal/mol when no adenine is bound (“off” state: -17.9 kcal/mol vs. “on” state: -16.5 kcal/mol, from secondary structure), whereas the “on” state should be more stable by ~ 6.6 kcal/mol with adenine bound. In the absence of adenine, tertiary structure in the “on” state (mostly the loop-loop interaction) brings another -1.5 kcal/mol of stability. (Since the loop-loop interaction can still form when P1 is prevented from folding (as seen by using the P1 blocking oligo), the difference between “on” and “off” states remains unchanged. Since adenine binding provides an additional stability of -8 kcal/mol, with adenine bound the “on” state being more stable by 6.6 kcal/mol, whereas without adenine the “off” state would be more stable by 1.4 kcal/mol.) Experimentally, the “off” state was occupied ~ 1.7 times more often than the “on” state in the absence of adenine, implying that the “off” state is more stable by ~ 0.3 kcal/mol. This agrees with the estimate above, within error (~ 1 kcal/mol), although the energy difference deduced from our measurements may be underestimated due to incomplete equilibration between structures arising from an insufficient delay time between each FEC.

Interestingly, the folding rate of the aptamer in the absence of adenine implies that the *add* riboswitch regulatory mechanism must, of necessity, be controlled thermodynamically. We determined the folding rate at $F = 0$, k_{fold} , from refolding FECs measured in the absence of adenine while ramping the force down to 0 pN from ~ 20 pN (where the aptamer is completely unfolded). 1,227 refolding FECs were measured, and the cumulative folding probability were analysed (Dudko et al., 2008) similar to the analysis in Figure 4.3B. The average refolding times plotted as a function of force were then extrapolated

to zero force by fitting to equation 3.7, resulting in an unloaded folding rate, $k_{\text{fold}} \sim 30 \text{ s}^{-1}$ (Figure 4.10A). Given a typical transcription rate *in vivo* of $\sim 50\text{--}100 \text{ nt/s}$ (Darzacq et al., 2007; Mooney et al., 1998) and the 40-nt length of the expression platform, this folding rate implies that the aptamer should fold completely while the repressor hairpin has only been partially transcribed, regardless of adenine binding. The riboswitch should therefore always start in the “on” state immediately after transcription. Since adenine binding locks the riboswitch into the ‘on’ state with a large binding energy and low unfolding rate, then if the riboswitch is to act as a switch, clearly the ‘on’ state must be able to equilibrate into the ‘off’ state rapidly in the absence of adenine, before the ribosome binds the Shine–Dalgarno sequence. To determine this equilibration rate, 5012 FECs were measured on 11 molecules while varying the time spent waiting at low force to permit equilibration, from 0.3–5 s. An exponential fit to the fraction of molecules found in the ‘off’ state as a function of time (Figure 4.10B) yielded an equilibration time of $\tau_{\text{eq}}=0.2\pm 0.1 \text{ s}$. The ribosome binding time is not well known, but a lower bound (even for coupled transcription and translation) is provided by the time required to transcribe the mRNA footprint for the ribosome. Footprinting studies and crystal structures (Huttenhofer and Noller, 1994; Yusupova et al., 2001) imply this involves another $\sim 15 \text{ nt}$ downstream of the 3'-end of the riboswitch expression platform, requiring $\sim 0.3 \text{ s}$ at typical transcription rates (or $1.5 \tau_{\text{eq}}$). Hence the ‘on’ and ‘off’ states should be substantially (if not completely) equilibrated by the time the ribosome binds, strongly supporting a thermodynamically controlled regulatory mechanism. The picture of the riboswitch mechanism suggested by our results is summarized in Figure 4.10C. During the transcription of the expression platform, the riboswitch first folds into the ‘on’ state. If adenine does bind to the ‘on’ state, then it becomes greatly favoured thermodynamically and the riboswitch stays ‘on’. If adenine does not bind to the aptamer by the time the expression platform is fully transcribed, the “on” state is no longer thermodynamically stable and the riboswitch is driven to the “off” state, repressing gene expression. However, the ‘on’ state is still populated at an appreciable level. Hence a switch in the thermodynamic minimum upon adenine binding is required to obtain any regulatory action.

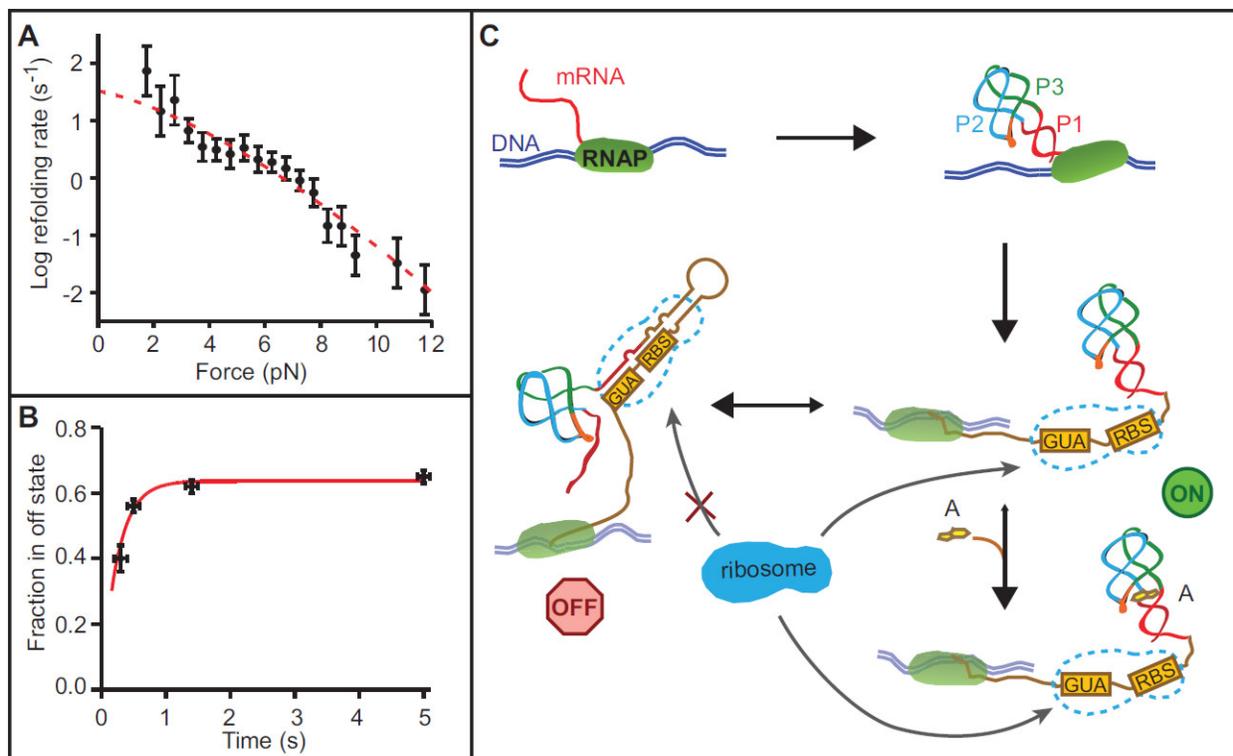


Figure 4.10 Folding kinetics and riboswitch mechanism.

(A) The aptamer refolding rate as a function of force without adenine was determined from the distribution of refolding forces when ramping the force down from denaturing values. Error bars show SEM. Red line: fit to equation (3.7). (B) The fraction of unfolding FECs in the off state as a function of refolding time at low force in the absence of adenine showed an exponential rise as the riboswitch structure equilibrated into the more stable ‘off’ state. (C) Schematic of the riboswitch mechanism. The aptamer folds rapidly before the expression platform is transcribed, regardless of adenine binding. If adenine binds to the aptamer, it stabilizes the ‘on’ state of the riboswitch (aptamer folded, ribosome binding site exposed). Without adenine binding, the ‘on’ state is unstable and equilibrates into the ‘off’ state (P1 unfolded, ribosome binding site sequestered). RBS: ribosome binding site.

Chapter 5: RNA pseudoknot conformational plasticity

The second class of RNA molecules whose dynamics I studied with optical tweezers were viral pseudoknots that stimulate programmed frameshifting. In this chapter, I describe smFS measurements of a panel of pseudoknots from different viruses, investigating what features of the pseudoknot folding are related to their ability to stimulate frameshifting to varying degrees. It was found that the parameters describing the energy landscape for mechanical unfolding of the pseudoknot (energy barrier height and distance to the transition state) could not be correlated to frameshifting efficiency. Alternatively, it was found that, increased frameshifting efficiency was correlated with an increased tendency to form alternate, incompletely-folded structures, suggesting a more complex picture of the role of the pseudoknot involving the conformational dynamics and a more complex folding landscape. The original model involved the ribosome being pushed back along the slippery sequence of the RNA by the pseudoknot acting as a large roadblock it could not unfold. Our work suggests a new way of thinking about the importance of mRNA structural dynamics in programmed frameshifting. A variation of this chapter was already published: Dustin B. Ritchie, Daniel A. N. Foster, and Michael T. Woodside “Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding” *PNAS* 109: 16167-16172 (2012). In this work, DBR and MTW conceived the experiments, DBR produced the samples, DBR and DANF performed experiments and all authors analysed the data.

5.1 Ribosomal Frameshifting

A ribosomal frameshifting event in messenger RNA (mRNA) is one where the reading frame of the ribosome changes along the mRNA thereby generating an alternate gene product from the same mRNA. It occurs in different ways in nature, being utilised by viral systems, retrotransposons (McDonald, 2012), and bacterial insertion elements (Farabaugh, 2012), for examples. (And can of course occur spontaneously at lower rates (Streisinger and Owen, 1985).) Ribosomes synthesise protein by reading the mRNA in 3-nucleotide (nt) steps to maintain a reading frame until a stop codon is reached. Because of the three letter amino acid coding relationship, three different reading frames are possible in

an mRNA. Shifting the reading frame by a single nucleotide yields a different amino acid sequence, so that a single gene track could code for 2 proteins (Ketteler, 2012) e.g. dnaX (Larsen et al., 1997). Viral frameshifting is known to use a frameshift in both the +1 and -1 direction (Dinman, 2006). In -1 programmed ribosomal frameshifting (-1PRF), the shift is triggered by a pseudoknot (PK) structure in the mRNA typically in combination with an upstream slippery sequence (rich in Us), the ribosome is forced backward by 1 nt, thereby resulting in the bypass of a stop codon and the decoding of a new reading frame that specifies a different amino acid sequence (Brierley et al., 2010; Giedroc and Cornish, 2009). It has important biomedical and commercial impacts as the functioning of many RNA viruses from plants and animals critically depend on the tight regulation of the optimum expression ratio of the frameshifted proteins. Many RNA viruses from plants and animals use -1 PRF to express two proteins from a single mRNA in coordination at a defined ratio (Brierley et al., 2010; Farabaugh, 1996; Giedroc and Cornish, 2009; Jacks and Varmus, 1985). Essential aspects of viral function, such as replication of the viral genome and packaging of the genome into a virion, depend critically on the tight regulation of the optimum expression ratio of the frameshifted proteins (Dinman and Wickner, 1992; Dulude et al., 2006). -1 PRF also occurs during the expression of cellular genes in a wide range of other organisms (Baranov et al., 2011; Farabaugh, 1996; Manktelow et al., 2005; Wills et al., 2006).

The factors that determine frameshifting efficiency are not yet fully understood. -1PRF depends on two specific components in the mRNA: a 7-nt “slippery sequence” at which -1 PRF occurs, and a stimulatory structure (usually but not always a pseudoknot) located 6–8 nt downstream of the slippery sequence (see Figure 5.1) (Brierley et al., 2010; Brierley et al., 1992; Dinman et al., 1991; Farabaugh, 1996; Giedroc and Cornish, 2009; ten Dam et al., 1990). This class of pseudoknot (known as H-type) is formed from the binding single-stranded loop of a hairpin base-pair with complementary nucleotides outside of that loop (Figure 5.2) (Dam et al., 1992). The slippery sequence can generate -1 PRF on its own, with some sequences increasing the intrinsic level of -1 PRF errors by up to 100-fold per codon; the pseudoknot component further stimulates frameshifting, 10–30 times the level produced by the slippery sequence alone (Giedroc and Cornish, 2009). Previous single-molecule work has suggested that

frameshifting efficiency is related to the resistance of the pseudoknot against mechanical unfolding, which is now a widely accepted model. Various models have been proposed to explain how -1 PRF could occur at various steps in the elongation cycle (Brierley et al., 2010; Giedroc and Cornish, 2009) (Jacks et al., 1988; Namy et al., 2006; Plant and Dinman, 2005; Plant et al., 2003). In one commonly-cited model, the pseudoknot is viewed as a mechanical roadblock hindering ribosome translocation just when the slippery sequence is in registry with the A and P sites of the ribosome (Kontos et al., 2001; Namy et al., 2006). The strain from the mechanical resistance of the pseudoknot is thought to deform the P-site tRNA, weakening the codon-anticodon base pairing and promoting a -1 shift in reading frame. Interestingly, recent measurements of translocating ribosomes show that the ribosome actively generates tension in the mRNA to open the junction of structured RNAs and promote unwinding at the mRNA entry site (Qu et al., 2011; Wen et al., 2008).

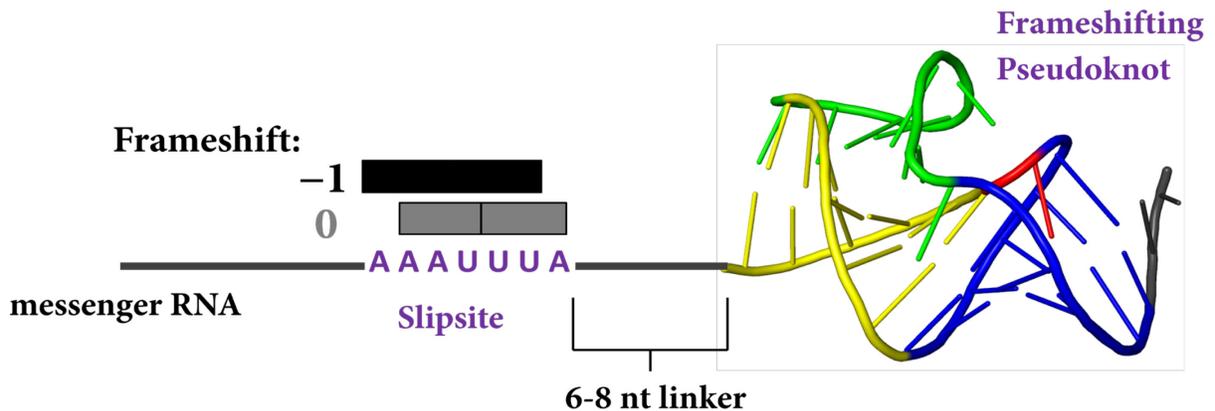


Figure 5.1 Schematic of -1 PRF.

While the ribosome is reading along the slipsite, upstream of the pseudoknot, an interaction of it with the pseudoknot (or structures within that sequence) cause the mRNA to be pulled back by 1 nt, changing the reading frame in the -1 direction. (Figure courtesy of Dustin Ritchie.)

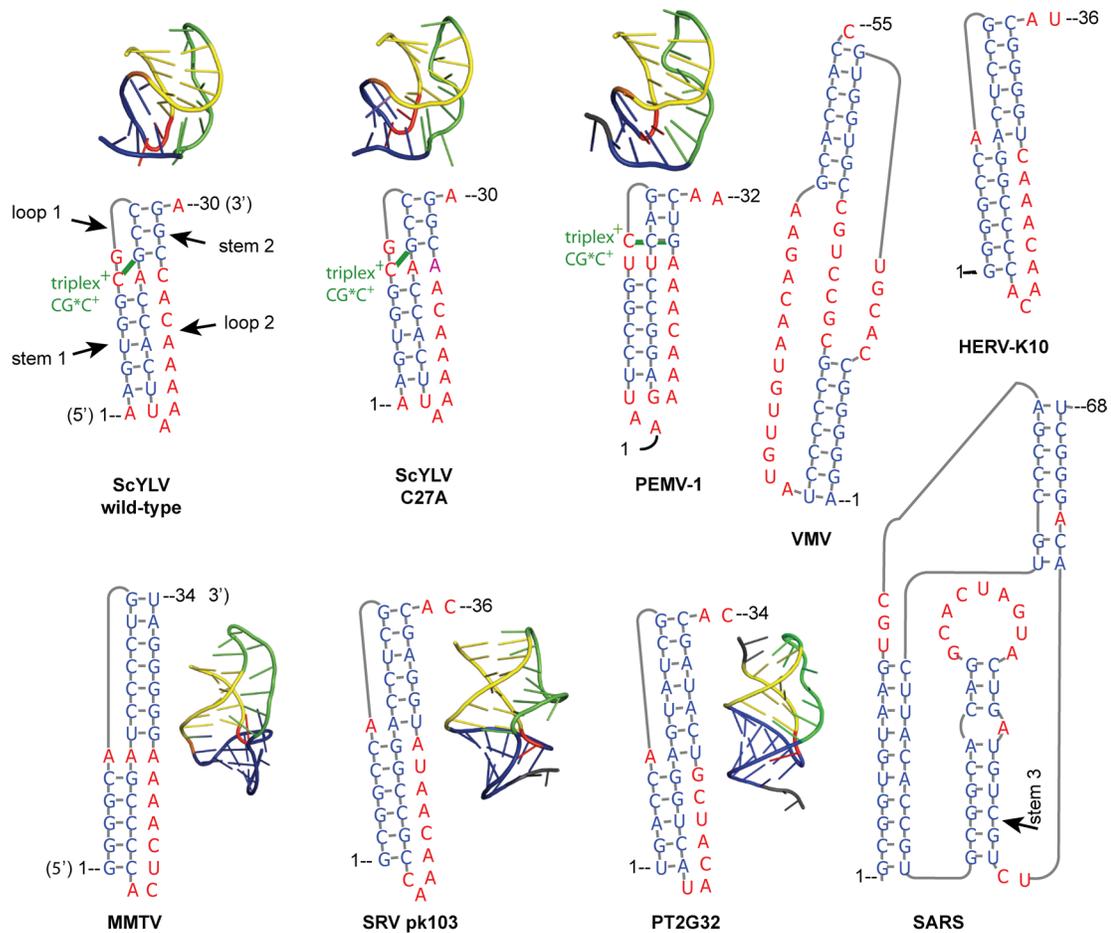


Figure 5.2 Secondary and tertiary structure of the pseudoknots.

Most of the pseudoknots are H-type, and consist of 2 stems and two loops, as indicated for the ScYLV wild-type pseudoknot. In addition, the VMV pseudoknot contains a lengthy inter-stem element (shown in magenta), and the SARS pseudoknot, a non-H-type, contains a third stem and loop. Base-paired nucleotides are indicated in blue in the secondary structure, unpaired nucleotides in red or purple. High-resolution structures are shown for the pseudoknots where available. The stems and loops are colour-coded in the tertiary structure: stem 1 yellow, loop 1 red, stem 2 blue, loop 2 green. 2D renderings generated with Pseudoviewer3 (Byun and Han, 2009; Ponty and Leclerc, 2015)

Given this mechanical model of pseudoknot-stimulated -1 PRF, where -1 PRF is dependent on tension induced in the mRNA when the translocating ribosome encounters a folded pseudoknot (Namy et al., 2006; Plant and Dinman, 2005), -1 PRF efficiency is expected to depend on how strongly the pseudoknot resists unfolding (Cao and Chen, 2008), which would produce higher energy barriers in the

unfolding landscapes for those higher forces. However, -1 PRF efficiency seems to be unrelated to pseudoknot thermodynamic stability (Chen et al., 1995; Kang et al., 1996; Naphine et al., 1999). A similar result is obtained when using duplexes formed by anti-sense oligos to induce -1 PRF (Howard et al., 2004), although there is conflicting evidence from -1 PRF induced by hairpin structures (Yu et al., 2011). Moreover, the extent of pseudoknot-induced ribosomal pausing is not strongly correlated with -1 PRF efficiency (Kontos et al., 2001), as might be expected from this picture. On the other hand, base triples formed between loop 2 and the minor groove of stem 1 (Figure 5.2), which should increase the pseudoknot stability, do stimulate efficient -1 PRF (Chen et al., 2009; Cornish et al., 2005; Kim et al., 1999; Liphardt et al., 1999; Michiels et al., 2001; Nixon et al., 2002; Olsthoorn et al., 2010; Shen and Tinoco, 1995; Su et al., 1999).

The possible correlation between the ability of pseudoknots to promote -1 PRF and their mechanical stability against unfolding was recently investigated directly using single-molecule force spectroscopy (smFS) (Chen et al., 2009; Chen et al., 2007; Green et al., 2008; Hansen et al., 2007; White et al., 2011), whereby tension is applied to the ends of the pseudoknot until it unfolds (Woodside et al., 2008). This approach is particularly appropriate because the mode of unfolding mimics how the ribosome unwinds RNA structure, by actively applying force to the mRNA (Qu et al., 2011; Wen et al., 2008). smFS may permit the observation of heterogeneity in the folding or unfolding, and might reveal the presence of metastable states. Two smFS studies found a correlation between -1 PRF efficiency and pseudoknot unfolding force (Chen et al., 2009; Hansen et al., 2007) by studying the effects of destabilizing mutations, especially mutations disrupting major-groove base triples (Chen et al., 2009). In contrast, two other studies found no such correlation (Green et al., 2008; White et al., 2011), although slower unfolding rates were correlated with higher -1 PRF efficiency over a narrow force range (Green et al., 2008), supporting a hypothesis that the rate of unfolding or barrier height might be important (Giedroc et al., 2000; Plant et al., 2003). It is unclear from such measurements whether the lowered frameshifting efficiency in fact results from the lower unfolding force, or rather from possible structural changes and the prevention of specific contacts with the ribosome when key stabilizing interactions are

removed. Each of these studies, however, was restricted to a single pseudoknot and its mutations; a broad survey of many different types of pseudoknots has yet to be made.

5.2 Pseudoknot frameshifting across a broad range of efficiencies

To investigate possible deterministic correlations between frameshifting efficiency and pseudoknot resistance to mechanical unfolding, we used a panel of 9 pseudoknots. We measured four retroviral pseudoknots, from the simian retrovirus-1, SRV1 (Michiels et al., 2001); human endogenous retrovirus-K10, HERV (Wang et al., 2002); Visna-Maedi retrovirus, VMV (Pennell et al., 2008); and mouse mammary tumor virus, MMTV (Chamorro et al., 1992; Shen and Tinoco, 1995). Luteoviral pseudoknots from the pea enation mosaic virus-1, PEMV1 (Nixon et al., 2002), and sugarcane yellow leaf virus, ScYLV, along with a ScYLV mutant (C27A) producing much reduced -1 PRF efficiency (Cornish et al., 2005; Cornish et al., 2006), were also included, as was a pseudoknot (non-H-type) from the SARS coronavirus (Baranov et al., 2005). Finally, a non-frameshifting (2% efficiency) pseudoknot from the bacteriophage T2 gene 32, PT2G32 (Holland et al., 1999), which is structurally similar to the SRV1 pseudoknot (Michiels et al., 2001), was included as a control. These pseudoknots were chosen for several reasons: (i) most (SRV1, PT2G32, ScYLV, PEMV1, MMTV, HERV) have similar size and topology (H-type) despite causing different -1 PRF efficiency; (ii) many have high-resolution structures, so the that lowest energy folded state is known for these pseudoknots; (iii) the effects of larger size (VMV and SARS), long inter-stem elements (VMV), and different topology (3 stems in SARS) (Park et al., 2011; Plant et al., 2005) can be explored; and (iv) they represent a wide range of -1 PRF efficiencies: 2–30% as measured in rabbit reticulocyte lysate. The sequences and structural properties of all pseudoknots are summarised in Figure 5.2 and Appendix table D.2.

5.3 Experimental pseudoknot extension changes under tension

RNA constructs were created by inserting the sequence for the pseudoknot into the pMLuc-1 plasmid between the SpeI and BamHI restriction sites. The resulting transcription template containing the pseudoknot or riboswitch sequence flanked on each side by kb-long “handle” sequences

was transcribed *in vitro* using T7 RNA polymerase. The RNA transcript was then annealed to single-stranded (ss) DNA complementary to the handles, and attached to beads held in optical traps (Figure 2.1), as described previously (Neupane et al., 2011). The handle lengths are 840 nt on the 3' end of the transcript and labeled with biotin, and other 2280 nt on the 5' end of the transcript and labeled with digoxigenin. The handles were produced by asymmetric PCR from double-stranded DNA PCR products corresponding to the flanking handle sequences (Saiki et al., 1986). The handles were annealed with the RNA transcript then incubated with 600 nm and 820 nm diameter polystyrene beads labeled with avidin DN (Vector Labs) and anti-digoxigenin (Roche), respectively, to create dumbbells. Dumbbells were placed in measuring buffer (50 mM MOPS, pH 7.0, 130 mM KCl, 4 mM MgCl₂, ≥50 U/mL Suprase-In RNase inhibitor (Ambion) and oxygen scavenging system: ≥40 U/ml glucose oxidase, ≥185 U/mL catalase, and 8.3 mg/mL D-glucose) and inserted into a sample chamber on a clean microscope slide in the optical trap. Buffer ionic strength, which can affect unfolding energies, rates, and pathways (Giedroc et al., 2000; Green et al., 2008; White et al., 2011), was chosen to be near-physiological, as required duration translation for the pseudoknots. The FECs were sampled at 20kHz, while the constant force was sampled at 50kHz.

For each pseudoknot, force extension curves (FECs) were measured with pauses near zero force for 3–10 s to permit folding of the RNA, then the beads were separated at constant velocity. Representative FECs are shown in Figure 5.3A for unfolding of the ScYLV C27A pseudoknot, where unfolding typically occurred at 20–40 pN as a two-state process, without intermediates (Fig. 1A, black, red). The change in contour length during unfolding, ΔL_c , was found to be 13.9 ± 0.7 nm (all errors represent standard error on the mean), which agreed well with the value 14.2 nm expected from the NMR structure (Cornish et al., 2006), indicating that the pseudoknot was natively folded. Very rarely, an unexpectedly short ΔL_c for unfolding was found in a FEC, indicating that the pseudoknot was not folded into the native structure before that pull and hence started from an alternate structure (Figure 5.3A, blue; WLC fit: green).

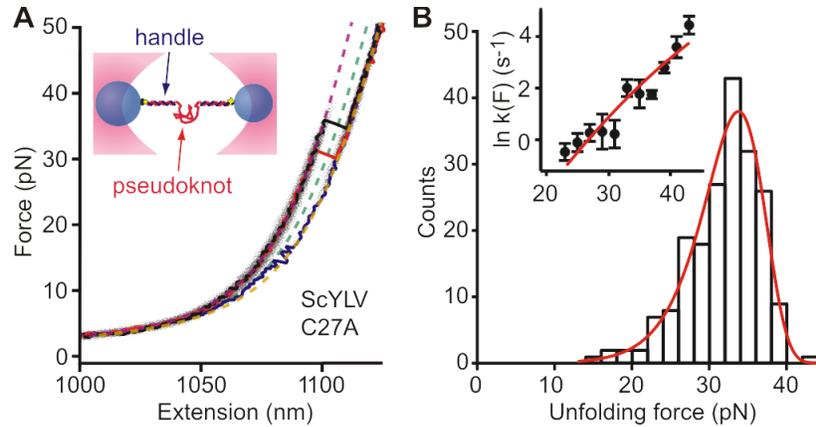


Figure 5.3 Force spectroscopy of ScYLV C27A pseudoknot.

(A) Inset: RNA containing the pseudoknot flanked by handle sequences was annealed to DNA strands complementary to the handles and attached to beads held in optical traps. Individual FECs (black, red, blue) are plotted above the aggregated data from 200 FECs (grey). Most FECs (black, red) show a monotonic rise of force with extension up to ~ 30 pN, at which point the extension increases abruptly as the RNA unfolds. A few FECs (blue) unfold at lower forces with a smaller length increase, indicating the RNA started in a different structure composed of fewer nucleotides. WLC fits to the elasticity of the handles and unfolded RNA, used to determine the contour length change upon unfolding, are shown for three different states of the RNA: fully folded (purple); fully unfolded (brown); incompletely folded (green). **(B)** The distribution of unfolding forces from the natively folded pseudoknot FECs (black) was well fit by equation 3.8 (red), yielding parameters describing the mechanical resistance to unfolding. Inset: Unfolding rate as a function of force (black) was well fit by equation 3.7 (red).

To quantify the resistance of the pseudoknot to mechanical unfolding, we examined the distribution of unfolding forces in the FECs, $p(F)$ (Figure 5.3B). The average unfolding force, here 32 ± 2 pN, provided the simplest measure of the mechanical stability. Other parameters related to the resistance to mechanical unfolding were obtained from fitting the shape of $p(F)$. As seen in Figure 5.3B (red), $p(F)$ was well fit by equation 3.6. A complementary analysis of the kinetics based on the cumulative probability of unfolding (Dudko et al., 2008) yielded the unfolding rate $k(F)$ as a function of force (Figure 5.3B, inset), which was well fit by the same type of landscape model using equation 3.7. Distributions measured at pulling rates ranging from 110-270 nm/s were analysed by both methods and averaged, yielding $\log k_{\text{off}} = -4.1 \pm 0.4 \text{ s}^{-1}$, $\Delta x^\ddagger = 1.9 \pm 0.2 \text{ nm}$, and $\Delta G^\ddagger = 49 \pm 6 \text{ kJ/mol}$ for this pseudoknot. There were no comparable measurements of ΔG^\ddagger for -1PRF pseudoknots until recently (de Messieres et al., 2014), but the values for k_{off} and Δx^\ddagger agree well with those for the IBV and telomerase pseudoknots,

$\log k_{\text{off}} \sim -4-5 \text{ s}^{-1}$ and $\Delta x^\ddagger \sim 1-2 \text{ nm}$ (36, 38). The small Δx^\ddagger is typical of RNA tertiary structures (Greenleaf et al., 2008; Liphardt et al., 2001; Neupane et al., 2011), differing from the larger values seen for secondary structure alone (Liphardt et al., 2001; Woodside et al., 2008).

Similar measurements were made for each pseudoknot in the panel. The FECs displayed the same characteristic behaviour as above (Figure 5.4), sometimes unfolding from the native structure, sometimes starting from an alternate conformation. For several of the pseudoknots (PT2G32, PEMV1, ScYLV WT and SRV1), ΔL_c values indicated that the pseudoknots were natively-folded (Figure 5.4, insets, and Figure 5.2) at the start of almost every FEC (Figure 5.4, black), as for ScYLV C27A. For MMTV, HERV, SARS, and VMV, however, many FECs were observed in which the ΔL_c from folded to unfolded was less than expected for the native state (Figure 5.4, blue), indicating that the pseudoknot started in an alternate conformation. The frequency with which unfolding occurred from such alternate states varied for different pseudoknots. The ΔL_c values expected from the native structure for each pseudoknot are listed in Table 5.1, along with the observed values.

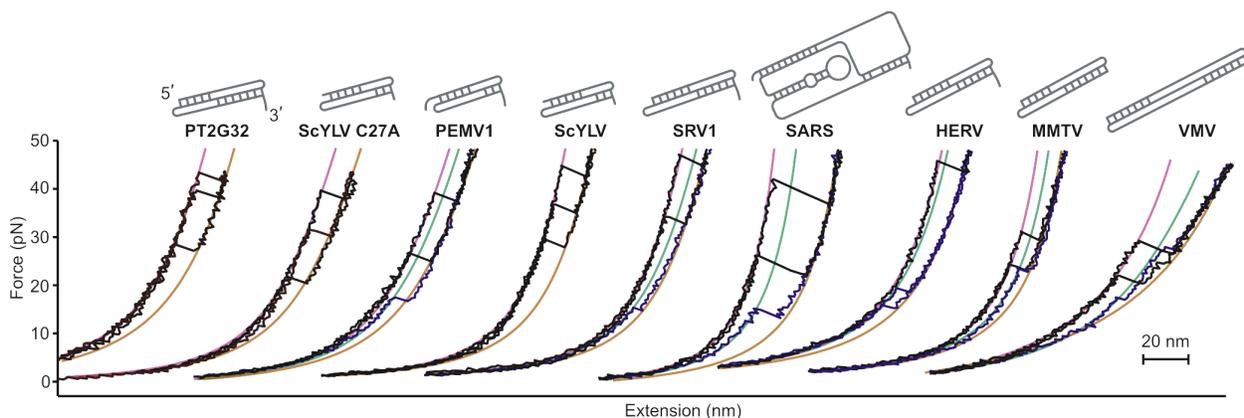


Figure 5.4 Representative FECs for the nine pseudoknots.

FECs from the different pseudoknots, arranged by increasing -1 PRF efficiency of the pseudoknot from left to right, show qualitatively similar unfolding. FECs in black show unfolding from the native structure; those in blue, from alternate structures. States were distinguished primarily by the contour length changes upon unfolding, using WLC fits (purple: native structure; brown: unfolded; green: alternate structure). Insets: secondary structures of the pseudoknots. 2D renderings generated with Pseudoviewer3 (Byun and Han, 2009; Ponty and Leclerc, 2015)

The unfolding force distributions were analysed for each pseudoknot as described above for ScYLV C27A, including only those FECs in which the pseudoknot was natively folded as determined by ΔL_c (Figure 5.5, grey). We first calculated the average unfolding force. Next, $p(F)$ was fit to Equation 3.8 (Figure 5.5, red) and the rates $k(F)$ were fit to equation 3.7, yielding the average values for k_{off} , Δx^\ddagger , and ΔG^\ddagger (Table 5.1). In the case of HERV and VMV, it was not possible to obtain sufficient FECs showing unfolding of the native pseudoknot structure from a given molecule before it broke, hence the force distributions could not be fit reliably to obtain landscape parameters. The results are summarised in Figure 5.6, plotting each quantity against the -1 PRF efficiency for the corresponding pseudoknot. No correlation is evident between the -1 PRF efficiency and the average unfolding force (Figure 5.6A); indeed, the highest average force (for HERV) and the lowest (for MMTV) occurred for the same -1 PRF efficiency of 20%, and all other pseudoknots unfolded in the range 30–40 pN. Similarly, -1 PRF efficiency was not correlated with any of the other parameters describing the mechanical unfolding: k_{off} (Figure 5.6B), Δx^\ddagger (Figure 5.6C), or ΔG^\ddagger (Figure 5.6D). The lack of correlation was confirmed by least-squares linear fits to the data in Figure 5.6, which in each case yielded a slope of zero (within error).

Pseudoknot	-1 PRF Efficiency (%)	ΔL_c (nm) complete unfolding	ΔL_c (nm) expected (native) ^h	ΔL_c (nm) alternate unfolding	Average unfolding force (pN)	$\log k_{off}$ (s ⁻¹)	Δx^\ddagger (nm)	ΔG^\ddagger (kJ/mol)
PT2G32	2 ^a	13.0 ± 0.6	13.8	8 ± 1	40 ± 2	-4.1 ± 0.2	1.6 ± 0.1	63 ± 10
ScYLV C27A	2 ^b	13.9 ± 0.7	14.2	9 ± 1	32 ± 2	-4.1 ± 0.4	1.9 ± 0.2	49 ± 6
PEMV1	9 ^c	14 ± 1	14.3	7 ± 1	31 ± 2	-3.5 ± 0.4	1.9 ± 0.2	39 ± 7
ScYLV WT	15 ^b	13 ± 1	14.6	10 ± 1	42 ± 2	-5.0 ± 0.3	1.7 ± 0.1	60 ± 4
SRV1	16 ^d	14.9 ± 0.6	15.6	9.8 ± 0.6	39 ± 2	-5.1 ± 0.3	1.9 ± 0.1	70 ± 20
SARS	17 ^e	32 ± 1	~34 ⁱ	21 ± 1	40 ± 2	-4.3 ± 0.4	1.7 ± 0.1	42 ± 5
MMTV	20 ^f	15 ± 1	16.3 ^j	9.3 ± 0.5	26 ± 3	-3.1 ± 0.4	2.0 ± 0.4	34 ± 5
HERV	20 ^a	16 ± 1	~16 ^k	8 ± 1	52 ± 3	n/a	n/a	n/a
VMV	28 ^g	27 ± 1	~27 ^l	19 ± 2	30 ± 2	n/a	n/a	n/a

Table 5.1 Summary of results.

The contour lengths from WLC fits to the FECs showed that complete unfolding matched the length expected from the native structure. An alternate structure sometimes formed with a different contour length. The parameters describing the resistance of the pseudoknots to mechanical unfolding are also listed. All errors

represent standard error on the mean. ^a (Wang et al., 2002). ^b (Cornish et al., 2005). ^c (Nixon et al., 2002). ^d (Michiels et al., 2001). ^e average (Park et al., 2011; Plant et al., 2005). ^f (Chamorro et al., 1992). ^g (Pennell et al., 2008). ^h Errors for expected ΔL_c are estimated as ± 0.2 nm based on the typical heavy-atom RMSD of RNA pseudoknot NMR structure bundles. ⁱ The end-to-end distance d_T for SARS was estimated at 6 nm, similar to IBV (Green et al., 2008). ^j d_T was taken from the structure of a related sequence optimized for NMR (PDB ID: 1RNK). ^k d_T was estimated as 4 nm. ^l d_T was estimated as 5 nm.

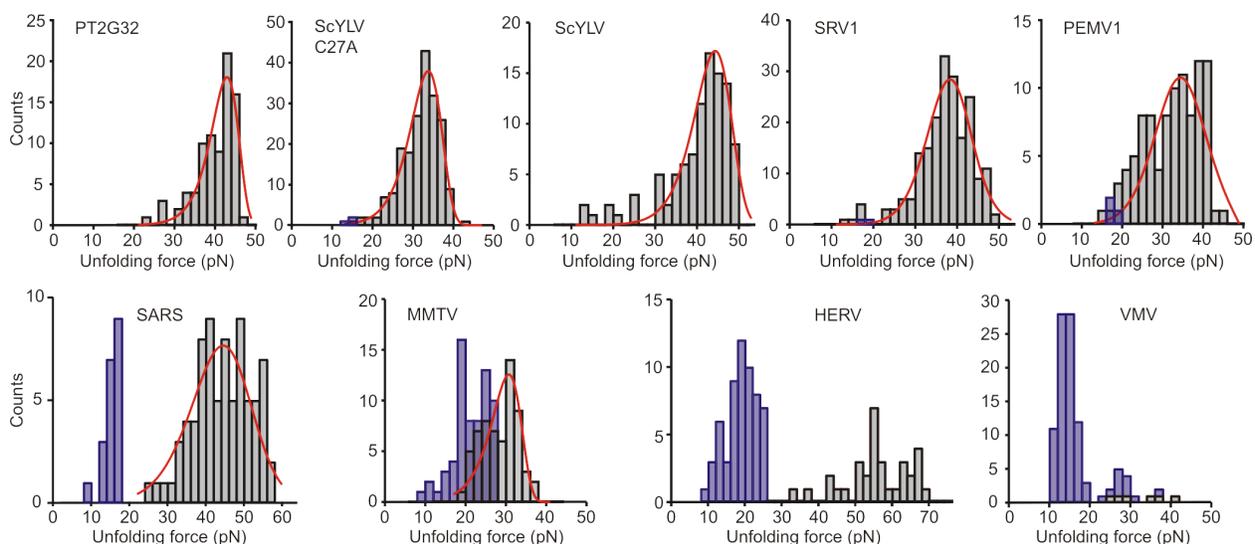


Figure 5.5 Representative unfolding force distributions.

The force distributions for unfolding the native structure of each pseudoknot (black) were fit by equation 3.8 (red) to obtain parameters describing the energy landscape for unfolding. For some pseudoknots, significant numbers of FECs unfolded from alternate structures (force distributions shown in blue). All distributions measured at ~ 270 nm/s pulling speed. No fits are shown for HERV and VMV because there was insufficient data to fit the unfolding force distributions for the native state.

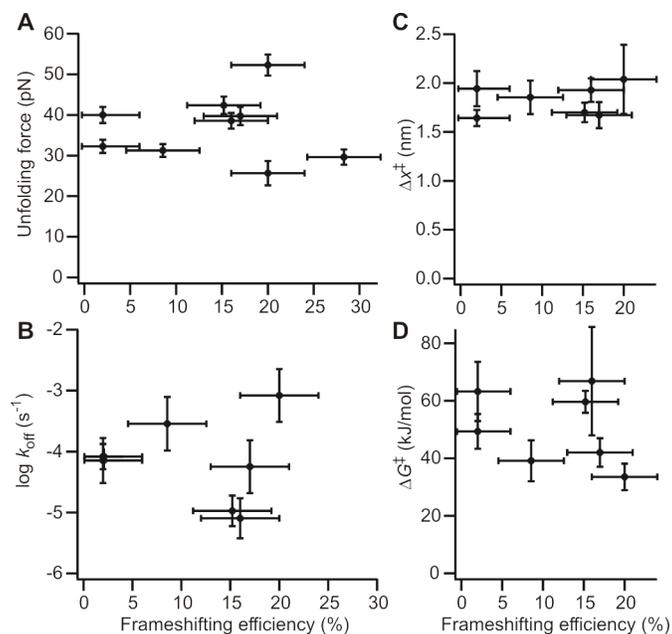


Figure 5.6 Frameshifting efficiency is not correlated with mechanical stability parameters.

The average unfolding force (A), unfolding rate at zero force, k_{off} (B), distance to the transition state, Δx^{\ddagger} (C), and height of the energy barrier, ΔG^{\ddagger} (D), were determined from the unfolding force distributions (as in Figure 5.4) and fits to equation 3.8. All quantities are effectively constant as a function of -1 PRF efficiency of the pseudoknot, indicating that there is no correlation and hence mechanical stability is not a primary determinant of -1 PRF efficiency. Error bars in the ordinate represent standard error; errors in the -1 PRF efficiency are an estimate of the variability in the values from the literature.

5.4 Pseudoknot alternate structures

Unexpectedly, however, a different property of the FECs was found to correlate well with -1 PRF efficiency: the tendency of the pseudoknot to fold into alternate structures. The percentage of FECs in which the pseudoknot started in an alternate state was determined from the number of curves in which the total ΔL_c during unfolding did not match the value expected for the native state. The pseudoknots exhibiting low or intermediate -1 PRF efficiency tended to fold reliably into the native structure, when allowing 3s for the RNA to refold at zero force between FECs; in contrast, pseudoknots stimulating higher -1 PRF efficiency tended to unfold more frequently from alternate structures (Figure 5.7, black). For the three pseudoknots with the highest -1 PRF efficiency (MMTV, HERV, VMV), the fraction of FECs starting from alternate structures did not change when increasing the refolding time between

measurements from 3 s to 10 s. Interestingly, this correlation extends to previous results for the telomerase pseudoknot, the only other pseudoknot for which alternate folding was similarly quantified (Chen et al., 2009). Although the telomerase pseudoknot is not naturally part of a frameshift signal, it stimulates -1 PRF efficiently and displays a correspondingly very high rate of folding into alternate conformations (Figure 5.7, blue).

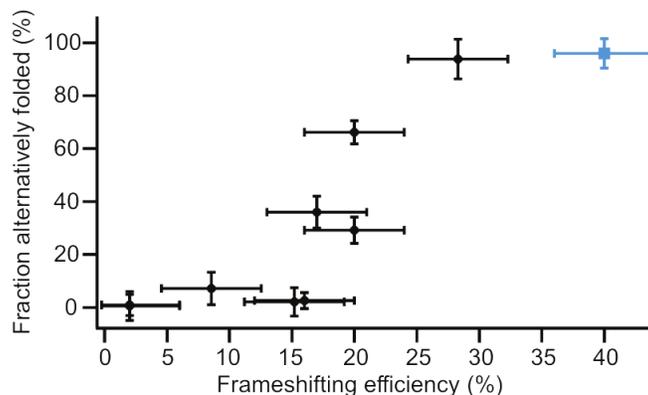


Figure 5.7 Frameshifting efficiency correlates with the formation of alternate structures.

The fraction of FECs demonstrating unfolding from an alternate structure is generally higher for pseudoknots causing higher -1 PRF efficiency. The correlation extends to the telomerase pseudoknot measured in (Chen et al., 2009) (blue square).

5.5 Observed contour length changes vs. frameshifting efficiency

Previous studies have provided contradictory evidence regarding correlations between -1 PRF efficiency and pseudoknot unfolding forces. The first to probe this question studied two near-wildtype pseudoknots derived from infectious bronchitis virus (IBV) having different unfolding forces. The one with higher unfolding force was found to be a more efficient -1 PRF stimulator (Hansen et al., 2007), but the pseudoknots were likely not natively folded (shown by shorter than expected ΔL_c values), making the interpretation uncertain. This study was also contradicted by measurements on a related set of pseudoknots, a “minimal” near-wild-type IBV pseudoknot with shortened loop 2 (Naphthine et al., 1999) and three different mutants, which did not find any clear correlation between unfolding force and -1 PRF efficiency (Green et al., 2008). A third study, on the beet western yellow virus (BWYV) pseudoknot and non-frameshifting mutants (White et al., 2011), again found little or no correlation between

unfolding forces and -1 PRF efficiency, although it too showed a discrepancy between observed and expected ΔL_c values.

In contrast, measurements on the telomerase pseudoknot and a set of mutations that systematically disrupted the 5 base-triples in the native structure (57), thereby reducing its mechanical stability, found a clear correlation between unfolding force and -1 PRF efficiency (Chen et al., 2009). The -1 PRF efficiency fell from ~ 40 – 50% for the wild-type pseudoknot to $\sim 0\%$ when all 5 base-triples were disrupted, dropping exponentially as the unfolding force fell from 50 pN to ~ 20 pN. These results showed that mutations mechanically destabilizing the pseudoknot can decrease the efficiency of -1 PRF. This conclusion was corroborated by our measurements on ScYLV and the C27A mutation, which negatively affects the stacking of triple base-pairs crossing the helical junction (Cornish et al., 2006), reducing the unfolding force from 42 ± 3 pN to 32 ± 3 pN and decreasing the -1 PRF efficiency markedly from 15% to 2%. However, it is unclear from such measurements whether the lowered unfolding force in fact causes the -1 PRF efficiency reduction or is merely an incidental byproduct; for example, the change in -1 PRF efficiency could be due to subtle structural changes and/or the prevention of specific contacts with the ribosome when key stabilising interactions are removed (Chen et al., 1996; Kang et al., 1996; Kim et al., 1999; Pallan et al., 2005).

Our measurements tested the correlation between -1 PRF efficiency and mechanical strength in a way that avoids these concerns; by determining whether wild-type pseudoknots producing different -1 PRF efficiencies also have different levels of resistance to unfolding. The result we found from our survey—that the -1 PRF efficiency was uncorrelated with unfolding force, rate, transition state location, and barrier height—indicates that resistance to mechanical unfolding is not, in fact, a key determinant of -1 PRF efficiency. Deleting critical tertiary interactions in a given pseudoknot may decrease both -1 PRF efficiency and unfolding force, but the mechanical strength is generally a poor predictor of -1 PRF efficiency when comparing different pseudoknots. Extending our results to include previous measurements on the high-efficiency telomerase (wild-type) and IBV (near wild-type) pseudoknots only reinforced this conclusion (Figure 5.8). Plotting the average unfolding force against -1 PRF efficiency for PT2G32, PEMV1, SRV-1, ScYLV, HERV, MMTV, VMV, SARS, IBV, BWYV, and telomerase

pseudoknots (11 in total) as in Figure 5.8, a straight-line fit to test for a linear correlation yields a slope of 0.1 ± 0.2 , *i.e.* zero. The average unfolding force is 36 pN, with a standard deviation of 10 pN.

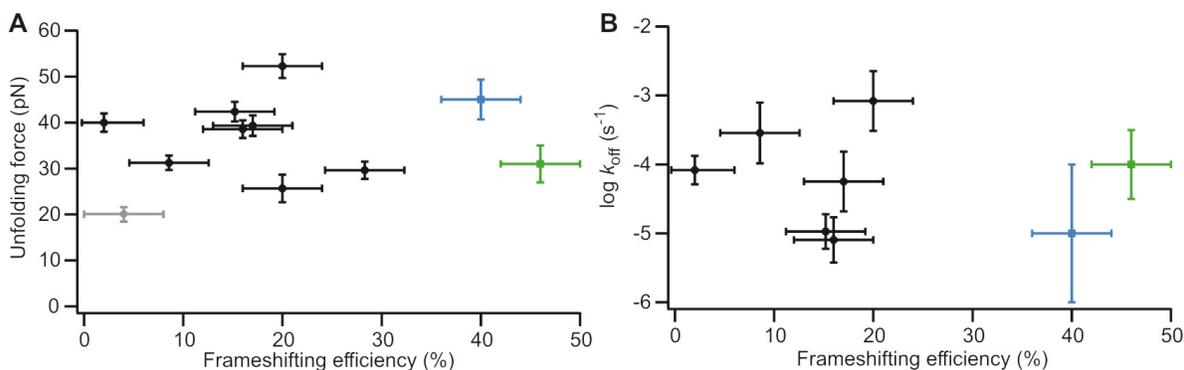


Figure 5.8 Frameshifting efficiency is not correlated with mechanical stability parameters.

The lack of correlation of -1 PRF efficiency with mechanical unfolding parameters found in Figure 5.6 was confirmed by the available data from other comparable studies of mechanical unfolding, including the IBV (cyan), telomerase (green), and BWYV (grey) pseudoknots (respectively from refs (Dam et al., 1992), (Brierley et al., 1992), (Jacks et al., 1988)). The low force for BWYV is in part due to a pulling rate 10–20 times slower than the other measurements. Uncertainties for the unfolding rates were not provided and hence are estimates only.

5.6 Possible correlations of frameshifting efficiency with mechanical parameters

We cannot rule out the possibility of a weak correlation between efficiency and mechanical unfolding parameters (force, rate,...) that would only be apparent with a larger sample size, nor that these parameters may play a significant role for certain subsets of pseudoknots that we have not yet tested. However, the results were inconsistent with simple mechanical strength being a principal determinant. Instead, there was a clear trend that pseudoknots stimulating high -1 PRF efficiency tend to fold more frequently into alternate structures.

Given that there is variability in the -1 PRF efficiency values reported in the literature, we tested the robustness of the correlation in Figure 5.7 to uncertainty in the reported -1 PRF efficiency. Monte Carlo simulations were conducted in which the true -1 PRF efficiency values for each pseudoknot were assumed to lie within a normal distribution around the reported values, with the widths of these distributions given by the uncertainties in the efficiency values. Simulations were run with the

uncertainties taken to be $\pm 4\%$ absolute, $\pm 25\%$ relative to the given efficiency value, and $\pm 40\%$ relative. In each simulation, Spearman rank correlation coefficient (r_s , for testing monotonicity) and Pearson correlation coefficient values (r_p , for testing linearity) were calculated for 100,000 combinations of the efficiency values chosen randomly within the distributions expected for each pseudoknot, and compared to the critical value indicating the 95% confidence level. With a $\pm 4\%$ absolute level of error, r_s was 88% likely to be higher than the critical values and r_p was over 99% likely to be higher. With a $\pm 25\%$ relative error on the efficiency values, r_s and r_p were respectively 84% and 95% likely to be higher than the critical values. Even if the error were as large as $\pm 40\%$ relative, the correlation coefficients were still respectively 64% and 76% likely to be higher than the critical values. The correlation in Figure 5.7 is quite strong: the Spearman rank correlation coefficient (testing monotonicity) is $r_s = 0.94$, whereas the Pearson correlation coefficient (testing linearity) is $r_p = 0.87$, both significantly above the 95% confidence level (0.63). It is also robust against uncertainty in the measurements of -1 PRF efficiency.

5.7 Possible underlying mechanisms behind the pseudoknot pulling results

How can these results be understood in terms of the role of pseudoknots in stimulating -1 PRF? This role is clearly more complex than simply providing a tuned mechanical resistance to unfolding by the ribosome. The pseudoknot is also likely acting as more than a passive roadblock for translation, given the suggestion from Figure 5.7 that conformational dynamics are important. Indeed, other evidence supports a role for conformational plasticity in the pseudoknot. For example, NMR measurements found the non-frameshifting pseudoknot PT2G32 to be more conformationally rigid than pseudoknots derived from SRV1 and MMTV, with a much lower breathing frequency of the base-pairs at the junction of the two stems (Wang et al., 2002), suggesting that a rigid structure prevents the pseudoknot from sampling a frameshift-competent conformation and leads to low -1 PRF efficiency. Recent work on the murine leukaemia virus pseudoknot has also suggested that a dynamic equilibrium exists between alternate structures, only one of which is active in recoding translation, leading to a recoding efficiency proportional to the time spent sampling the active conformation (Houck-Loomis et al., 2011). Additionally, it was shown that the correlation extends to PRF variations induced by a ligand that abolished PRF in SARS; in this case, the correlation was found to be perfectly linear (Ritchie et al., 2014). A final recent example involves CCR5-mediated -1 PRF directed by an mRNA pseudoknot: the

frameshifting is stimulated by at least two microRNAs, and experiments suggest that it manifests several distinct unfolding pathways when mechanically destabilized and that there is the formation of a triplex RNA structure to simulate the frameshifting event (Belew et al., 2014; Ritchie and Woodside, 2015).

The fact that the correlation in Figure 5.7 is not perfectly linear makes it unlikely that the alternate structures in our FECs are active (although they might lead to an active conformation when complexed with the ribosome); this view is supported by the fact that in several cases the alternate ΔL_c values are consistent with a hairpin formed from stem 1, which does not typically stimulate efficient -1 PRF. However, in the cases of HERV, MMTV, and VMV, the observed ΔL_c were also consistent with non-native, off-pathway secondary structures predicted by *mfold* (Zuker, 2003), such as an alternate hairpin in the HERV and MMTV pseudoknot sequence, or extended versions of stem 1 in the MMTV and VMV sequences. In the case of ScYLV, the length change was consistent with a hairpin based on stem 2, but the unfolding force was too high for such a marginally-stable structure (Woodside et al., 2006b). In the case of SARS, the observed ΔL_c is too large for hairpin structures based on either stem 1 or stem 2, and hence must represent some other structure. For VMV, 2 different alternate structures likely formed, as suggested by an extra peak in the force distribution (Figure 5.5), but with similar ΔL_c values. It appears more likely that the correlation is an indirect reflection of the relevant behaviour, with the alternate structure formation acting as a proxy for the property that does determine frameshifting efficiency.

One possibility is that the conformational fluctuations themselves, rather than specific structures, play an important role. It has been suggested that the ribosome senses the tension that it actively generates in the mRNA as structure is unfolded (Qu et al., 2011). A dynamic conformational equilibrium might then trigger frameshifting by causing fluctuations in this tension which are communicated to the tRNA-mRNA complex, similar to the previous proposal that refolding of a partially-unfolded pseudoknot during accommodation might induce a frameshift by pulling back on the mRNA (Plant and Dinman, 2005). Measurements of pseudoknot extension under constant tension near the average unfolding force do show a dynamic conformational equilibrium. In the case of the telomerase (Chen et al., 2007), IBV (Green et al., 2008), and SARS (Figure 5.9) pseudoknots, the structural fluctuations

occurred on the ms to s timescale, likely rapid enough to affect the ribosome while it is paused at the slippery sequence. In the case of HERV, however, which also stimulates -1 PRF efficiently, the fluctuations were very slow (Figure 5.10), suggesting that this explanation is incomplete. Complicating the picture, the ribosome is known to be an active helicase that interacts with the mRNA structure it is unwinding to facilitate melting (Qu et al., 2011). Such interactions may play an important role by biasing the dynamic equilibrium in favour of certain structures or speeding up the equilibration rates, but they were not probed in our measurements, which lack the ribosome.

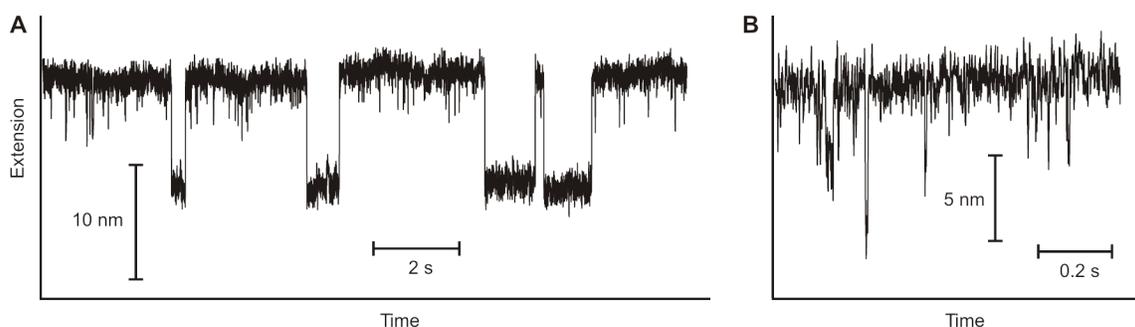


Figure 5.9 Structural dynamics of SARS pseudoknot at constant force.

When held under constant tension, the SARS pseudoknot fluctuates dynamically between multiple structures at different extension values. Structural fluctuations occur on multiple timescales, from seconds (A) to milliseconds (B). Data sampled at 20 kHz at a force of ~ 15 pN, median filtered in a 5 ms window.

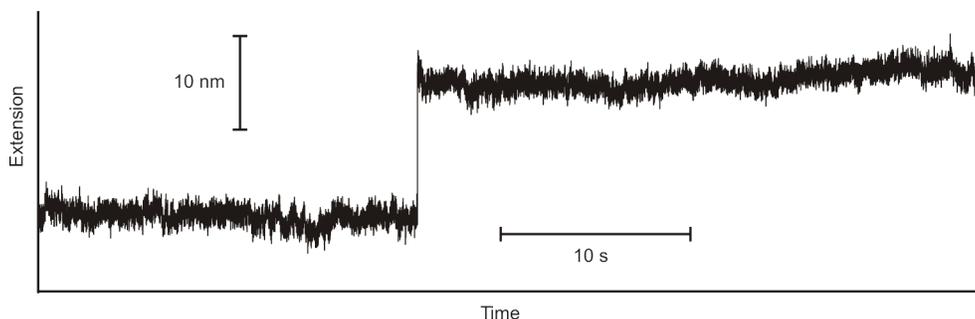


Figure 5.10 Structural dynamics of HERV pseudoknot at constant force.

Under constant tension, the HERV pseudoknot undergoes very slow dynamics. Seen here is a single unfolding transition. Data sampled at 20 kHz at a force of ~ 30 pN, median filtered in a 5 ms window.

A variety of evidence indicates a possible role for specific interactions with the ribosome in -1 PRF. Structural and functional studies suggest that triplex structures and unpaired, exposed loop nucleotides may make or direct specific contacts to the ribosome that help determine -1 PRF efficiency, explaining why efficiency is reduced by removing or altering these structures (Chen et al., 2009; Chen et al., 1995; Cornish et al., 2005; Kim et al., 1999; Liphardt et al., 1999; Michiels et al., 2001; Nixon et al., 2002; Olsthoorn et al., 2010; Shen and Tinoco, 1995; Su et al., 1999). Footprinting analyses of frameshift signals complexed with prokaryotic ribosomes indicate that specific contacts are indeed made with the pseudoknot when the slippery sequence is in the ribosomal decoding centre (Mazauric et al., 2009), but the details of these interactions and their role in regulation of -1 PRF efficiency are not well understood. Proteins at the mRNA entry tunnel of the ribosome, which could interact with the pseudoknot, have also been implicated in regulating -1 PRF efficiency (Kirthi et al., 2006). Again, however, we do not probe such interactions here. Programmed -1 frameshifting is clearly a complex phenomenon, regulated by many factors involving both the mRNA and the ribosome. The recent development of single-molecule assays of ribosome translocation along mRNA (Qu et al., 2011; Wen et al., 2008) holds out the promise of directly observing interactions between the pseudoknot and ribosome during the actual frameshifting event, leading to a more complete understanding of -1 PRF mechanisms.

Chapter 6: Folding transition path times for nucleic acids

We next turn from studying how the properties of the folding relate to function, to examining the physics of the folding in greater detail. In this chapter, I describe studies of the transition paths that constitute the most important part of the folding trajectories: the part where the molecule moves over the barrier in the energy landscape. The transition path time results from the energy landscape analysis of nucleic acids (NAs) and its direct measurement is presented. They are all determined from single-molecule trajectories recorded with optical tweezers. Much of the material in sections 6.2 through 6.3 was published as K. P. Neupane, D. B. Ritchie, H Yu, D. A. N. Foster, F. Wang, M. T. Woodside “Transition path times for nucleic acid folding determined from energy landscape analysis of single-molecule trajectories” *Phys. Rev. Lett.*, 109: 068102 (2012). For this paper, DBR, DANF and FW produced the samples, KN, DBR, HY, and DANF performed the experiments, KN, DBR, and MTW worked on the analysis. The remainder of the chapter is not yet published. To make direct measurements of the DNA hairpin transitions, the time resolution of the trap was improved by increasing the stiffness and recalibrating the instrument (see appendix C). The direct measurements of the transition times (6.4 and 6.5) have been submitted as: K. Neupane, D. A. N. Foster, D.R. Dee, H. Yu, F. Wang, M. T. Woodside “Direct observation of transition paths during the folding of proteins and nucleic acids” For this more recent work, FW prepared the DNA hairpin samples, multiple authors prepared the prion samples (Yu, 2013; Yu et al., 2012a; Yu et al., 2012b), KN and DANF performed the DNA hairpin measurements, while multiple authors including DRD and HY measured the prion samples, the analysis was completed by KN and MTW.

6.1 Transition paths

As we saw in the introduction, a quantitative, microscopic description of how biological macromolecules fold into complex three-dimensional structures remains one of the grand challenges of biophysics. The folding trajectory, from the unfolded state to the final folded state, has a particular characteristic time, but the actual time for a folding molecule to transition over a barrier, the transition time, is of great importance, as it contains the key microscopic information about the folding reaction mechanisms (see Figure 6.1). Ensemble biochemical and biophysical measurements have provided

significant insights into the folding problem (Buchner and Kiefhaber, 2005), complemented in recent years by single-molecule approaches offering ever more detailed pictures of the microscopic behaviour (Borgia et al., 2008; Woodside et al., 2008). Such measurements, however, have invariably concentrated on characterizing stable or metastable states and their lifetimes, which range from milliseconds to minutes (Buchner and Kiefhaber, 2005; Thirumalai and Hyeon, 2005), rather than the paths taken between the states and the duration of the structural transitions. The transitions themselves are extremely challenging to observe experimentally because they are very brief (Chung et al., 2009; Chung et al., 2012; Lee et al., 2007). Until recently, only all-atom simulations could provide insight into the mechanistic details of the transition paths (Best, 2012; Bolhuis et al., 2002; Lindorff-Larsen et al., 2011). Moreover, such transition paths (see Figure 6.1) are inherently a property of single molecules, presenting additional technical challenges for observing them. Thus, advances in high-resolution single-molecule techniques are uniquely placed to enable experimental characterization of transition paths.

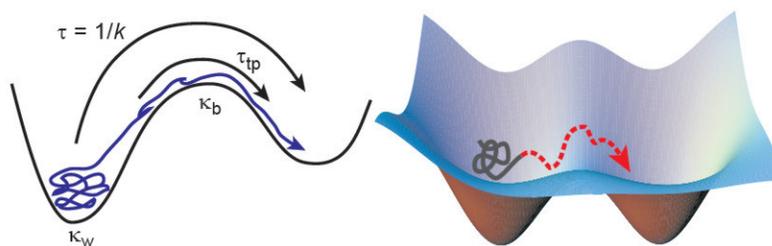


Figure 6.1 Schematic free-energy profile for a folding reaction.

The folding rate given by Kramers' theory (k) is set primarily by the length of time spent diffusing/fluctuating within one of the potential wells. The transition path time required to cross the barrier (τ_{tp}) is much shorter. κ_b : curvature of the barrier; κ_w : curvature of the well.

Kramers' theory was introduced in chapter 3, where in Figure 3.3 (Hänggi et al., 1990), the curvatures are represented by κ_w for the potential well, κ_u for the unfolded well and κ_b for the barrier. The time to traverse from κ_u to κ_b is the inverse of the rate $\tau = 1/k$. Most of the time spent between folding/unfolding transitions, is however, mostly taken up by diffusion within the potential wells; the actual transition time, τ_{tp} , is generally much faster. Here τ_{tp} represents the average transition time, while t_{tp} represents an individual transit. Because of the extreme technical difficulty of measuring τ_{tp} , only a handful of

measurements exist. A Förster resonance energy transfer (FRET) study of folding in one RNA molecule estimated τ_p for folding at 240 μs , based on time-apertured photon cross-correlation of the fluorescence, but was unable to estimate it for unfolding despite the expected symmetry (Lee et al., 2007). More recently, the analyses of photon statistics from high time-resolution FRET determined τ_p for two-state folding in proteins by analysing the trajectories photon by photon as the proteins moved across the energy barrier (Chung and Eaton, 2013; Chung et al., 2012). The results, $\tau_p \sim 2 \mu\text{s}$ and $< 10 \mu\text{s}$ for two different proteins, were considerably faster than the earlier RNA study and agreed well with expectations based on the $\sim 0.1\text{--}1 \mu\text{s}$ “speed limit” for protein folding (from unfolded state to native state) (Kubelka et al., 2004). In a similar vein, an upper bound of 2.5 μs for the transition time for a DNA hairpins with a short stem has been published recently (Truex et al., 2015).

6.2 Average transition times from energy landscape reconstructions

The first attempts at measuring τ_p directly from force spectroscopy, by measuring the time required to cross the distance between folded and unfolded states (Figure 6.2) in a constant-force trajectory, was limited by the time response of the force probe. By holding a reference construct consisting just of handles in the traps, then jumping the traps apart suddenly, and then measuring the resulting motion of the beads (Figure 6.2 c), the response time under these conditions was found to be about 50 μs , not as fast as needed. (Later (in 6.7) we increased the stiffness of the traps, improving the time resolution of the instrument.) To see if the finite transition time between the folded and unfolded states could be observed directly, the extension of a hairpin held under constant tension was measured at a bandwidth of 50 kHz (Figure 6.2b). The effects of Brownian noise, which might obscure the transition, were reduced by aligning each 1-ms segment containing a transition of a given type (folding or unfolding) on the centre of the transition (Figure 6.2c, red) and averaging all the segments (Yu et al., 2012a). The average of 2,529 unfolding transitions (Figure 6.2c, black) is the same as the time-reversed average of 2,529 folding transitions (Figure 6.2c, yellow), indicating that the transition is symmetric. The total time for the transition was then estimated by measuring the time required to move between the inflection points of the Gaussian extension distributions of the two states (Figure 6.2c, blue lines). The result from averaging

14,766 transitions was $49 \pm 3 \mu\text{s}$, 5 times faster than the transition time found for RNA using FRET (Lee et al., 2007).

Such measurements do not truly give the trajectory of the hairpin during the transitions, however, because the optical trap filters the hairpin trajectory, owing to the finite time required for the beads and handles to respond to motions of the hairpin (Yu et al., 2012a). To estimate the time resolution of the trap, we measured the “transition time” needed to move beads tethered by dsDNA without any hairpin through a distance similar to the extension change of the hairpin (Figure 6.2c, cyan). The transition time for this reference construct was identical to the transition time measured with the hairpin, $50 \pm 5 \mu\text{s}$ indicating that τ_{tp} for the hairpin must be considerably smaller. Hence the 50- μs transition time obtained directly from the trajectories represents an upper bound for the transition time.

We therefore developed an alternate approach for determining τ_{tp} , making use of the ability of smFS to measure energy landscapes. Two methods were applied: (i) constant-force extension trajectories were used to determine both the full landscape profile (by an inverse Boltzmann transform) (Woodside et al., 2006a) and the kinetic rates, the diffusion coefficient D was found from equation 3.4, and then τ_{tp} was calculated from equation 3.5 (ii) FECs were used to determine the landscape parameters and rates from the theory of Dudko *et al.* (Dudko et al., 2006, 2008) (approximating the landscape profile as linear-cubic), the value for D implied by these results was calculated from equation 3.4, and τ_{tp} was again found from equation 3.5.

We applied both methods to determine τ_{tp} for DNA hairpins of varied size and sequence, and then the Dudko methods to various RNA pseudoknots, and the *add riboswitch* aptamer. These choices allowed us to investigate the effects of different properties on τ_{tp} , such as molecule size, secondary versus tertiary structure, and ligand-induced interactions. In each case, the nucleic acids being studied were attached to kb-long ds “handles” held by beads in a high-resolution dual-beam optical trap (Neupane et al., 2011) (Figure 6.2a). Non-equilibrium force-extension curves (FECs) were measured at pulling rates of 10–220 nm/s and trap stiffness of 0.3–0.9 pN/nm, sampling data at 20 kHz after analog filtering online at

the Nyquist frequency (10kHz). Equilibrium measurements were sampled at 50 kHz with a trap stiffness of 0.3 pN/nm, using a passive force clamp to maintain constant force (Greenleaf et al., 2005).

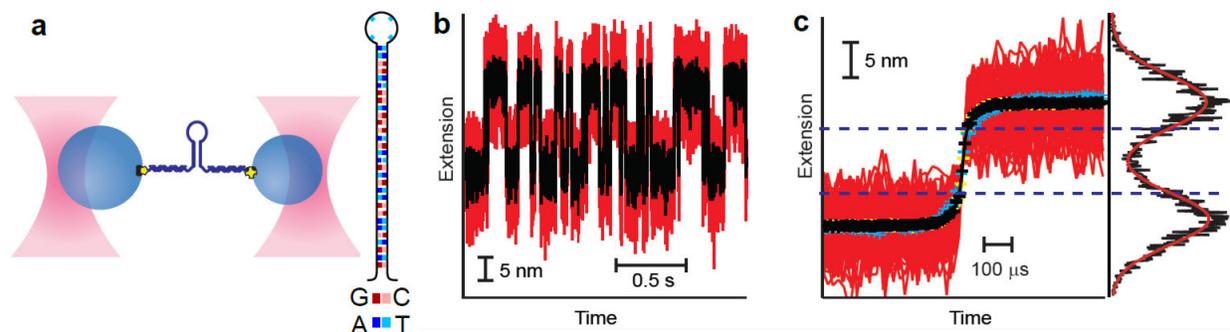


Figure 6.2 Extension trajectories of DNA hairpins.

(a) Single nucleic acid molecules attached to duplex handles are held under tension between beads in two traps. A DNA hairpin attached to handles (purple) was linked to beads (blue) held in laser traps (pink) applying tension. [Inset: hairpin sequence for 30R50/T4.] (b) The extension of hairpin 20TS06/T4 as a function of time at a constant force shows sudden changes as the hairpin folds and unfolds (red: data sampled at 50 kHz; black: filtered at 10 kHz). (c) One-ms records straddling the transitions (red) were aligned and averaged to reduce Brownian noise. The averages of 2 529 unfolding (black) and refolding (yellow, time-reversed) transitions overlap with each other and with the instrument response signal from fast ($< 1 \mu$ s) motions of the traps (cyan), indicating that the apparent transition time of $\sim 50 \mu$ s between the inflection points of the extension probability distributions (blue lines) is instrument-limited.

6.2.1 Transition times for DNA hairpins

We first looked at four DNA hairpins (30R50/T4, 20TS06/T4, 20TS10/T4, 20TS18/T4) whose folding under tension as cooperative two-state systems has been extensively characterized (Engel et al., 2014; Gupta et al., 2011; Woodside et al., 2006a; Woodside et al., 2006b). We removed the resolution-limiting instrumental compliance effects by deconvolution (Woodside et al., 2006a) from the energy landscape profiles reconstructed from an inverse Boltzmann transform of the extension probability distribution. Measuring the barrier height and potential well curvatures from the energy profiles (Figure 6.3a, b) and using the rates measured directly from the trajectories, we found D and hence τ_{tp} for each hairpin from equations 3.4 and 3.5 (Table 6.1). D is similar for all hairpins and lies within the range 10^7 to $10^5 \text{ nm}^2/\text{s}$ (10^{-11} – $10^{-13} \text{ m}^2/\text{s}$) inferred from measurements on single-stranded DNA and DNA hairpins (Ansari et al.,

2001; Kaji et al., 2009; Wang and Nau, 2003), confirming the validity of the energy landscape analysis approach. The values for τ_p , $\sim 6\text{--}30\ \mu\text{s}$, are all lower than the upper bound established by direct examination of the folding trajectories. They are similar in magnitude to τ_p for small proteins (Chung et al., 2012) but somewhat higher, likely due to the different folding mechanism in nucleic acids compared to proteins (Thirumalai and Hyeon, 2005). For each hairpin, τ_p for folding and unfolding agreed well (e.g. $\tau_p = 30 \pm 6\ \mu\text{s}$ for folding 30R50/T4, $\tau_p = 33 \pm 8\ \mu\text{s}$ for unfolding), as expected, hence only the averaged value was reported in Table 6.1.

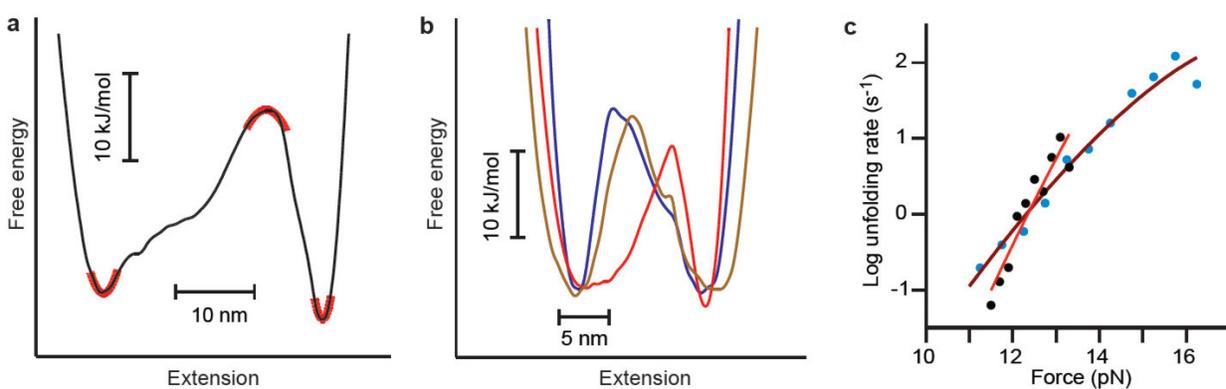


Figure 6.3 Energy landscape analysis of DNA hairpins.

(a) The deconvolved landscape profile of hairpin 30R50/T4 (black) allows the barrier and well curvatures to be measured (red: quadratic fits), as well as the barrier heights, thereby determining D and τ_p . (b) Examples of deconvolved landscape profiles for hairpins 20TS06/T4 (blue), 20TS10/T4 (brown), and 20TS18/T4 (red). (c) Fitting the force dependent unfolding rates for 30R50/T4 (black) and 20TS06/T4 (cyan) to equation 3.7 (red and brown, respectively) reveals that the hairpins have very different unfolding rates at zero force.

It can be technically very challenging to reconstruct full landscape profiles, thus we analysed the unfolding force distributions and/or force-dependent rates with a more limited knowledge of the key parameters of the landscape and compared those results to the results for D and τ_p from above. By making some reasonable assumptions about the shape of the profile, for example, assuming a linear-cubic (three unique roots) potential (Dudko et al., 2006) (locally quadratic in the well and the barrier), D

is given in terms of ΔG^\ddagger , k_{off} (the unfolding rate at zero force), and Δx^\ddagger (the distance to the barrier from the folded state) by equation 1, whereas τ_{ip} can be estimated from equation 3.5 assuming that $\kappa_w \approx \kappa_b$ (as is often done (Chung et al., 2009; Hummer, 2004)).

$$D \approx \frac{k_{\text{off}} (\Delta x^\ddagger)^2}{\Delta G^\ddagger / k_B T} \exp\left(\frac{\Delta G^\ddagger}{k_B T}\right) \quad (6.1)$$

The three parameters needed for these calculations can be found for a linear-cubic profile by analysing the distribution of unfolding forces or kinetics in FECs (Dudko et al., 2006, 2008) (see equations 3.7 and 3.8).

Molecule	D (nm ² /s)	τ_{ip} (μ s)
DNA hairpins		
30R50/T4	$4.6 \pm 0.5 \times 10^5$	31 ± 5
20TS06/T4	$5 \pm 3 \times 10^5$	26 ± 9
20TS10/T4	$1 \times 10^{5 \pm 0.4}$	$1.6 \times 10^{1 \pm 0.4}$
20TS18/T4	$2 \times 10^{5 \pm 0.4}$	$6 \times 10^{0 \pm 0.4}$
Pseudoknots		
MMTV	$4 \times 10^{2 \pm 1}$	$5 \times 10^{2 \pm 1}$
PEMV1	$6 \times 10^{2 \pm 1}$	$3 \times 10^{2 \pm 1}$
ScYLV	$6 \times 10^{4 \pm 1}$	$3 \times 10^{0 \pm 1}$
ScYLV C27A	$8 \times 10^{3 \pm 1}$	$2 \times 10^{1 \pm 1}$
PT2G32	$1 \times 10^{6 \pm 2}$	$5 \times 10^{-2 \pm 2}$
<i>add</i> riboswitch		
without adenine	$2 \times 10^{5 \pm 0.3}$	$5 \times 10^{0 \pm 0.3}$
with adenine	$2 \times 10^{4 \pm 1}$	$6 \times 10^{1 \pm 1}$

Table 6.1 Diffusion constants and transition path times from energy landscape analysis.

For the hairpins, D and τ_{ip} were calculated for folding and unfolding separately, to ascertain that they were symmetric with respect to the transition direction, then averaged. Uncertainties represent standard error on the mean.

To confirm that this indirect landscape analysis gives results consistent with the analysis of the full landscape profiles, 2,048 FECs were measured for hairpin 30R50/T4 and 2,966 FECs for hairpin 20TS06/T4. The landscape parameters obtained from fitting the force-dependent unfolding rates (Figure

6.3c, black: 30R50/T4, cyan: 20TS06/T4) to equation 3.7 and the distribution of unfolding forces to equation 3.8 (Table 6.2) yielded $D = 4 \times 10^{5 \pm 1} \text{ nm}^2/\text{s}$ and $2 \times 10^{6 \pm 1} \text{ nm}^2/\text{s}$ for hairpins 30R50/T4 and 20TS06/T4, respectively. The uncertainty is larger than when directly analysing the full profile, but these values nevertheless agree well with the values listed in Table 6.1. Equation 3.5 yielded $\tau_p = 2 \times 10^{1 \pm 1} \mu\text{s}$ and $1 \times 10^{0 \pm 1} \mu\text{s}$ for hairpins 30R50/T4 and 20TS06/T4, respectively. Again, these values agree well (within error) with those listed in Table 6.1, indicating that the two methods indeed give consistent results.

Molecule	$\ln k_{\text{off}} (\text{s}^{-1})$	$\Delta x^\ddagger (\text{nm})$	$\Delta G^\ddagger (k_B T)$
Hairpins			
30R50/T4	-48 ± 1	19.2 ± 0.3	59 ± 2
20TS06/T4	-21 ± 2	9 ± 1	35 ± 2
Pseudoknots			
MMTV	-7 ± 1	2.0 ± 0.4	14 ± 2
PEMV1	-8 ± 1	1.9 ± 0.2	16 ± 3
ScYLV	-11.5 ± 0.7	1.7 ± 0.1	24 ± 2
ScYLV C27A	-9 ± 1	1.9 ± 0.2	20 ± 2
PT2G32	-9.4 ± 0.5	1.6 ± 0.1	26 ± 4
<i>add</i> riboswitch			
without adenine	-11.0 ± 0.2	6.1 ± 0.5	20.5 ± 0.7
with adenine	-15 ± 1	6.2 ± 0.6	27 ± 3

Table 6.2 Energy landscape parameters for NAs found from fits to $p(F)$ and $k(F)$.

Results from the two fits were averaged. Errors represent the standard error on the mean over all molecules.

6.2.2 Diffusion and transition times from pseudoknots and the *add* riboswitch

Hairpins contain only secondary structure. To investigate how tertiary structure affects D and τ_p , we also studied RNA pseudoknots, which consist of two intercalated stem-loop structures (Brierley et al., 2007b). Five different pseudoknots were measured, from MMTV, PEMV1, ScYLV and its C27A mutant, and PT2G32. In each case, the full free-energy profile could not be obtained from constant-force measurements, but the key parameters needed to estimate D and τ_p could still be found from FECs, as for the DNA hairpins. 300–600 FECs were measured for each of the pseudoknots. The landscape parameters

obtained from fitting the distribution of unfolding forces to equation 3.8, as illustrated for representative distributions (Figure 6.4a-c), and also from fitting the complementary force-dependent lifetimes derived from these histograms (Dudko et al., 2008), are listed in Table 6.2. The resulting diffusion constants and transition times (Table 6.1) are more variable than for the hairpins, owing to larger uncertainties. However, averaging the values over all the pseudoknots (which have identical topologies and very similar sizes, 30–34 nucleotides) yields $\tau_{tp} = 1 \times 10^{1 \pm 0.7} \mu\text{s}$, similar to the hairpins, and $D = 1 \times 10^{4 \pm 0.7} \text{ nm}^2/\text{s}$, slightly lower than for the hairpins.

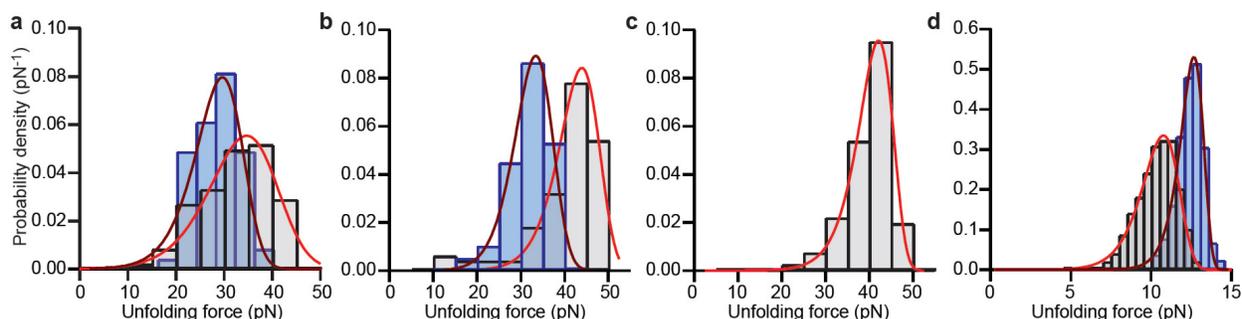


Figure 6.4 Unfolding force distributions for pseudoknots and riboswitches.

D and τ_{tp} were determined from the key landscape parameters obtained by fitting the unfolding force distributions from FECs to equation 3.8. (a) MMTV (grey) and PEMV1 (blue) pseudoknots. (b) ScYLV (grey) and C27A mutant (blue) pseudoknots. (c) PT2G32 pseudoknot. (d) The add adenine pseudoknot with (blue) and without (grey) ligand bound.

To probe effects arising from different fold topologies and different tertiary interactions, we also investigated τ_{tp} for another class of RNA: the aptamer domain from a bacterial riboswitch. Riboswitch aptamers bind ligands that can induce structural changes and thereby alter the energy barrier height and unfolding rate (Greenleaf et al., 2008; Neupane et al., 2011). Unfolding force distributions from FECs of the *add* adenine riboswitch aptamer were analysed as for the pseudoknots (Figure 6.4d), based on 5,200 FECs measured without the adenine ligand bound and 3,000 with ligand bound. The fit results (Table 6.2) yielded values for D and τ_{tp} that are similar (within error) to those found for the hairpins and the pseudoknots (Table 6.1). They are also the same (within error) whether ligand is bound or not, even though ligand binding changes the barrier height significantly ($7 k_B T$).

We note that the average transition times found for the different types of nucleic acids investigated here are roughly the same, on the order of 10 μ s. This is ~ 10 times faster than the only previous estimate of τ_{tp} for a nucleic acid (Lee et al., 2007), but comparable to τ_{tp} from measurements of small proteins (Chung et al., 2012). Interestingly, τ_{tp} for nucleic acids is slightly higher than for proteins: the average over all the different molecules is $16 \pm 8 \mu$ s, compared to $\sim 2 \mu$ s for the WW domain of the formin-binding protein (Chung et al., 2012). This difference likely reflects the different microscopic mechanisms for folding in nucleic acids compared to proteins (Thirumalai and Hyeon, 2005). Nucleic acid duplex formation is often modeled as a zippering process (Ansari et al., 2001; Cocco et al., 2003; Pörschke, 1974), which might be expected to lead to a τ_{tp} that depends linearly on the duplex length. This is indeed what is seen comparing τ_{tp} for hairpins with different stem lengths: $\tau_{tp} = 31 \pm 5 \mu$ s with a 30-basepair (bp) stem, whereas $\tau_{tp} = 16 \pm 4 \mu$ s on average for 20-bp stems. The ratio of transition times, 1.9 ± 0.6 , thus agrees well with the expected ratio of 1.5. The correlation of τ_{tp} with duplex length seems to extend to the RNA, as well, although the experimental uncertainty is sufficiently large that the comparison can only be made for the aptamer without ligand bound. The transition state for aptamer unfolding involves unfolding helix P1 (Neupane et al., 2011), which is only 9 bp long, and τ_{tp} for the aptamer is correspondingly shorter, at $\sim 5 \mu$ s. A plot of τ_{tp} against the length of the duplex being unfolded (Figure 6.5) implies an estimate for the zippering time of $\sim 0.9 \mu$ s/bp, in reasonable agreement with previous estimates of ~ 0.1 – 0.3μ s/bp from temperature-jump measurements of double helices (Pörschke, 1974) and modeling of hairpin folding rates under tension (Cocco et al., 2003). Our results disagree with a different estimate of ~ 1 – 20 ns/bp from modeling hairpin folding rates under temperature jumps (Kuznetsov and Ansari, 2012), however this latter estimate is unable to account for the observed τ_{tp} length-dependence. We note that recent theoretical work suggests that the length-dependence for helix zippering should in fact be superlinear, with an exponent of ~ 1.6 (Frederickx et al., 2014), but the experimental uncertainty is currently too large to distinguish between linear and superlinear cases.

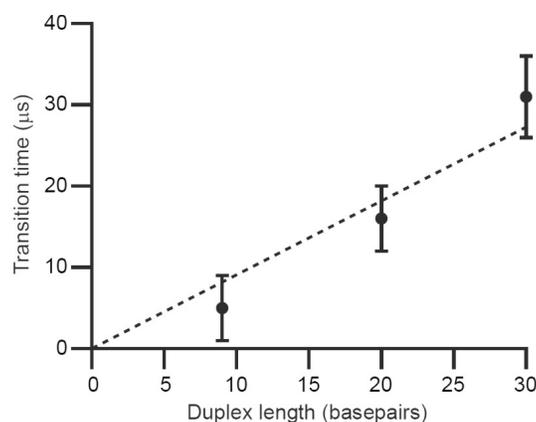


Figure 6.5 Stem-length dependence of τ_p .

The transition time for unfolding the DNA hairpins (30- and 20-bp stems) and the riboswitch aptamer (the critical helix unfolded to reach the transition state has 9 bp) varies linearly with the length of the helix stem.

6.3 Comparing the transition time and diffusion coefficient across nucleic acids

Despite the noticeable stem-length dependence of τ_p , the more notable fact is that τ_p is very similar for all the molecules, despite unfolding rates that differ by many orders of magnitude. The unfolding rates at zero force ranged from $\sim 10^{-3} \text{ s}^{-1}$ for the MMTV pseudoknot to $\sim 10^{-21} \text{ s}^{-1}$ for hairpin 30R50/T4, or 18 orders of magnitude (due largely to differences in the barrier heights, $\sim 45 k_B T$). A similar effect was also seen for proteins (Chung et al., 2012), although over a much smaller range of rates. Intuitively, this can be pictured in terms of the molecule trying to jump over the barrier with a certain initial “velocity” across the landscape: to get over a higher barrier requires a higher initial velocity (which fewer molecules have, hence lowering the total rate), but the time taken to get over the barrier is hardly changed, analogous to what happens to a projectile thrown in a parabolic trajectory. Mathematically, the effect is explained by the weak barrier-height dependence of τ_p in equation 3.5.

Most remarkably, there is no significant difference in τ_p for molecules with very different topologies, despite the fact that folding mechanisms are believed to be determined primarily by the topology of the native fold (Baker, 2000). Single stem-loops (hairpins), intercalated stem-loops (in pseudoknots), and triple-helix junctions (in the aptamer) all produce transition times on the order of 10 μs , suggesting that τ_p is relatively insensitive to the details of the folding mechanism. The values of D for the RNA structures

are on average somewhat lower than for the DNA hairpins, $2 \times 10^{4 \pm 0.6}$ vs $3 \times 10^{5 \pm 0.2}$ nm²/s. Since D is lower for a rougher landscape, this suggests that the tertiary interactions in the pseudoknots and riboswitch aptamer roughen the landscape slightly. Assuming a random roughness distribution (Zwanzig, 1988), this additional roughness is approximately $2 \pm 1 k_B T$ on average. We note that using energy landscape analysis to obtain τ_p as described here does, of course, depend on the validity of equation 3.5. Although this equation is expected to be rigorous, since it is derived from the well-established Kramers theory, it has not yet been validated formally by comparing predicted τ_p values to those measured directly.

6.4 Direct measurement of transition times in DNA hairpins

From the estimates of τ_p obtained by energy landscape analysis, we can see that a relatively modest improvement in the response time of the instrument would allow transition times to be measured directly. To achieve this increased time resolution, we moved away from constant-force measurements using the passive force clamp, because the low system stiffness (owing to the requirement of operating one trap in the zero stiffness regime) reduced the time response. Instead, we measured at constant trap separation, with the trap stiffnesses set as high as possible (0.63 and 1.12 pN/nm) (see appendix C for more detail). Under these conditions, the time response improved over 5-fold from the previous work, to about 8 μ s (Figure 6.7, Figure 6.8). We first studied the two-state DNA hairpin 30R50/T4 (Figure 6.2a, inset), in the two stiffer optical traps (Figure 6.2a) measured in equilibrium (near $F_{1/2}$).

From the equilibrium trajectories of the extension of the molecule (Figure 6.4c), individual transitions (Figure 6.6a, red) were identified as those crossing between boundaries that defined the barrier region separating the folded and unfolded states (Figure 6.6b, dotted lines). Trajectories across the barrier for unfolding (Figure 6.9a) and refolding (Figure 6.9b) revealed that transitions occurred over a wide range of times, from less than 10 μ s to over 100 μ s. Moreover, many diverse shapes were observed for the transition paths: some crossed the barrier at relatively constant speed, whether slow or fast, showing roughly uniform extension distributions across the transition (Figure 6.9, blue), but more commonly the speed varied greatly along the paths, with noticeable pauses often occurring at one or

more points in the transition (Figure 6.9, red). These measurements thus allow direct visualisation of a host of transient, high-energy intermediates in the barrier region.

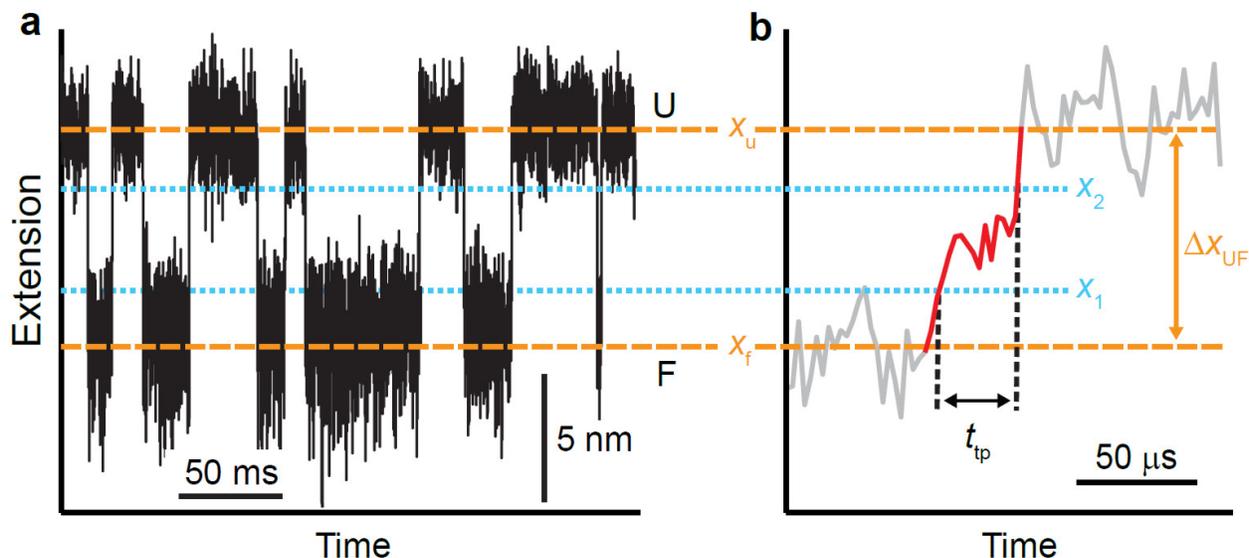


Figure 6.6 Transition paths in force spectroscopy measurements.

(a) End-to-end extension of a hairpin fluctuating in equilibrium between folded (F) and unfolded (U) states under conditions of constant trap separation. (b) Transition paths were identified as the parts of the trajectories (red) moving between U and F states (dashed lines). The transit time, t_{tp} , was defined as the time required to cross between the boundaries x_1 and x_2 (cyan).

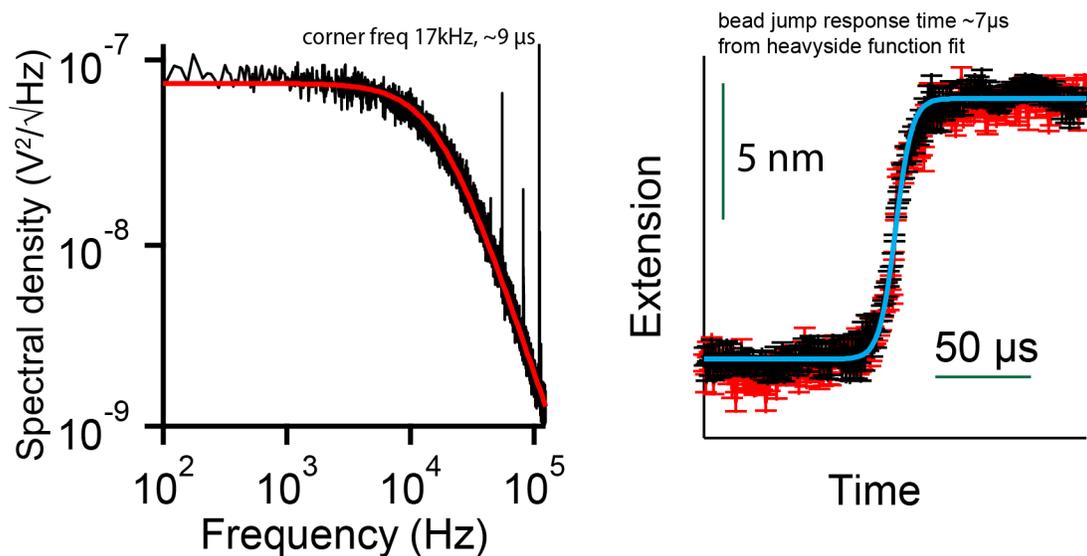


Figure 6.7 Response time measurement of tethered beads.

Left: The power spectrum of a trapped 820-nm diameter bead. The 3-dB corner frequency (17 kHz) translates to a response-time resolution of $\sim 9 \mu\text{s}$. (The result was similar using 600-nm beads). Right: the bead response time measured by jumping the trap position is $\sim 7 \mu\text{s}$. The sampling rate for both measurements was 256 kHz.

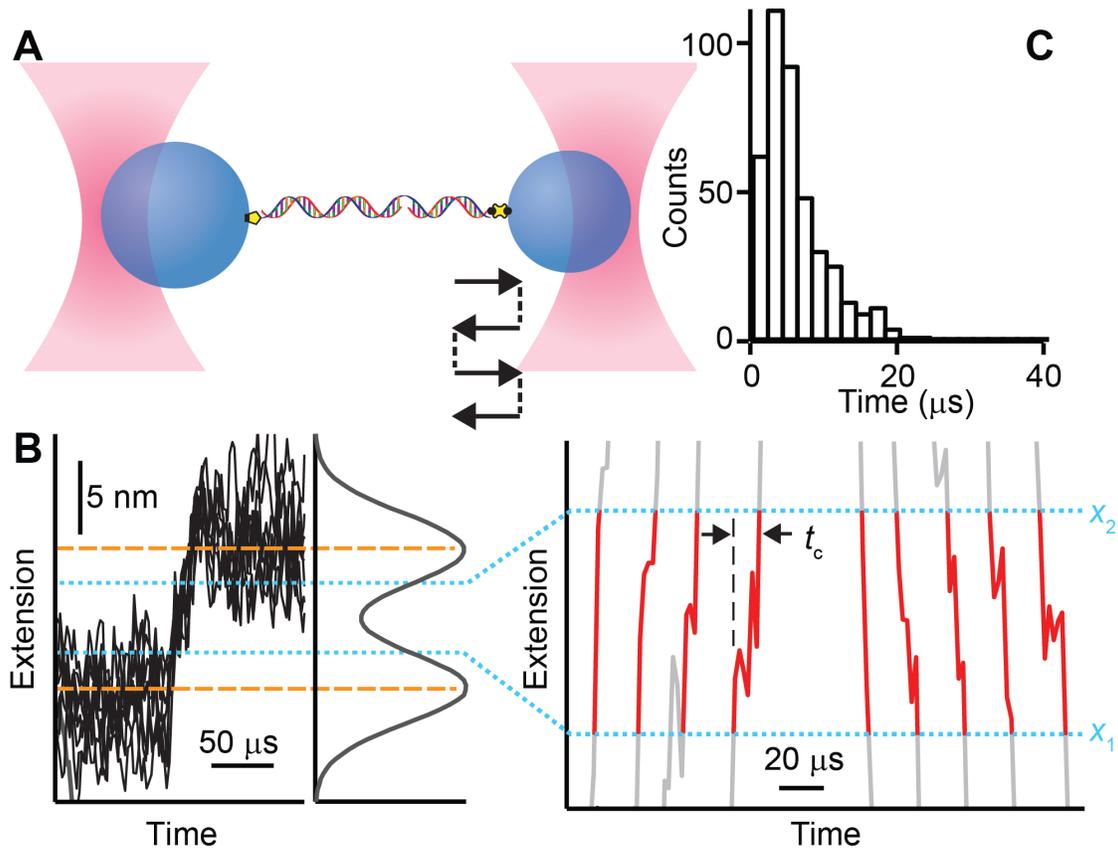


Figure 6.8 Instrumental response time

(A) The response time of the optical tweezers to changes in extension, t_c , was measured using a reference construct consisting of DNA handles only. The construct was held at about 14 pN and one trap was jumped abruptly back and forth to cause the extension of the molecule to change by an amount equivalent to the extension change in the folding of the hairpin 30R50/T4. (B) Extension trajectories of the reference construct (black) were analysed in the same way as the folding/unfolding transitions, measuring t_c from individual transitions (red) as the time required to move between the boundaries x_1 and x_2 (dotted lines). Back-and-forth motion in the trajectories reflects the diffusive motion of the bead. (C) The distribution of response times was peaked near $4 \mu\text{s}$, decaying to 0 by approximately $20 \mu\text{s}$, and had an average of $6 \pm 1 \mu\text{s}$.

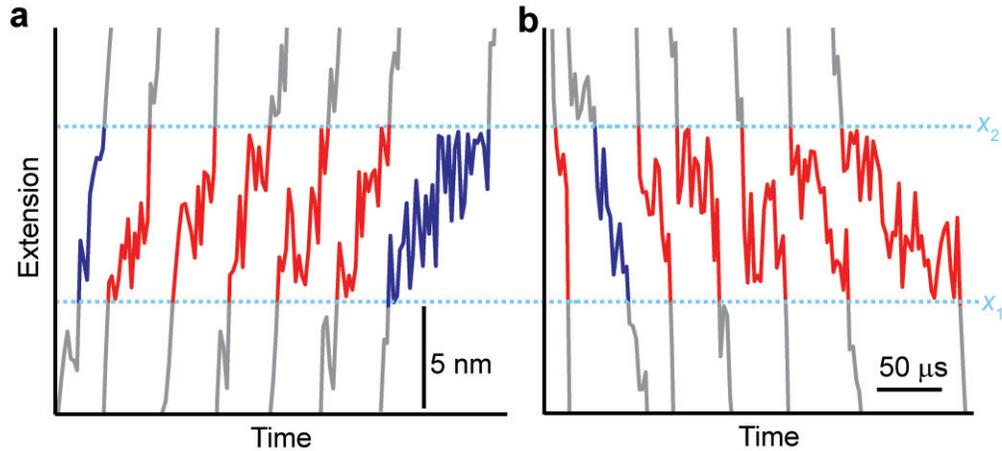


Figure 6.9 Transition paths for a DNA hairpin.

Selection of transition paths for (a) unfolding and (b) refolding. Boundaries x_1 and x_2 (blue) demark barrier region. Transition paths display a wide variety of shapes and transit times.

6.5 Testing diffusive theories of folding

To test at a microscopic level the basic physical picture of folding as a diffusive search over the energy landscape, we focused in on the duration of the transition paths. The transit time for barrier crossing in each transition, t_p , was measured directly from the extension trajectory simply as the time required crossing from one boundary to the other (Figure 6.6d). For consistency, the boundaries were chosen to define the barrier region as the middle half of the total extension change between the folded and unfolded states, Δx_{UF} (Figure 6.6d, dashed lines). Measuring transit times individually for 24 591 unfolding transitions and 24 600 refolding transitions, the average value for the transit time, τ_{tp} was found to be $27 \pm 2 \mu\text{s}$ for unfolding and $28 \pm 2 \mu\text{s}$ for refolding. These average times were slower than the upper bound of $4 \mu\text{s}$ for τ_{tp} estimated for a much shorter DNA hairpin from photon statistics (Truex et al., 2015), but similar to the value for an engineered protein (Chung and Eaton, 2013).

These results now allowed us to test theories of the transit time quantitatively. Above (in section 6.4), the measured landscape profile (Woodside et al., 2006a) and rates (Woodside et al., 2006b) for hairpin 30R50/T4 were used to calculate D from Kramers' equation for diffusive barrier crossing (Hänggi et al., 1990), and τ_{tp} . These results agree very well with those from the direct measurements, validating equation 3.6. Having validated equation 3.6, we then used it to refine the above estimate of D , since τ_{tp} is

in principle a more robust measure than approaches like estimating D from rates using Kramers' theory (Chung and Eaton, 2013; Woodside et al., 2014). Using the barrier parameters from the reconstructed landscape for this hairpin (Woodside et al., 2006a) ($\Delta G^\ddagger = 9.1 \pm 0.1 k_B T$, $\kappa_b = 0.29 \pm 0.02 k_B T/\text{nm}^2$), we found $D = 4.4 \pm 0.4 \times 10^5 \text{ nm}^2/\text{s}$, very close to the indirect value ($4.6 \pm 0.5 \times 10^5 \text{ nm}^2/\text{s}$).

We also tested a proposed relationship between the rates for folding/unfolding (respectively, k_F and k_U) and τ_{tp} :

$$\tau_{tp} = \frac{p(TP)}{2k_U P_F} = \frac{p(TP)}{2k_F P_U}, \quad (6.2)$$

where P_F and P_U are the equilibrium probabilities to be in the folded or unfolded states (respectively) and $p(TP)$ is the fraction of time spent on transition paths (Chaudhury and Makarov, 2010; Hummer, 2004). From the extension trajectories, we found $p(TP) = 6.7 \times 10^{-4}$. The properties of the hairpin in the unfolded state ($k_F = 25 \pm 1 \text{ s}^{-1}$, $P_U = 0.54 \pm 0.03$) then predicted $\tau_{tp} = 25 \pm 2 \mu\text{s}$ for folding from equation 6.2, whereas the properties in the folded state ($k_U = 30 \pm 1 \text{ s}^{-1}$, $P_F = 0.46 \pm 0.03$) predicted $\tau_{tp} = 25 \pm 2 \mu\text{s}$ for unfolding. These results agreed very well with the directly measured values, validating equation 6.2.

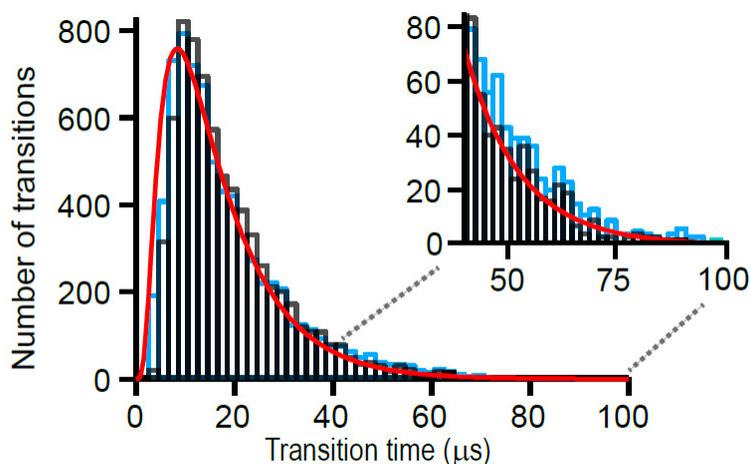


Figure 6.10 Distribution of transit times for DNA hairpin.

The distribution of transit times for barrier crossing is the same for both folding (green) and unfolding (black) transitions. The full distributions are well fit by equation 6.3 (red: unfolding; cyan: folding) and the tails (inset) are separately well fit by equation 6.4 (brown: unfolding; blue: folding), with both fits returning the same results within error.

Significantly, in addition to determining the average value τ_{tp} , the distribution of transit times, $P_{TP}(t)$, could also be measured for the first time, because transit times were found for individual

transitions. As expected from the time-reversal symmetry of the problem (Chaudhury and Makarov, 2010), the distributions had the same shape for both unfolding (Figure 6.9, black) and refolding (Figure 6.9, green) transitions. The transit times were broadly distributed, with a peak around 10 μs and a long exponential tail (Figure 6.9, inset). This behaviour was similar to that predicted for transit over a harmonic barrier in the Kramers regime: $P_{\text{TP}}(t)$ should have the form (see appendix in: (Chaudhury and Makarov, 2010))

$$P_{\text{TP}}(t) \approx \frac{\omega_{\text{K}} \sqrt{\beta \Delta G^{\ddagger}}}{1 - \text{erf} \sqrt{\beta \Delta G^{\ddagger}}} \frac{\exp\left[-\beta \Delta G^{\ddagger} \coth(\omega_{\text{K}} t/2)\right]}{\sinh(\omega_{\text{K}} t/2) \sqrt{2\pi \sinh(\omega_{\text{K}} t)}}, \quad (6.3)$$

where $\omega_{\text{K}} = \beta D \kappa_b$ sets the decay timescale for the exponential tail. Fitting the two distributions (Figure 6.6, dashed lines), we found $\omega_{\text{K}} = 6 \pm 3 \times 10^4 \text{ s}^{-1}$ for both folding and unfolding, implying $D = 2 \pm 1 \times 10^5 \text{ nm}^2/\text{s}$ (using the previously measured barrier stiffness), close to the result found from τ_{tp} via equation 3.6 as well as that estimated from rates via Kramers' theory. The barrier height returned by the fit, $\Delta G^{\ddagger} \approx 0.4 k_{\text{B}} T$, was however too low, reflecting the fact that there were more fast transitions than would be expected from the theory for harmonic barriers. Following the fitting of the distributions to equation 6.3, we have an approximation for the exponential tail (see Appendix E for derivation):

$$P_{\text{TP}}(t) \approx 2\omega_{\text{K}} \beta \Delta G^{\ddagger} \exp(-\omega_{\text{K}} t) \quad (6.4)$$

The distributions of transit times thus agreed quite well, for both folding and unfolding, with the expectations from 1D harmonic approximations to the previously measured landscapes for these two molecules. The primary discrepancy is that the transit time distributions seem to be biased somewhat to shorter times. This bias might arise from a breakdown in the approximations used in the theory (Chaudhury and Makarov, 2010), such as anharmonicity in the barriers or the need to include higher dimensionality in the landscape, or it could reflect the influence of the dynamics of the beads and handles to which the molecules are tethered (Hinczewski et al., 2010b; Makarov, 2014), which are ignored in the analysis (the effects of the tethering on the measured transition time are non-trivial and remain unexplored theoretically (Nam and Makarov, 2015)). The precise origin of the bias to short times is thus still unclear. The overall agreement between different measurements of the diffusion coefficient,

however, provides additional evidence supporting the notion that 1D descriptions of folding are quite good (Gupta et al., 2011; Manuel et al., 2015; Neupane et al., 2015; Truex et al., 2015; Yu et al., 2012a), despite the vast simplification that they embody.

The ability to observe and characterise transition paths directly opens up many exciting avenues to explore in folding studies by allowing more direct investigation of transition states. Previously invisible microstates along the transition paths may now be detectable, permitting their properties to be characterised directly. It may, moreover, be possible to distinguish different classes of transitions paths having different properties such as barrier heights, intermediates, or roughness. There is significant potential for deeper integration of experiment and simulation through direct comparisons of the transition path properties found experimentally to the results of atomistic simulations (Chung et al., 2015). Because the transition time is so sensitive to the diffusion coefficient D (Chung and Eaton, 2013; Woodside et al., 2014), such measurements also hold great promise for investigating the effects of solvent viscosity and internal friction (Brierley et al., 2007a; Chung and Eaton, 2013; Hagen, 2010). By determining the time required for structural transitions to take place in molecules of different size and topology, these measurements open a new window on the microscopic events occurring during biomolecular folding.

Chapter 7: Diffusion Coefficient analysis from single molecule trajectories

7.1 Introduction

The folding of biomolecules such as proteins and nucleic acids can be well described by energy landscape theory. As a result of this, the structural dynamics can be described as diffusive motion over the hypersurface of this landscape (Bryngelson and Wolynes, 1987; Dill and MacCallum, 2012). This notion of diffusion is rather straightforward to describe: it is not taking a direct path. The diffusive search of a folding molecule from state to state is analogous to a molecule's diffusive movement through solution. We may like to picture this diffusion as a random walk in many dimensions, much like our inebriated lab colleagues on a late night after celebrating the movement of bubbles in beer; they waltz out of the bar onto the street in search of a lamppost. For a movement where each step along its path is uncorrelated to the previous ones (said to be Markovian, (alternatively see: (Makarov, 2013; Plotkin and Wolynes, 1998)), the diffusion coefficient sets the speed (or timescale) of this motion. A simpler, alternative way to state it is that the time required to travel a distance depends on the square of that distance (approximately). Further, in opposition to classic transition-state theory, the motion can cross back and forth over the barrier more than once, since the transition over the barrier is not a point of no return (Hänggi et al., 1990). We are greatly interested in the diffusivity: it provides the connection between the kinetics and the energy landscape. The diffusion coefficient may reflect the underlying viscosity or 'roughness' of the energy landscape. It helps determine the kinetic properties of our biological molecule under study, such as the reconfiguration time for the polymer chain (Gopich et al., 2009), the rates for contact formation and folding (Kramers, 1940), and of course, the transition paths across the landscape's energy barriers (Chaudhury and Makarov, 2010; Chung et al., 2009)(and previous chapter).

When the molecule's motion in conformational space is projected down onto a one-dimensional reaction coordinate, this changes the observed dynamics, hence the diffusion coefficient D generally becomes position dependent, and the details depend on the nature of the projection. Moreover, the roughness of the energy landscape depends on the axis (choice of reaction coordinates), so depending on

the circumstances, D may or may not end up being constant. Measuring this position dependence of D is a point of great difficulty, as it is unclear how reliable previous results relating D 's change over position have been. Moreover, it has been challenging to measure D over the folding/unfolding transition barrier(s) separating state(s) (Woodside et al., 2014). Previously we determined transition times with the assumption that D can be approximated as a constant, but at higher resolution it we should not expect it to be a constant, e.g. FRET studies showing a decrease in D as it approaches the transition state (Borgia et al., 2012). The folding kinetics of the villin subdomain has been measured with nanosecond laser temperature jump, and the fitting of its data with an Ising-like model produced only a relatively small position dependence for the diffusion coefficient, compared to the predictions of theory (Cellmer et al., 2008). It has been pointed out that single-molecule trajectories might be the best way to explore the position dependence (Best and Hummer, 2011), however few experiments have been carried out so far.

With smFS we can determine D over the barrier, crucially, and over a wider range of the extension (away from said barrier) of the molecule by reconstructing the energy landscape profile along the extension reaction coordinate (Lannon et al., 2013; Woodside and Block, 2014; Yu et al., 2012a) (previous chapter). Recent AFM studies including D in their analysis have shown little position dependence (Berkovich et al., 2012; Lannon et al., 2013). These results are not without controversy, as they do not take into account effects that the instrumentation itself could have made on the measurements. Recent work has suggested that the tethering to an AFM cantilever, or beads and handles in the case of OTs, contributes to the microscopic dynamics being measured (Berkovich et al., 2012) because of its limited response time and other properties (Makarov, 2014; Nam and Makarov, 2015; Woodside et al., 2014). There has been some debate about the effects of the chosen reaction coordinate, but extension has recently been confirmed to be a good reaction coordinate, at least in the limiting cases of DNA hairpins being studied (Neupane et al., 2015). As such, we have explored an alternative approach put forward by R. Netz's group (Hinczewski et al., 2010a), based on the average time it takes for a molecule to return to its starting position on the reaction coordinate. (Specifically, it makes use of a one-dimensional Fokker-Planck approach and assumes an underlying Markovian process.) We also investigated a second approach from J. Brujic's group, based on the average fall time in collapsing

trajectories to the molecules folded states, to determine the diffusion coefficient as a function of position (Lannon et al., 2013).

7.2 Round-trip time analysis of diffusivity in DNA hairpins

The round-trip time approach to calculating the diffusivity $D(x)$ assumes that the reaction coordinate undergoes stochastic time evolution in 1D as described by the Fokker-Planck equation of the probability configuration. In our analysis, this can be written as:

$$\frac{\partial \Psi(x, t)}{\partial t} = \frac{\partial}{\partial x} D(x) e^{-\beta G(x)} \frac{\partial}{\partial x} \Psi(x, t) e^{+\beta G(x)} \quad (7.1)$$

Here $\Psi(x, t)$ is the probability of a configuration with respect to extension (x) at time t , so that $\langle \Psi(x) \rangle = \rho(x)$ is the time-averaged probability distribution, while $\beta G(x) = -\ln \langle \Psi(x) \rangle$ is the free energy profile.

Defining the mean first passage time to go from an extension value x to a final state x^f without a recrossing as $\tau_{FP}(x, x^f)$, the round trip time, defined as the time needed, starting from x , to travel to x^f

and back again to x (see Figure 7.1), is given by: $\tau_{RT}(x, x^f) = \text{sign}(x - x^f) \cdot \{ \tau_{FP}(x, x^f) - \tau_{FP}(x^f, x) \}$,

such that $\tau_{RT}(x, x^f) = Z \int_x^{x^f} \frac{e^{\beta G(x')}}{D(x')} dx'$, where Z is the partition function. Solving for diffusivity, one

obtains: $D(x) = Z e^{\beta G(x)} \frac{\partial \tau_{RT}(x, x^f)}{\partial x}$ (Hinczewski et al., 2010a).

This approach is well-suited for application to single-molecule folding trajectories measured in equilibrium, since the numerator is nothing more than the inverse of the extension probability distribution, $\rho(x)$, whereas τ_{RT} can be calculated empirically straight from the trajectory:

$$D(x) = \left(\rho(x) \frac{\partial}{\partial x} \langle \tau_{RT}(x, x^f) \rangle_{\text{average}} \right)^{-1} \quad (7.2)$$

We can note that $D(x) \propto^{-1} \frac{\partial}{\partial x} \langle \tau_{RT}(x, x^f) \rangle$. In principle, this approach should yield the same answer

regardless of the choice of the reference point x^f , assuming that the folding is well-described by 1D diffusion along the reaction coordinate. As stated in the previous section, end-to-end extension has been

proven to be a good reaction coordinate for DNA hairpins (Manuel et al., 2015), and indeed the hairpin folding statistics has been shown to agree well with the expectations for pure 1D diffusion along the measured landscape, making DNA hairpin folding the ideal system for testing this approach for measuring the position-dependence of D .

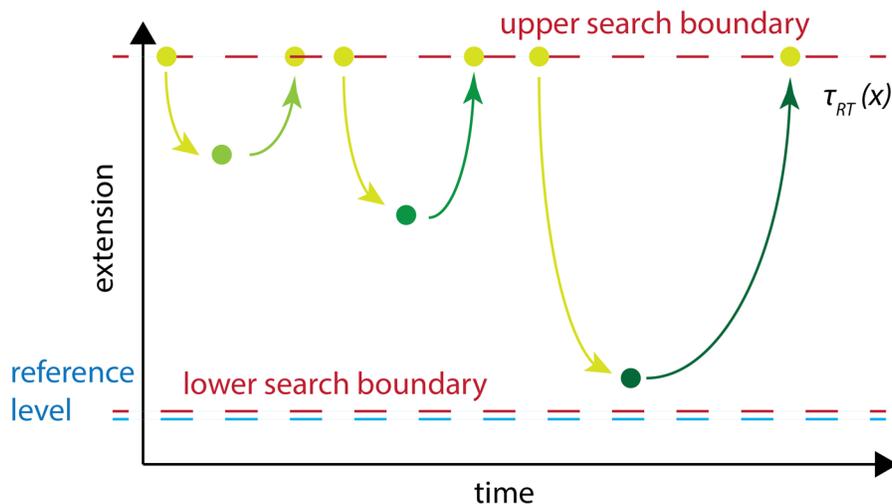


Figure 7.1 Round trip time as defined in an extension trace

We applied this approach to constant-force measurements of two DNA hairpins (HPs): 30R50/T4 and 20TS06/T4 (Figure 7.2). Data from hairpin 30R50/T4 are shown in Figure 7.3. We calculated τ_{RT} in two ways: once with respect to a reference point located at the peak of $\rho(x)$ corresponding to the folded state, x_f (forward direction), and then again with respect to a reference point located at the peak of $\rho(x)$ corresponding to the unfolded state, x_b (reverse direction) (As Hinczewski *et al.* discuss, the choice of placement of our adsorbing boundary and the precise placement of our starting point is somewhat flexible since the round trip time, τ_{RT} will scale by an additive constant if our placements vary a little, and will not change D .) The results are shown in Figure 7.4. To reduce noise, we smoothed $\tau_{RT}(x)$ sparingly before taking its derivative (Figure 7.5). The final result for the diffusivity for hairpins 30R50/T4 and 20TS06/T4 is shown in Figure 7.6 and Figure 7.7. The trap stiffnesses were 0.52 and 0 pN/nm for the experimental data from constant force. When comparing the diffusion from constant force data collected from different molecules of 30R50/T4, the results are qualitatively similar, varying over the same two orders of magnitude.

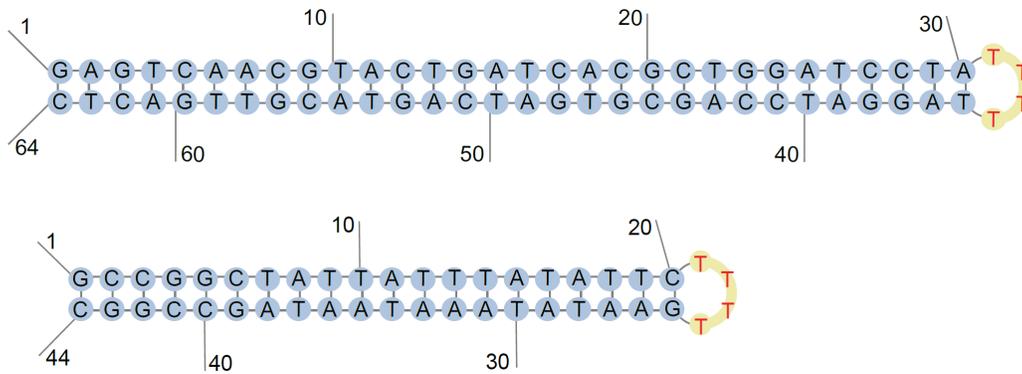


Figure 7.2 Schematics of DNA hairpins 30R50/T4 (top) and 20TS06/T4 (bottom)

Figure generated with Pseudoviewer3 (Byun and Han, 2009; Ponty and Leclerc, 2015)

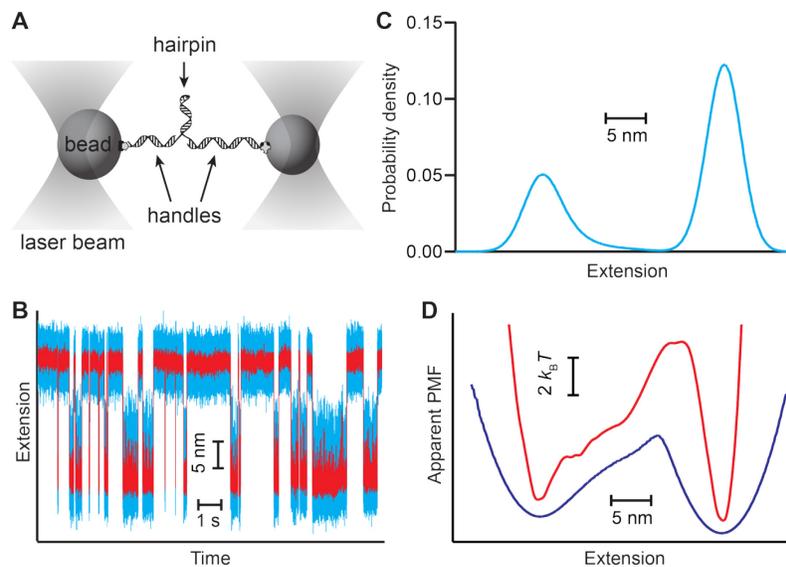


Figure 7.3 Constant force of HP 30R50/T4

A: schematic of measurement setup, **B:** representative trajectory of constant force data trace from HP 30R50/T4, with filtering to better show separation of states and **C:** the normalised histogram for the distribution. **(D)** Apparent PMF found from an inverse Boltzmann transform of the extension distribution (**blue**), and the landscape after deconvolution to remove the effects of the compliant handles and probes (**red**) (Adapted from figure 1 from(Woodside et al., 2014).) See Table 6.2 on page 87 for previously measured values for this hairpin.

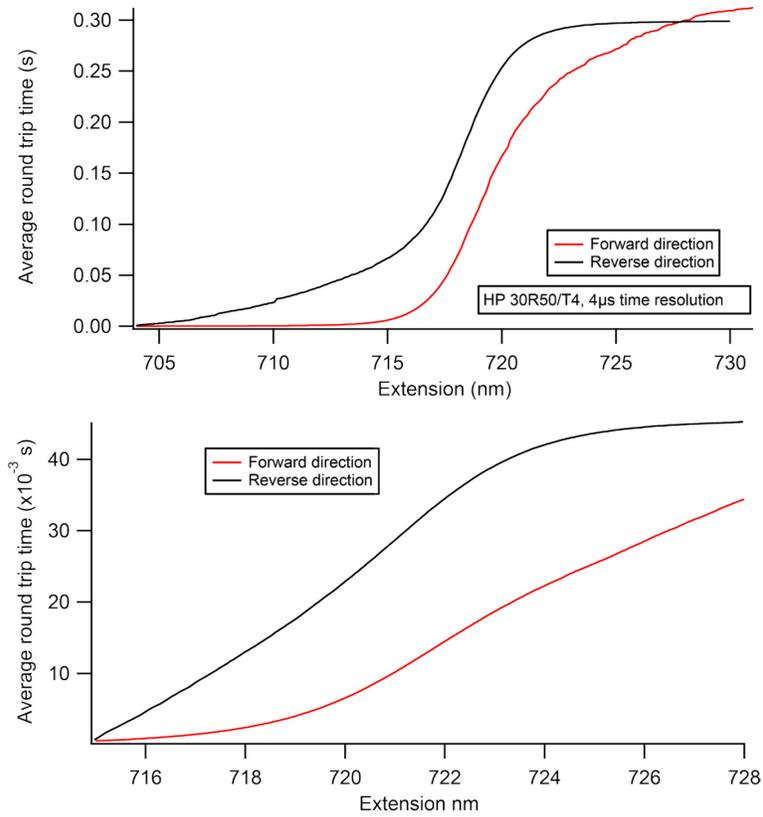


Figure 7.4 Average round trip times example.

Top: τ_{RT} with reference point at the folded state (F) in red (**forward**), the vertically offset τ_{RT} referenced from the unfolded state (U) in black (**Backward**). HP 30R50/T4 data data sampled at 250 kHz. **Bottom** equivalent data from HP 20TS06/T4 data sampled at 250 kHz. All data are from constant-force measurements at each hairpin's respective $F_{1/2}$.

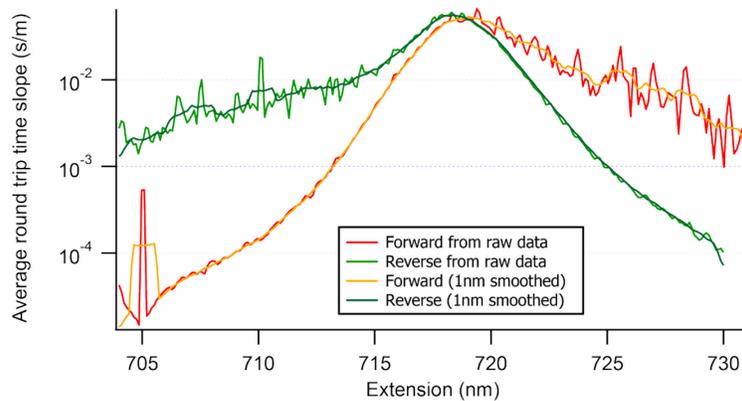


Figure 7.5 The derivatives of the average τ_{RT} .

With box (mean) smoothing of τ_{RT} at 1 nm width, overlaid. **Red:** unsmoothed folded to unfolded path (referenced at F), **green:** unsmoothed unfolded to folded path (referenced at U). HP 30R50/T4 data sampled at 250 kHz. Data are from constant-force extension records at $F_{1/2}$.

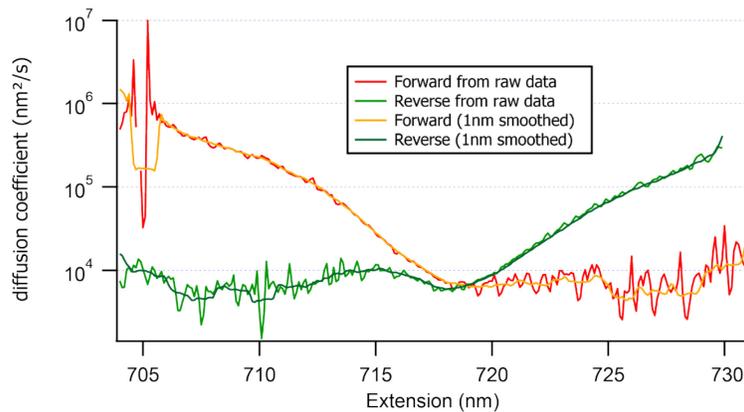


Figure 7.6 Round trip time diffusion coefficient for HP 30R50/T4 .

with box (mean) smoothing of τ_{RT} at 1 nm width, overlaid. **Red:** unsmoothed folded to unfolded path (referenced at F), **green:** unsmoothed unfolded to folded path (referenced at U). HP 30R50/T4 data sampled at 250 kHz. The trap stiffnesses were 0.52 and 0 pN/nm for the experimental data from constant force.

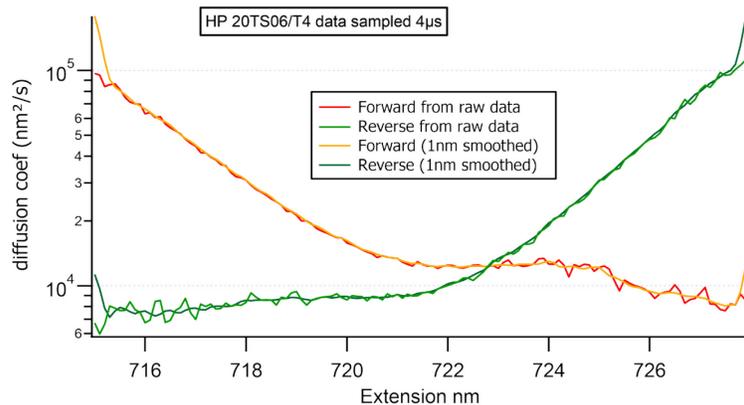


Figure 7.7 Round trip time diffusion coefficient for HP 20TS06/T4

with box smoothing of τ_{RT} at 10 points, or 1 nm, overlaid. **Red:** unsmoothed folded to unfolded path (referenced at F), **green:** unsmoothed unfolded to folded path (referenced at U). HP 20TS06/T4 data sampled 250 kHz. Data are from constant-force extension records at $F_{1/2}$.

It is immediately evident that this calculation suggests that D is not constant, varying over as much as two orders of magnitude, depending on the hairpin. However, the analysis is not self-consistent,

since the calculations reference at the folded (F) and unfolded states (U) do not agree, and in fact they diverge qualitatively. Our data are not directly comparable to the simulations on a 15 residue alanine-based peptide of Hinczewski *et al.*. According to the theory, differences in $D(x)$ calculated for different reference starting points should not occur, since the folding is known to be well-described as ideal 1D diffusion with Markovian behaviour (Manuel et al., 2015; Neupane et al., 2015). Furthermore, the values of D obtained near the barriers for the two hairpins ($4.6 \pm 0.5 \times 10^5$ and $5 \pm 3 \times 10^5$, respectively for 30R50/T4 and 20TS06/T4) (see Table 6.2, p. 87), are considerably larger than the results found previously from both Kramers theory (HP 30R50/T4 : $9 \pm 1 \times 10^3$, HP 20TS06/T4: $6.3 \pm 0.5 \times 10^3$) (Woodside et al., 2014) and from the more detailed analysis of τ_p ($D = 4.4 \pm 0.4 \times 10^5$ nm²/s for HP 30R50/T4) (Chapter 6, Neupane *et al.* 2015 *under review*). One possible explanation here may be the influence of the handles and/or beads, which have been shown to have important effects on the observed dynamics, in some cases changing them considerably because the molecule has to now drag the dsDNA and bead with it as it folds/unfolds, increasing the time required (Makarov, 2014; Nam and Makarov, 2015).

7.3 Testing the round-trip method with simulated hairpin data

In an effort to further understand the results, we applied the round-trip time calculated to simulated folding on a 1D landscape, where a known diffusion coefficient was imposed. Simulated data generated by John Lambert as described in: (Woodside et al., 2014) were used. Briefly: the experiment was simulated with a toy model consisting of a bead of specified sizes (20, 100, 400, 2000nm radii; results here are only with 400nm beads) that was subjected to a constant force (with a linear spring on the bead to mimic the optical trap) and attached to a handle of given compliance, which was in turn attached to a “molecule” having a 1D energy landscape similar to hairpin 30R50/T4. The stochastic forces on the molecule and bead are drawn from a Gaussian distribution and then the fluctuating bead position was treated with Langevin dynamics. As for the fluctuations in the molecule position, these were treated with pure Brownian dynamics (because of the negligible inertia term). Simulated trajectories, using a velocity Verlet algorithm (Grønbech-Jensen and Farago, 2013), were down-sampled from the calculation step size (10ps timestep) to a rate similar to the experimental sampling rate (40 and 5 μ s, 320s traces, 3-4weeks of

simulation time per trace). To reduce the computational time, the barrier is lower and the extension is shortened compared to the DNA hairpins.

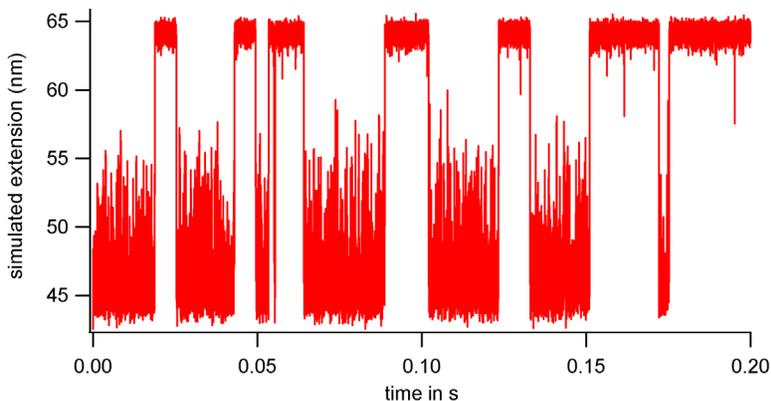


Figure 7.8 Sample simulated lone HP extension trace with constant D ($5 \mu\text{s}$ time res.)

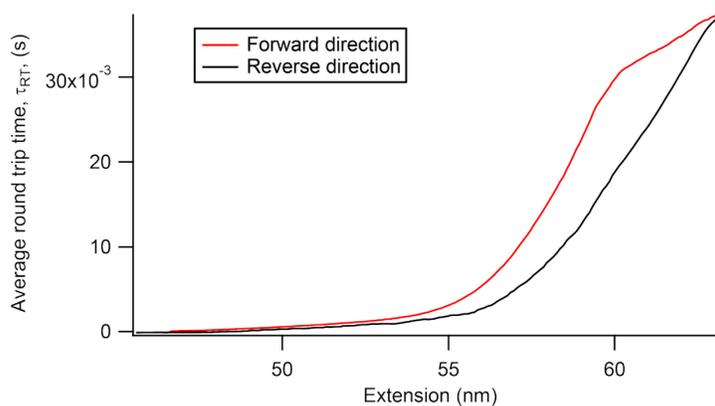


Figure 7.9 Average round trip time from simulated HP with constant D at $3 \times 10^5 \text{ nm}^2/\text{s}$

The red trace is with the starting point reference at the folded peak (F) and the black is with the reference at the unfolded peak (U).

To test if the problems indeed arose from the handles and/or beads, we first looked at simulations in which the handle and bead were absent, and force was applied directly to the end of the molecule (Figure 7.8). A constant diffusion coefficient of $3 \times 10^5 \text{ nm}^2/\text{s}$ was imposed in the simulation. τ_{RT} and $D(x)$ calculated from this trajectory are shown in Figure 7.9 and Figure 7.10 (upper panel). Significantly, the calculations with reference points at the folded state (F) and unfolded state (U) yield

effectively the same result, in contrast to the case with the experimental data. Moreover, the imposed constant value of $3 \times 10^5 \text{ nm}^2/\text{s}$ is recovered reasonably well over most of the range of the reaction coordinate, except at the edges, within a few nm of the folded and unfolded states. We can note that the diffusion is roughly constant, near the imposed value, with a notable ‘hump’ at the unfolded end of the extension. The sharp upturn at the short extension end is a smoothing artifact, but no obvious effect from the barrier between the unfolded and folded states. We also imposed a non-constant diffusivity in the simulation, rising from a basal value of $0.3 \cdot 10^5 \text{ nm}^2/\text{s}$ into a Gaussian peak of $3 \times 10^5 \text{ nm}^2/\text{s}$ at 60.15 nm. Again, $D(x)$ calculated from the simulated trajectory recovers the imposed D quite well (Figure 7.10-lower panel, imposed D indicated in light blue). However, for the reverse trace with the reference point at the unfolded state, substantially more smoothing (15nm) of the τ_{RT} is required for a visible line to be seen on the plot.

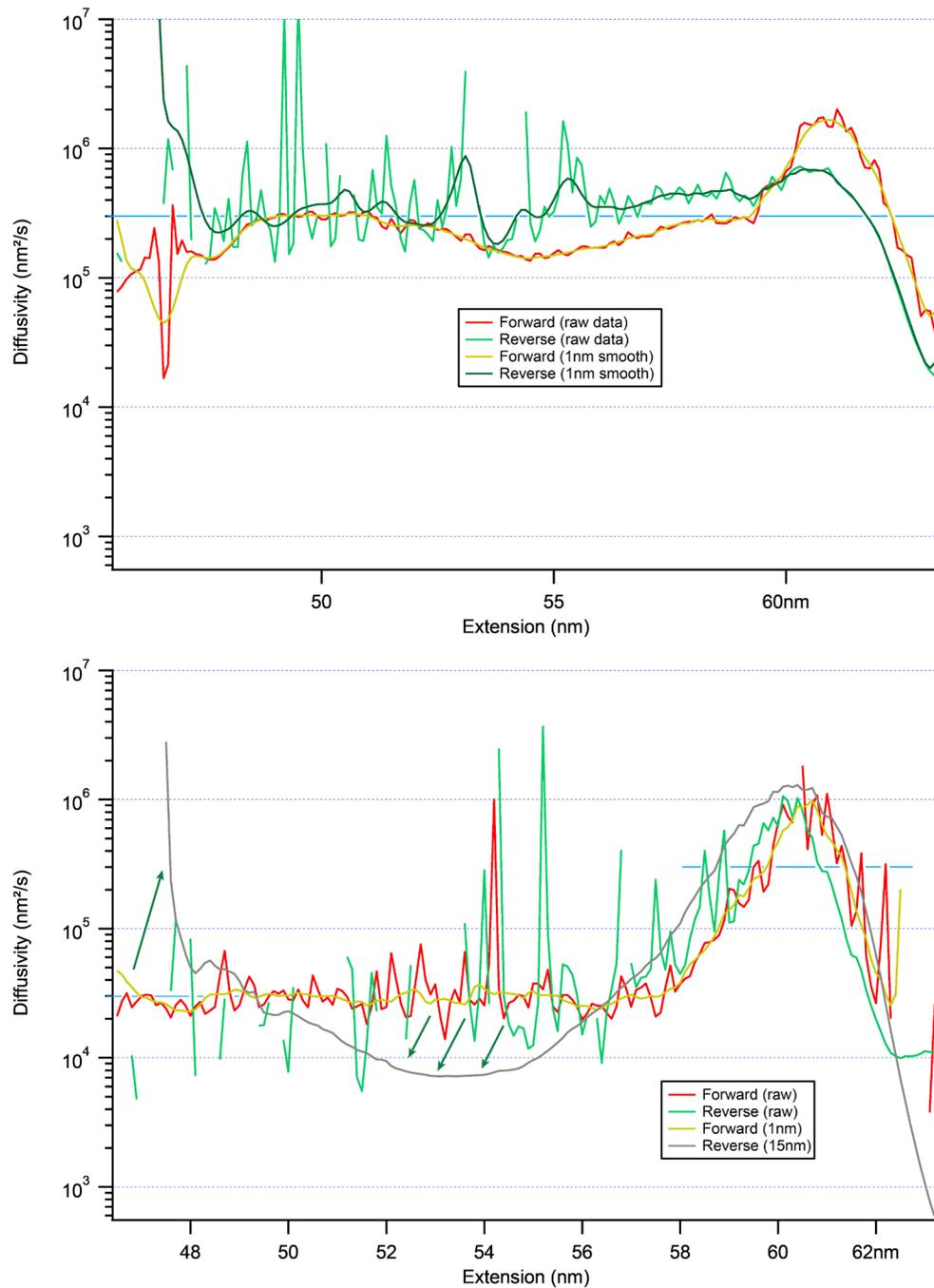


Figure 7.10 Diffusion coefficients for simulations without beads and handles.

Imposed constant D at 3×10^5 nm²/s (in light blue)(upper panel), and Gaussian variation from 0.3 to 3×10^5 nm²/s, peaking at ~ 60 nm (lower panel). Smoothing over 1 nm with the exception of the Gaussian reverse direction (reference at unfolded state (U), which is very noisy: overlay of D calculated from 15 nm smoothed τ_{RT} . (green arrows point to where the smoothed curve is diverging from the unsmoothed data)

Since the simulations without handles and beads support the notion that the inconsistencies observed in the analysis of the experimental data are an effect of the handles, we next added a 400-nm bead and a handle (stiffness varying from 0.2 – 1.0 pN/nm) to the simulation. A constant diffusion coefficient of $3 \times 10^5 \text{ nm}^2/\text{s}$ was imposed in these simulations. Calculating the diffusivity as above, the results are shown in Figure 7.11. Apart from the edge effects, we can note that the high-stiffness result is similar to the diffusivity calculated without a handle and bead. In contrast, the results with low-stiffness handles (0.2 and 0.3 pN/nm) look qualitatively more similar to the result found from the experimental data (Figure 7.6). Of course the correspondence is not exact, nor would we necessarily expect it to be so, since the simulations do not reflect the full complexity of the experiments (which include dynamics within the handles, as well as two beads and two handles each subject to their own fluctuations). Nevertheless, these results support the hypothesis that one must account for the effects of the beads and handles in the round-trip time analysis.

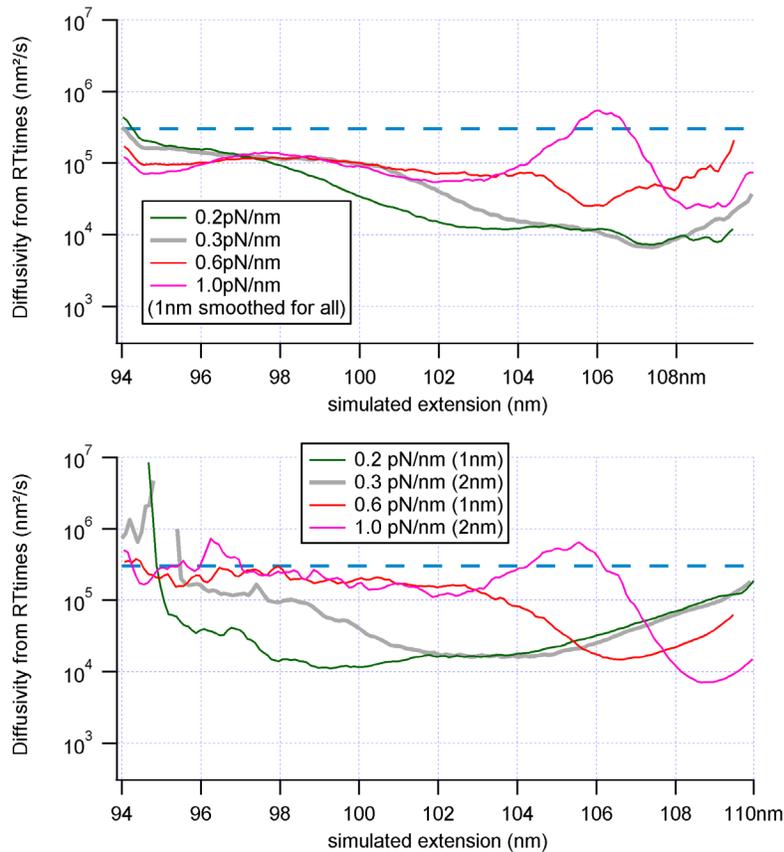


Figure 7.11 Diffusion coefficient from simulated HP landscape with varying stiffness.

Top: Reference starting point at folded state (F), bottom: reference at unfolded state (U). 0.2, 0.3, 0.6, and 1.0 pN/nm, 40 μ s resolution traces. The traces have been horizontally aligned for the purpose of comparison. The underlying imposed diffusion coefficient is a constant 3×10^5 nm²/s in all cases, indicated with the dashed blue line.

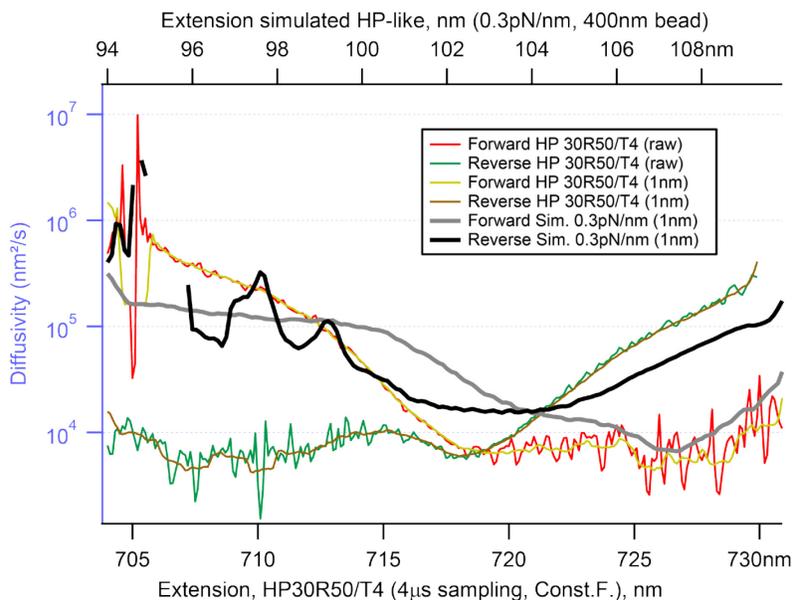


Figure 7.12 Diffusivity from simulation and experiment.

A comparison of the highest sampled constant force data of HP 30R50/T4 (bottom axis, red: (F), green: (U), 1nm box smoothing lines are overlaid), and the simulated HP 30R50/T4 -like landscape with a bead of a 400 nm radius in a 0.3 pN/nm potential (top extension axis, grey: (F), black: (U)).

We note that correcting for the effects of beads and handles on τ_{RT} is not as simple as it is for something like the energy landscape derived from the inverse Boltzmann transform, where a straightforward deconvolution of the compliance effects can be used (Hinczewski et al., 2010b; Woodside et al., 2006b), because a deconvolution of both spatial and temporal effects is needed. A general theory of dynamic deconvolution for mechanical networks has been developed (Hinczewski et al., 2010b), but it has proven challenging to apply to experimental data, yielding inconsistent results (Soong, 2014). Very recently, the question of how bead-linker connections to a molecule change the folding dynamics has been addressed in a more practical way by Makarov and co-workers (Makarov, 2014; Nam and Makarov, 2015). This work explored the effects of linker stiffness on the apparent diffusion coefficient during

motion across a barrier, showing that only when the linker is compliant compared to the stiffness of the barrier (which happens to be the case for the hairpin 30R50/T4) will the intrinsic diffusivity of the molecule be recovered from measurements of rates; otherwise, the apparent molecular diffusivity will be modified by that of the bead. However, the effects of this coupling on the transit time, which is of greater relevance to the calculation of τ_{RT} than is the folding rate, were deemed non-trivial, and thus no theoretical treatment of the question is currently available. As of yet, it remains an open question how best to correct the effects of handle/bead attachments for round-trip time calculations of the diffusivity.

7.4 Diffusivity from average fall times in force-jump measurements

An alternate approach for estimating the position-dependence of D emerged from work by Brujic and colleagues – described in more detail in the next chapter – to reconstruct energy landscapes from force-jump measurements (Lannon et al., 2013). This method was based on the time required for the molecule to reach an extension x , when starting in the unfolded state at high force, after the force was jumped down to a value sufficiently low for folding to occur. The diffusivity was related to this average collapse or fall time, τ_c , via

$$D(x) = \frac{1}{\tau_c(x) \frac{\partial}{\partial x} \rho(x)} \quad (7.3)$$

Here $\rho(x)$ is the non-equilibrium probability distribution of the molecular extension from the force-jump measurements, the out-of-equilibrium analogue of the equilibrium extension probability distribution used in the inverse Boltzmann transform method for landscape reconstructions.

We tested this fall-time method by measuring the refolding of hairpin 30R50/T4 under force-jump conditions, as described in greater detail in the next chapter. Briefly, starting initially with a force a ~ 2 pN above $F_{1/2}$, the force was abruptly jumped down to 2 pN below $F_{1/2}$ (~ 4 pN force drop), so that the hairpin was induced to refold. For each refolding transition, the fall time to a given value of extension, x , was measured directly from the data record, and the average over hundreds of transitions was computed (Figure 7.13).

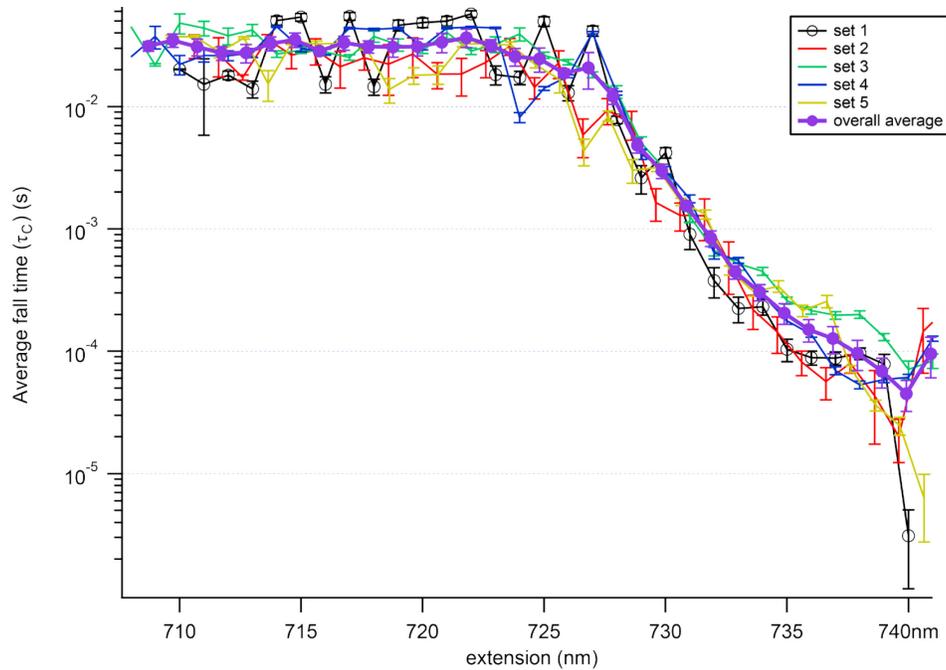


Figure 7.13 Average fall time along collapsing extension traces (τ_C)

With the average across sets (~100 jumps per set) in violet overtop. Traces start from the unfolded length near 740 nm. (Data concatenated from HP 30R50/T4, (200 and 100kHz sampling in sets), 16.6 to 12.3 pN force)

$D(x)$ was then calculated from equation (7.3), with result shown in Figure 7.14. Notably, there are gaps where D becomes negative, because of the direct dependence on the slope of $\rho(x)$. D also appears to vary considerably between the folded and unfolded states, over at least two orders of magnitude, and may often be irregular and noisy between samples. It is therefore difficult to conclude that the fall-time analysis is reliable. As discussed in Chapter 8, we believe these difficulties likely once again relate to the effects of the beads and handles on the measurement, this time affecting the calculation both by altering the distribution $\rho(x)$ and the dynamics of the transition (and hence the fall time). Interestingly, the fall-time analysis yields a diffusivity that is dissimilar to the result of the round-trip time analysis, which suggests that the handle/bead effects are different for the two analyses. Finally, we note that since neither of the methods we have tested can be applied reliably to experimental data, determining the position dependence of the diffusion coefficient remains an ongoing challenge in folding studies.

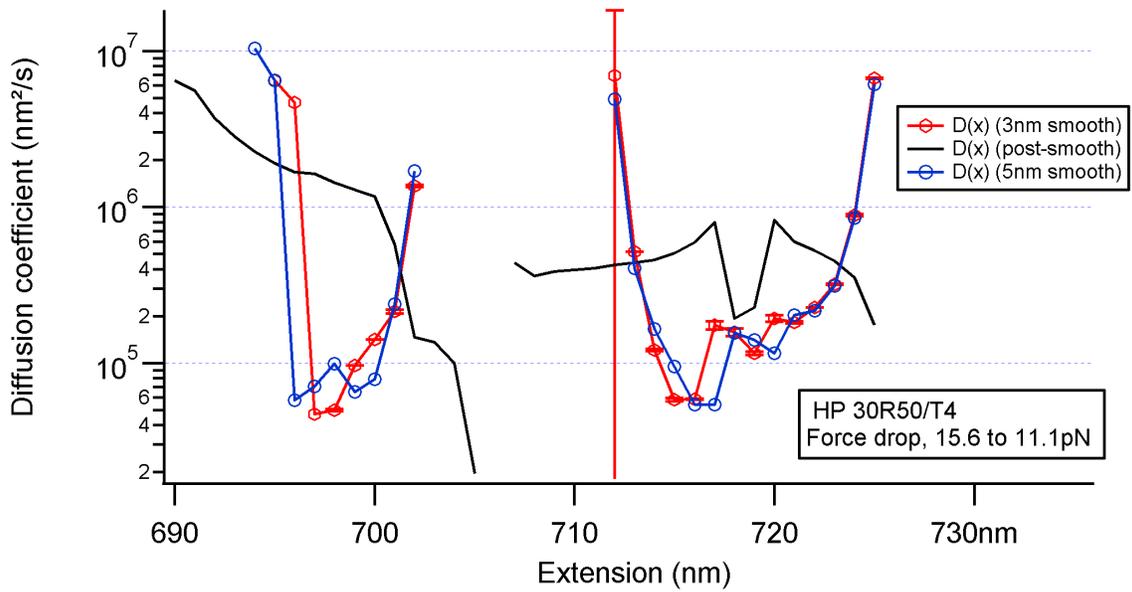


Figure 7.14 Example of diffusion coefficient from force jumps for 4.5pN drop.

D in (nm^2/s) calculated from the average fall time $\tau_c(x)$ for HP 30R50/T4 (100 kHz sampling) Its range is from the unfolded state on the right to the first arrival at the folded state on the left (short extension). Smoothing is applied in different ways: first to the $\tau_c(x)$ with 3 and 5 nm lengths, then a smoothing of 1 nm after the $D(x)$ calculation instead. The discontinuity from ~ 702 to 712 nm is due to a negative slope in that region of the non-equilibrium PDF.

Chapter 8: Energy landscape reconstruction from force-jump measurements

8.1 Motivation

If the energy landscape of the folding of a biomolecule is precisely known, it can be used to predict the different states of the folding that molecule. Methods of landscape reconstruction have been validated for equilibrium measurements (e.g. constant force (Woodside et al., 2006a; Woodside et al., 2006b)) and non-equilibrium measurements force ramp measurements (e.g. from FECs (Gupta et al., 2011)). However, many smFS studies make extensive use of force-jump methods, for which the previously-mentioned landscape reconstruction techniques are not applicable. Force-jump measurements are nevertheless very useful, as they not only allow for the study of folding in cases where it would be impractical to make equilibrium measurements (e.g. owing to slow folding rates), but they also in principle allow for unexpected pathways to be explored owing to the large applied force bias (Barsegov et al., 2006; Hyeon and Thirumalai, 2006; Hyeon and Thirumalai, 2008; Li et al., 2006; Pincus et al., 2008). Recently, a method was proposed for reconstructing landscapes from force-jump measurements (Zhang et al., 2011), and it was demonstrated on measurements of polyubiquitin (Lannon et al., 2013). However, this method has not yet been validated by comparing its results to those of other, more established methods. Here we seek to do so, by applying the method to force-jump measurements of a DNA hairpin whose landscape has been reconstructed previously by multiple methods and is thus well-known.

8.2 Outline of the method

Systems that have been driven far from equilibrium by a discontinuous change in the experimental control variable, such as by jumps in force, produce a distribution of non-equilibrium trajectories that present a difficult inverse problem. This problem has been approached with stochastic differential equations (Guillaume and Tom, 2004) and by the use of path integrals (Stuart, 2010). More recently, a simpler inversion procedure for non-equilibrium landscape reconstruction was proposed by Zhang *et al.* (Zhang et al., 2011) and applied experimentally to AFM data of protein folding by Brujić and colleagues (Lannon et al., 2013). In this method, the end-to-end extension of the molecule is treated as undergoing overdamped Langevin dynamics, and the standard probability density function (PDF) of the

extension used in equilibrium analysis is replaced with a non-equilibrium stationary PDF, so that the free energy can be written as (Zhang et al., 2011):

$$G(x) = -k_B T \left[\ln \rho(x) + D(x_{refolded}) \rho'(x_{refolded}) \int_x^{x_u} \frac{dx'}{D(x') \rho(x')} \right] \quad (8.1)$$

Here $\rho(x)$ is defined as the PDF of the collapsing trajectories from the moment the force is quenched ($x_{unfolded}$) until the moment that the molecule first reaches its folded length at low force ($x_{refolded}$), and $D(x)$ is the diffusivity. Lannon *et al.* put forward an approach to calculate the diffusion coefficient, D , from these same collapsing traces by using the average time (τ_c) to collapse to a point x from the unfolded state, in equation (7.3). As discussed in Chapter 7, in general D should have some dependence on position (Best and Hummer, 2010), but D is often approximated as constant (as indeed was done in the work by Lannon et al.), in which case the dependence of $G(x)$ on D in 8.1 vanishes:

$$G(x) = -k_B T \left[\ln \rho(x) + \rho'(x_{refolded}) \int_x^{x_{unfolded}} \frac{dx'}{\rho(x')} \right] \quad (8.2)$$

(A derivation of Equation 8.2 based on work by Zhang et al. (Zhang et al., 2011) is shown in Appendix F.) To help reduce noise, the probability density function in this calculation was fit using a Gaussian kernel density estimate (KDE) (e.g. (Wied and Weißbach, 2012)), a non-parametric way of estimating the PDF of a random variable, which has the advantage of producing a smooth and continuous function. As was seen in the last chapter, the inversion of a rough function can produce undesirable, erratic discontinuities. In order to produce a KDE of a reasonable kernel size, the Gaussian KDE is compared to the conventional histogram (a rectangular kernel) with narrow bins.

8.3 Application of the force jump techniques

To test this landscape reconstruction method, we used the DNA hairpins 30R50/T4 and 20TS06/T4 (see Figure 7.2 on p.102), whose landscapes have been reconstructed by multiple methods (Engel et al., 2014; Gupta et al., 2011; Manuel et al., 2015; Woodside et al., 2006a; Woodside et al., 2006b). Since large jumps in force can lead to the molecule exploring widely different parts of the energy landscape (Barsegov et al., 2006; Hyeon and Thirumalai, 2006; Hyeon and Thirumalai, 2008; Li et al., 2006; Pincus et al., 2008), and we aimed for quantitative comparison to previous equilibrium or near-

equilibrium measurements, we initially restricted the jump in force to be relatively small: a range of 4 pN, from 2 pN above $F_{1/2}$ for the hairpin to 2pN below (in terms of the equilibrium behaviour, from 99% unfolded to 99% folded). For comparison, the AFM work of Lannon *et al.* involved jumps of over 100 pN. Our experiments have a fundamental limit of ~ 60 pN at which point the DNA or RNA/DNA duplex begins to melt (shear) (Gross et al., 2011) and the dig-antidig linkage may break near that same force (Neuert et al., 2006). (The other linkage, biotin-avidin bond, withstands up to ~ 80 pN before dissociation becomes an issue, (Pincet and Husson, 2005)). In our force-jump measurements, the force is kept constant after the jump using the passive force clamp, where one of the beads sits in the zero-stiffness region of one trap. The force is jumped by abruptly changing the laser intensity in the trap being operated in the zero-stiffness region (see appendix A for more detail).

As explained in section 7.4, the force-jump data consist of extension trajectories that involve repeated one-way transitions (Figure 8.1, top) (only refolding, for jumps to lower forces, or unfolding, for jumps to higher forces). The analysis began by excising the individual transitions from the full extension traces (which involve many transitions concatenated together, see appendix B). Next, sets of transitions measured under the same conditions were pooled and aligned (to compensate for instrumental drift) (Figure 8.1, bottom). The distribution of extension values in the pooled data was then calculated and the KDE generated (using a fast Gaussian transform in Igor Pro), adjusting the kernel width so that the KDE agreed reasonably with the box transform) (Figure 8.2). The resulting KDE was used as our best estimate of the extension PDF. Note that the extension values for the folded and unfolded states were identified from the original extension record *before* the excision of transitions, so that they could be used later for the $G(x)$ calculation.

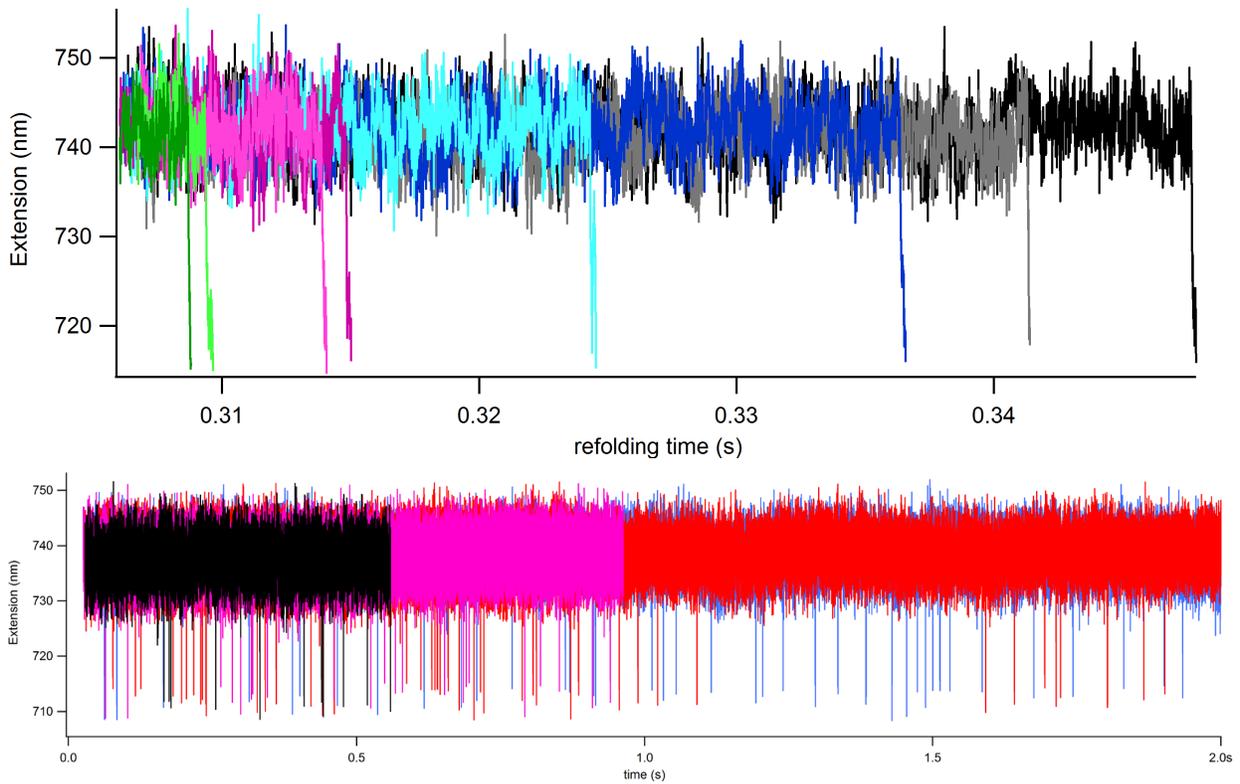


Figure 8.1 example extension pieces

Top: Examples of individual collapsing traces, overlaid and distinguished by arbitrary colour. **Bottom:** concatenation of individual traces into single trace and aligning concatenated sets (HP 30R50/T4 100kHz sampling sets, 16.6 to 12.3pN force drop)

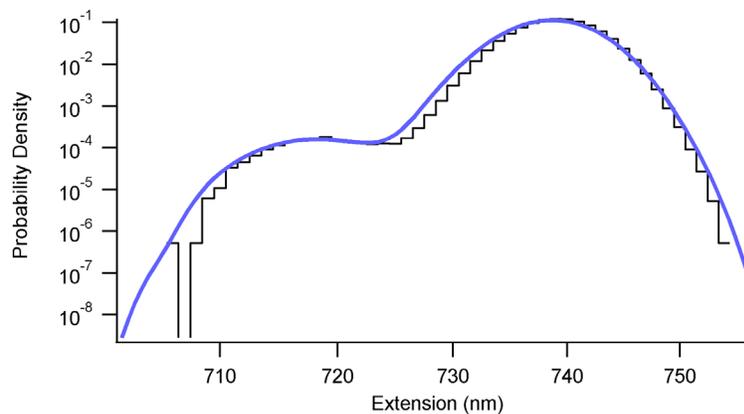


Figure 8.2 Example non-equilibrium $\rho(x)$, the histogram and KDE.

(Data from HP 30R50/T4 sampled 100kHz sampling sets, 16.6 to 12.3 pN force drop)

Using the KDE (for the non-equilibrium PDF), we calculated the free energy landscape from these data using equation (8.2), assuming a constant D , which is a reasonable approximation based on past work (Manuel et al., 2015). The result is shown in figure 8.3A. This landscape reconstruction was repeated for different choices of the initial and final forces on either side of $F_{1/2}$ (14.5 pN, (Woodside et al., 2006b)), but always with the same 4.4 pN force jump (figure 8.3B). We replicated this whole reconstruction procedure with force-jump data for hairpin 20TS06/T4, a shorter hairpin with a qualitatively different landscape (Woodside 2006). The resulting landscape reconstruction is shown in Figure 8.4; in these measurements, the force dropped from 14.3 pN, was twice as large, 9 pN.

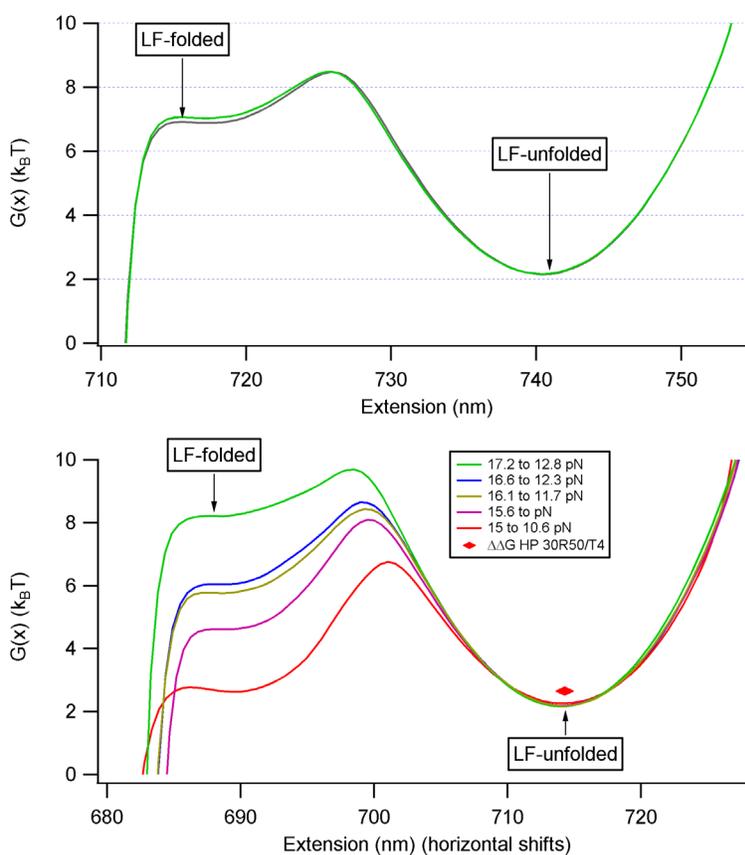


Figure 8.3 Energy landscape reconstruction for HP 30R50/T4

A: first example of reconstructed free energy as a function of extension. **B:** Different force drops horizontally offset to unfolded well for comparison, and the ‘◀▶’ marker for the calculated energy difference the bottom of the folded state for this 4.4 pN force difference (Woodside et al., 2006b) High force: 17.2, 16.6, 16.1, 15.6, 15 pN, low force: 12.8, 12.3, 11.7, 11.1, 10.6 pN, HP 30R50/T4 ($F_{1/2}$: 14.5pN), Data sampled at 100 kHz.

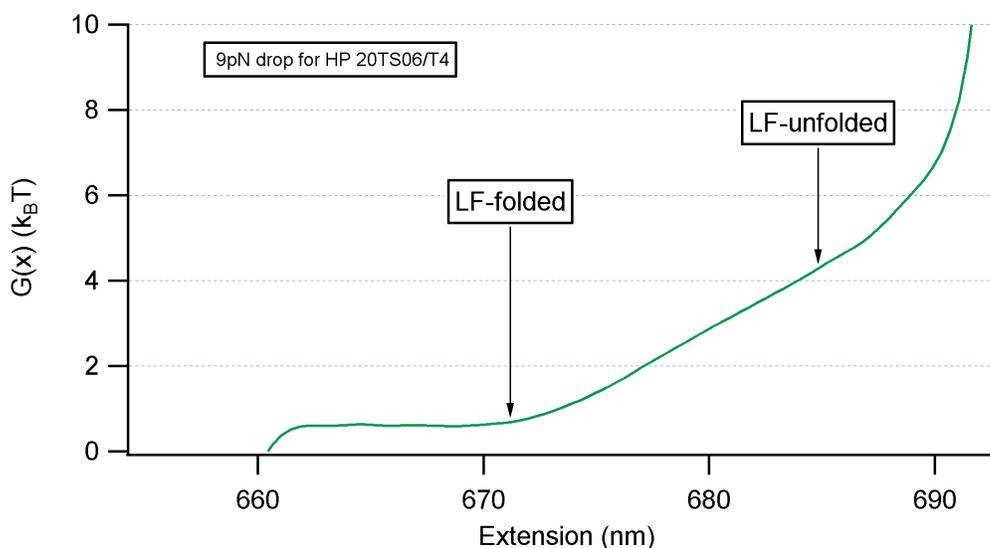


Figure 8.4 Example result for the non-equilibrium $G(x)$ calculation with HP 20TS06/T4.

9 pN drop, from 14.3 pN for HP 20TS06/T4 Data sampled at 100 kHz.

Considering first the results for hairpin 30R50/T4, we note that they differ from what we would expect in two important ways. First, although the landscapes reconstructed by this method do feature two states separated by a barrier, the energy of the folded state is higher than the energy of the unfolded state, even though the force at which the landscape was reconstructed (the low-force value after the jump) was in every case below $F_{1/2}$ and hence the opposite must have been true: folded state should have lower energy than the unfolded state. (Another way of saying the same thing is that the red curve in Figure 8.3 implies that $F_{1/2}$ is ~ 10.5 pN, several pN below the actual value (14.5 pN).) Second, the energy barrier is located closer to the folded state than the unfolded, the opposite of what was found in multiple reconstructions based on different approaches: from FECs using extensions of the Jarzynski equality (Gupta et al., 2011), the weighted histogram and inverse Weierstrass transform methods (Engel et al., 2014), and from constant-force measurements using the inverse Boltzmann transform (Woodside et al., 2006a; Woodside et al., 2006b) and splitting probability analysis (Manuel et al., 2015). The results from the first three of these previous methods (all reconstructions at $F_{1/2}$) are shown in Figure 8.5—whereas all three agree very well, they collectively disagree both quantitatively and qualitatively with the results in Figure 8.3. In the case of hairpin 20TS06/T4, similar considerations hold, although here it is in fact difficult to discern any barrier at all (other reconstruction methods reveal a barrier located closer to the folded state than the unfolded). Again, the relative energy of the folded state is considerably higher than it

should be, considering the force at which the landscape is reconstructed (5.7 pN below $F^{1/2}$ (11 pN) for this hairpin).

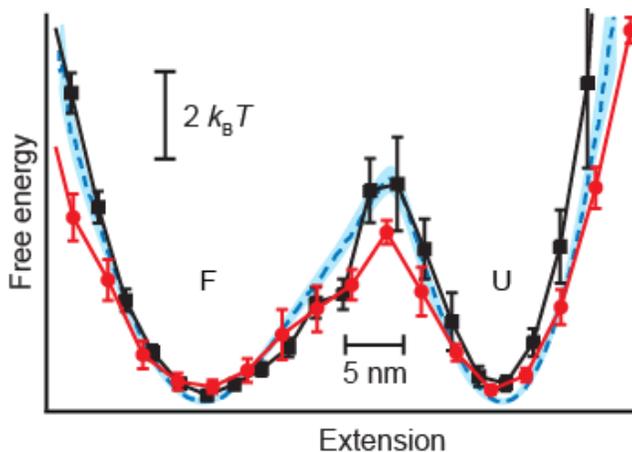


Figure 8.5 Profile of the free-energy landscape along the molecular extension of HP 30R50/T4.

The profile calculated from the force-extension curves using a weighted histogram (black) agrees well with the profile found from constant-force measurements via an inverse Boltzmann transform (dashed line). The profile found using the inverse Weierstrass transform approach (red) agrees reasonably well with the other two profiles, but underestimates the barrier height slightly owing to incomplete defiltering of smoothing caused by the force probe. F: folded, U: unfolded. Error bars and shaded blue region show standard errors (Engel et al., 2014).

To determine if the problems with the reconstructions came from assuming that D is constant in equation (8.2), we recalculated the landscape without this assumption with equation (8.1). We used the position dependent diffusion determined from the force-jump fall-time analysis described in section 7.4, since this calculation should be self-consistent with the calculation of $G(x)$ (Lannon et al., 2013). Not surprisingly, given the fact that the fall-time analysis did not appear to yield a good estimate of $D(x)$ (Figure 8.6), the landscape profile derived in this way was not an improvement over the initial estimate (Figure 8.7). Indeed, the discontinuities in $D(x)$ produced unphysical discontinuities in $G(x)$. Attempts to remove the discontinuities by smoothing were not successful (Figure 8.6).

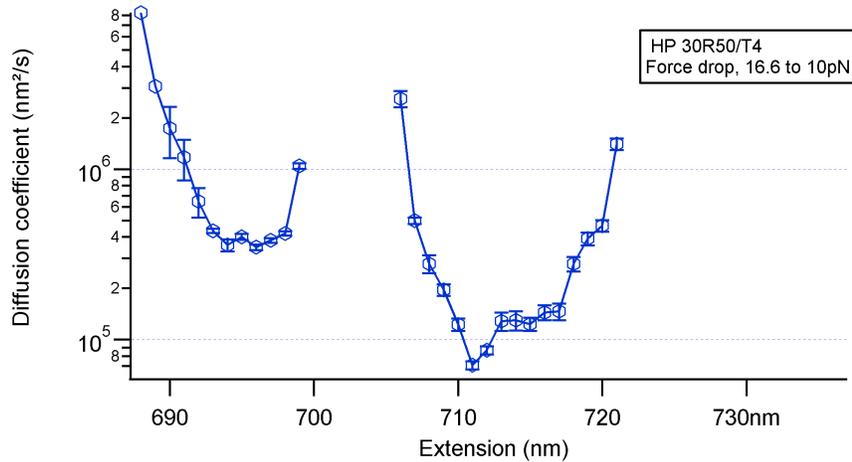


Figure 8.6 Example of diffusion coefficient from force jumps.

D in (nm^2/s) calculated from the average fall time $\tau_C(x)$ for HP 30R50/T4 (data sampled at 100 kHz) Its range is from the unfolded state on the right to the first arrival at the folded state on the left (short extension).

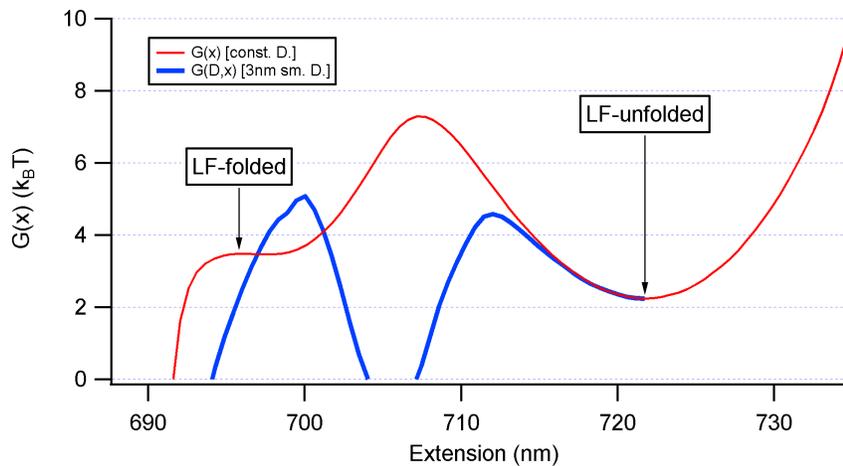


Figure 8.7 Comparison of energy landscape with and without position dependence in D .

Example of calculation of $G(D,x)$, from equation 8.3, using a 3 nm box smoothed $D(x)$ from 8.7, left panel. (Force drop, 16.6 to 10 pN $F_{1/2}$: 14.5 pN)

8.4 Analysis of the $G(x)$ calculations

In an attempt to understand the source of the disagreement between the force-jump reconstructions and the other reconstruction methods, we explored various features of the data analysis that might affect the result. We tried altering the input non-equilibrium PDF for the $G(x)$ calculation by

manually truncating the extension distribution at the point where the folded state is first reached, using both the calculated KDE as well as histograms with 1-nm and 1-ångström bins (Figure 8.8). However, little change was seen when switching from the KDE to histograms with 1-nm bins, and limitations in the size of our datasets distorted the calculation when attempting to use 1-ångström bins.

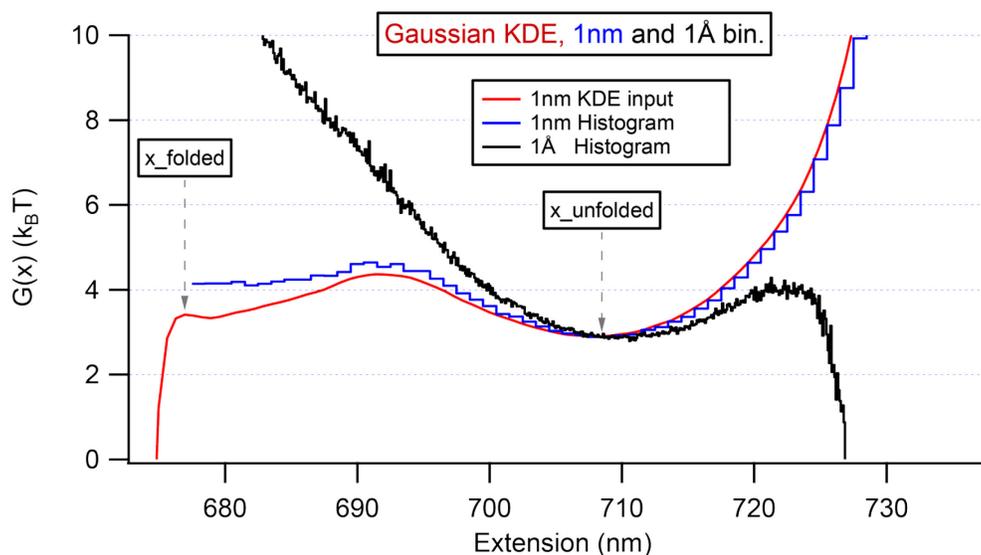


Figure 8.8 Manually restricting the extension distribution for $G(x)$ calculation.

The concatenated extension traces are manually cut off where the folded state is first reached before calculating the non-equilibrium PDFs in the form of histograms and KDEs. (HP 30R50/T4, 10pN force drop from 16.5pN, 100 kHz sampling)

We also tried modifying the strength of the correction term on the RHS of equation (8.2), ranging from 1-fold to 2-fold, in case it was being misestimated in our calculation. An additional weight factor of 5/3-fold appeared to provide an empirical fix correcting for the relative energy expected for the folded state relative to the unfolded state after a 4.5 pN drop (Figure 8.9 top). With this correction, the barrier location, notwithstanding, remains displaced from the barrier position of the tilted PMF from constant-force data on the same hairpin (Figure 8.9 middle). However, there lacks a good justification for this scaling, and indeed the required scaling is a function of force, as can be seen from applying the same exercise to calculations based on a larger, 10 pN force-jump for the same hairpin. In this latter case, a larger multiplier of ~4-fold is needed (Figure 8.9 bottom). We replicated the above rescaling procedure

with the data for hairpin 20TS06/T4, and an additional weight factor was increased to 5/3 produced results similar to before, for a 7 pN force drop, bringing the location of the folded state to approximately the right place, but lacking a defined barrier (Figure 8.10). A similar correction was attempted with a more continuous $G(D(x),x)$ result from a 10pN force drop (equation 8.1), increasing the multiplied weight on the correcting term on the RHS of the equation (Figure 8.11). Here the weight has to be increased similarly (4× for a 10 pN drop) to the free energy that assumed a constant diffusion, so the misalignment of the folded state in the energy landscape appears to be similar whether or not a constant D is assumed.

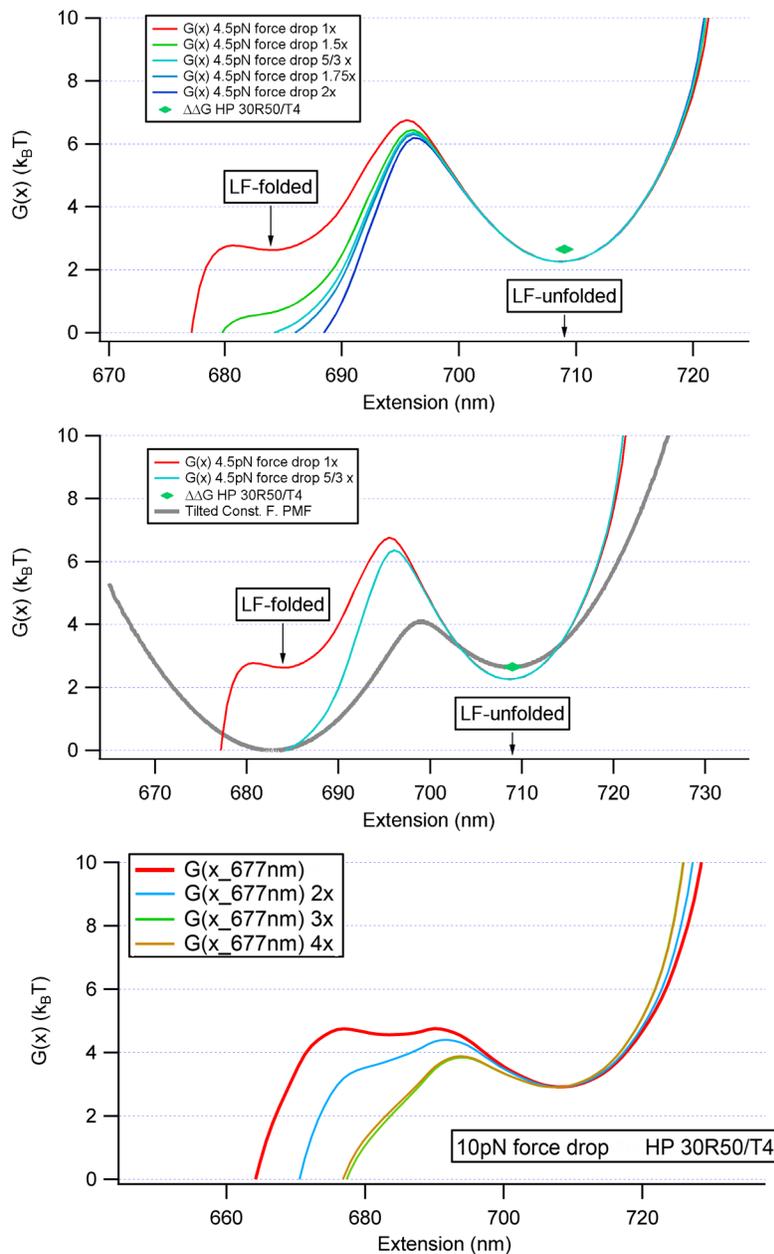


Figure 8.9 Varying the multiplier of RHS term in $G(x)$.

Top – varying the multiplier on the correcting term on the RHS for $G(x)$ in equation (8.2), from the original 1 \times , to 3/2, 5/3, 1.75, and 2 \times . 5/3rds comes closest to collapsing to the folded state (4.5 pN force drop), with the calculated energy level difference from unfolded to folded for this force difference: $\blacktriangleleft\blacktriangleright$. **Middle** – with the force-tilted PMF from constant force data at $F_{1/2}$ overlaid in grey to show the barrier location (250 kHz sampling). **Bottom:** 10 pN force drop, multipliers 1-4 \times , calculation folded target 677 nm, unfolded well: 708.5 nm

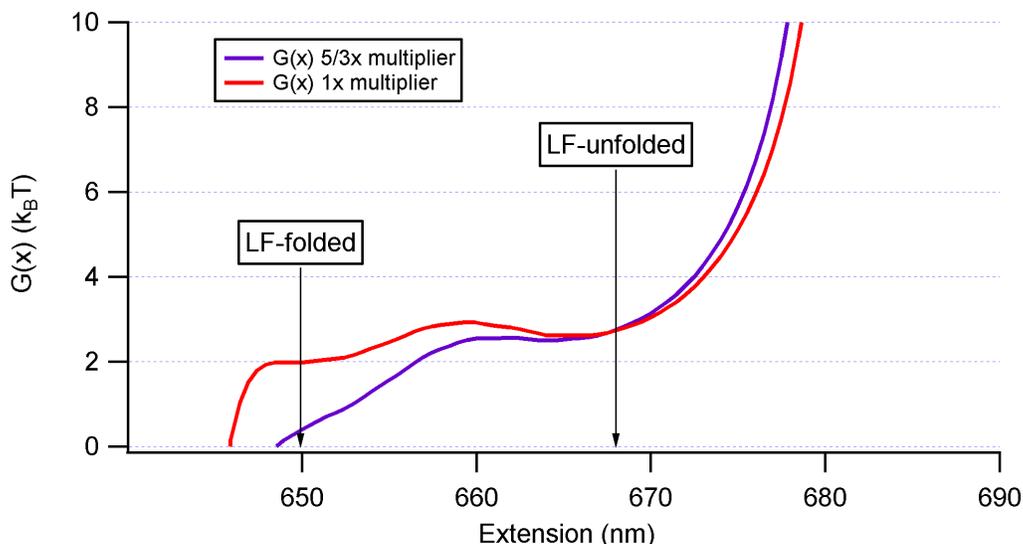


Figure 8.10 The results for the $G(x)$ calculation with HP 20TS06/T4.

7 pN force drop from 14.5 pN with the effect to increasing the multiplier to 5/3 on the correcting term on the RHS of equation (8.2) as was done for HP 30R50/T4. (Data sampled at 100 kHz.)

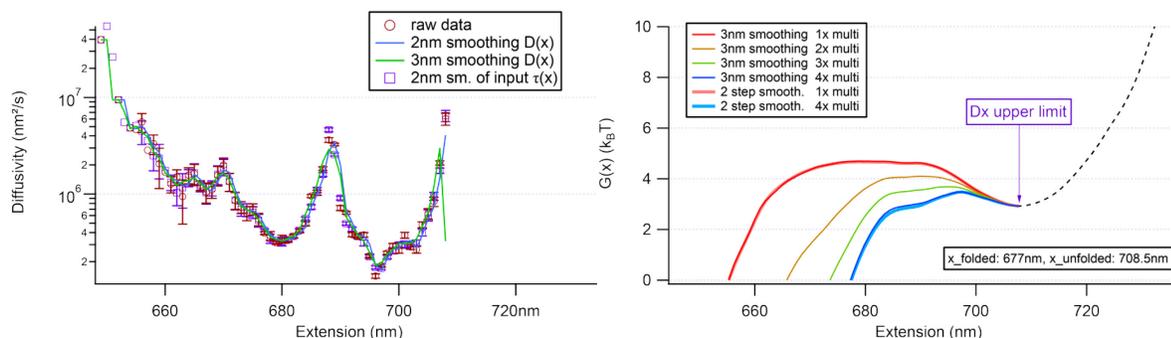


Figure 8.11 Varying the multiplier of RHS term in $G(D,x)$.

Similarly to the previous figures, the multiplier on the correcting term on the RHS for $G(D,x)$ in equation (8.3) was varied from the original 1× up to 4×. (HP 30R50/T4, 10 pN force drop, 100 kHz sampling)

In the above results, increasing the weight of the correction factor on the RHS of equation (8.1), with or without a constant diffusion, can bring the end point of the curve to the folded state's location, relative to the unfolded state's position, in extension. Increasing the multiplier does not address any other qualities in the $G(x)$ calculated from the non-equilibrium PDFs, i.e. for HP 30R50/T4, the energy level of the folded state above the zero level in the force jump landscape, should be 2.65 $k_B T$ or 6.57 kJ/mol for a jump down from 14.9 to 10.5 pN (see marker '◀▶' Figure 8.3 bottom, Figure 8.9) (Woodside et al., 2006a; Woodside et al., 2006b). As was mentioned in section 7.4, the calculation of the diffusion

coefficient from the average fall times is limited by the appearance of any negative slope in the non-equilibrium PDF. Lastly, Zhang et al. presented a form for $G(x)$ that considers a generalisation to a random initial point for the refolding trajectories.

$$G(x) = -k_B T \left[\ln \rho(x) + \rho'(x_{refolded}) \int_x^{x_{unfolded}} \frac{\int_{x'}^{\infty} \rho_0(z) dz}{\rho(x')} dx' \right] \quad (8.4)$$

where an initial distribution of the random starting points, $\rho_0(x)$ is placed in a nested integral (Zhang et al., 2011). When this form is applied to data from HP 30R50/T4, using the initial unfolded length distribution for $\rho_0(x)$, the final $G(x)$ curve differed by less than 0.1nm compared to previous calculations for $G(x)$.

8.5 Instrumental effects on the analysis

The *ad hoc* correction factors explored above that are needed to make the reconstructions agree with the bare minimum of the facts of the transition, namely the approximate energy changes between the states and their locations with respect to extension, suggests that something is missing from the analytical theory being applied to the data. Given that we know that instrumental factors like the trap stiffness and the attachments to the handles and beads can affect analysis of folding trajectories (Engel et al., 2014; Woodside et al., 2006b; Woodside et al., 2014), these seem to be likely candidates. In the analysis above, no corrections for the effects of the beads and the handles on the system have been implemented, for example as done for equilibrium landscape reconstructions using dynamic mechanical network models of the measurement (Hinczewski et al., 2013; Hinczewski et al., 2010b). It is not unreasonable to suppose that both the dynamical properties of the handles/beads and their compliance properties may be convolved with the intrinsic dynamic response of the molecule in such a way as to confound the landscape reconstruction from force-jump data. Indeed, as pointed out by Nam and Makarov, the molecular dynamics observed in smFS measurements can be strongly affected by the hydrodynamic drag on the beads in OTs, or the cantilever in the case of AFM (Nam and Makarov, 2015). The bead can slow the motion of the molecule and act as a low pass filter on the molecular trajectory. Moreover, in our case, we are in the regime where our linkers (dsDNA handles) are more compliant than

the molecule itself, so we could estimate that the molecular motion is decoupled from the force probe during a transition event. In that case then the intrinsic molecular coefficient can be recovered from a correction factor provided that the stiffnesses of the handles and molecule are known (Makarov, 2014).

Lastly, a potential limitation in our measurements may be in the time resolution. DNA hairpins are believed to fold up quickly by a zippering mechanism (Neupane et al., 2012), which might constrain the observable timescale over which the DNA hairpin should be measured, as Best and Hummer point out: the sampling time must be shorter than the time it takes for the molecule to cross its folding barrier (Best and Hummer, 2011). We could sample the DNA hairpins faster than we have, but as was seen in Chapter 6 for the direct measurements of τ_{fp} , we had to maximize the trap stiffnesses to reach a 6 μ s time resolution with the bead and handle response. Increasing the trap stiffnesses in a similar manner would prevent the use of the constant force modality of the trap, so that the refolding could not take place at constant force. This would narrow the non-equilibrium distributions further, hampering the reconstruction of the energy landscape.

In conclusion, we have been unable to adequately reconstruct the energy landscape with the non-equilibrium technique as applied to our experimental conditions. This suggests that the method cannot be used reliably until the effects of the force probe and linkers have been more thoroughly explored. Intuitively, it appears that this method might be the most sensitive to issues of handles and beads, since the non-equilibrium distributions we are examining will be altered by both the compliance (as they do to equilibrium distributions) and the dynamics (of which do not affect those distributions). It may be of value to measure a larger DNA hairpin, perhaps with designed intermediate states, to see how the reconstructed non-equilibrium landscape would turn out. Otherwise, the results of this method cannot be parsed with previous reconstruction results from equilibrium measurements and FECs, suggesting that the approach is incomplete, and must most likely compensate for the compliance and dynamics of the beads and handles.

Chapter 9: Summary and Future work

9.1 Summary

In this thesis, single molecule force spectroscopy was first applied to study the folding of nucleic acids. Looking at RNA molecules, I investigated how folding relates to their regulatory function. In the case of the *add* adenine riboswitch, gene regulation involves a switch between ‘on’ and ‘off’ conformations, where the ‘on’ state becomes thermodynamically favoured with the binding of the adenine ligand to the pocket formed by the three helix junction. In the case of the –1 PRF pseudoknots, we discovered that the hypothesis that the frameshifting efficiency correlates with the mechanical strength of the pseudoknot is incorrect. The frameshifting efficiency instead correlates with the propensity of the pseudoknot to sample alternate conformations, suggesting a more complex mode of interaction of the pseudoknot with the ribosome.

I also used force spectroscopy to test the physical picture of biomolecular folding, as diffusion over a conformational energy landscape, in unprecedented detail at the level of single molecules. Using DNA hairpins as a model system, I tested a novel method for reconstructing energy landscapes from force-jump measurements, as well proposed methods for determining the position-dependent diffusion coefficient governing the molecular dynamics from both equilibrium and non-equilibrium data. In all cases, we found that these new methods produced results that were inconsistent with expectations, likely because the effects of the beads and linkers attached to the molecules being studied are not accounted for in the theory. Work remains to be done to understand better how instrumental effects from linkers and the force probes themselves alter the experimental outcomes, so that the intrinsic molecular properties can be properly recovered.

Finally, the transition time over the barrier in the energy landscape was calculated for the first time from energy-landscape measurements, and the first direct measurements of transition time distributions were made. This work opens up a new frontier, allowing experimentalists to examine the dynamics of folding in microscopic detail. There are many opportunities to test the basic physics of folding by transition path measurements for different molecules.

9.2 Future work

9.2.1 Folding and function in RNA

Several interesting questions remained to be explored regarding the folding of riboswitches across the board. In the *add* riboswitch chapter, the switching mechanism (thermodynamic or kinetic), between the ‘off’ and ‘on’ states, was discussed with arguments about the timing of the ribosome’s binding to the RBS, and how it limits the equilibration time. The time to transcribe from the riboswitch would require ~ 0.3 s, which was within the uncertainty of our experimental equilibration time. We have further worked with our OT to improve our time resolution, with the use of high stiffness, which could be used to probe beyond that time limitation. Moreover, as was discussed in chapter 2, the force jump method using the EODs would be applied here, by moving the lower trap, to circumvent the fast signal change issues of the upper trap (see Appendix A).

With the improved time resolution of force jump experiments we would be able to more clearly construct the equilibration curve that is shown in Figure 4.10b. Moreover, the effect of adenine concentration on the shifts between the “on” and “off” populations of the riboswitch could be examined with ligand titration curves using both FECs and force jump measurements (as has been done with *pbuE* (Greenleaf et al., 2008) and TPP riboswitch (Anthony et al., 2012)). To examine the state switching (“on”/“off”), the full length sequence would be measured with constant force and constant trap separation. Extension changes may provide insight into how the stem P1 is pulled apart so that the sequestering hairpin can form, and vice versa. The mechanism by which the state switching occurs has not yet been studied in detail for *add*, and we aim to observe how the state switching occurs with these experiments. This would complement the OT study of the related full length *pbuE* riboswitch by the Block lab (Frieda and Block, 2012), which demonstrated a kinetic control mechanism.

Some disagreements remain for the *add* riboswitch (Savinov et al., 2014): in AFM experiments, the “off” state was claimed to include a pseudoknot that further stabilises the structure (Heus et al., 2011), and given their large uncertainties, OTs measurements with the full length structure should be able to further probe alternate structures are possible.

Since the purine riboswitches have among the simplest structures of all known riboswitches (apart from the preQ1)(Kim and Breaker, 2008), it would be useful to study other riboswitches with more complex structures. This could provide a deeper understanding of the interplay between structure and function in these molecules. A riboswitch of particular interest is the SAM-I leader sequence of *Thermoanaerobacter tengcongensis* (Montange and Batey, 2006). This riboswitch is built around a four-helix junction and a tertiary pseudoknot (figure 9.1) (Gilbert et al., 2006a) (Wang and Breaker, 2008). A related SAM-I riboswitch has been studied by smFRET (Heppell et al., 2011), with varying magnesium concentrations and then on some mutants (Eschbach et al., 2012). This structure would offer a unique laboratory for a detailed understanding of the folding of relatively complex RNA molecules through smFS, complementing and refining previous studies. Being more complex, this particular SAM riboswitch may also offer more opportunities to observe differences between RNAs folded co-transcriptionally and those refolded after transcription.

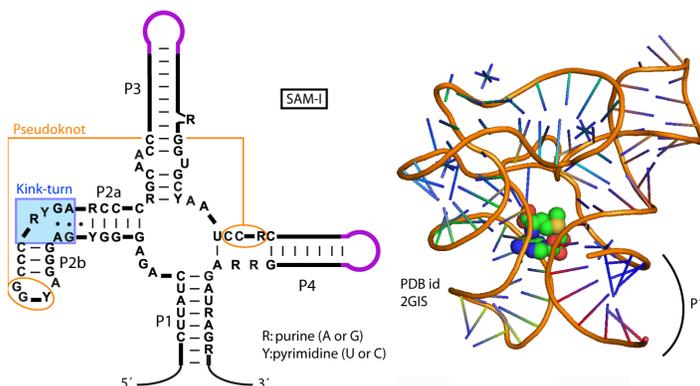


Figure 9.1 Structure of the SAM-I riboswitch aptamer (Wang and Breaker, 2008).

Many of the future approaches to the riboswitches and frameshifting pseudoknots would be similar. Structural studies from crystallography and NMR would not show transient intermediates or alternate conformations, and would hide a complex energy landscape. Constant trap separation experiments can be used to better elucidate intermediate states that have so far been studied with partial constructs. Rather than hybridizing to the 3' and 5' ends of the RNA, other sites could be to conjugate

handles by the use of techniques such as ‘click’ chemistry (Yu et al., 2012c), which could offer the placement of a third handle on a molecule, as we have so far been limited to pulling only at the end points. Another extension to our experiments with optical tweezers could be with smFRET (Savinov et al., 2014), for example, to track the interactions of loops.

9.2.2 Diffusion and landscape reconstruction

The generalisation of a constant coefficient of diffusion is unlikely to be true for molecules more complex than DNA hairpins, and since diffusion a basic folding property of proteins and nucleic acids, it is highly desirable to ascertain its position dependence (notwithstanding that D may also be a function of both position and velocity/time, e.g. $D = D(x, dx/dt)$).

The transition path studies open up many exciting avenues of study. Looking at the transition paths directly could offer the opportunity to see where pauses occur (representing microstates in the barrier region). It offers the chance to study diffusion at the microscopic scale, perhaps offering a direct measure of the position dependence, as well as the characterisation of different barriers.

The proper recovery of the properties of a molecule within a measurement system will require more study to understand how the instrument’s properties, such as the probe compliance, affect the measurements. This was illustrated in our work with the round trip time diffusion and the non-equilibrium reconstruction from force jumps techniques, which have not been successful with our experimental data. In a similar vein to the constant force Langevin dynamics simulation that have been done with and without beads and handles (Woodside et al., 2014), the generation of refolding traces from a force jump could be generated to attempt to clarify where the theory of the Brujic group encounters difficulty with our data (Lannon et al., 2013; Zhang et al., 2011). It would then be straightforward to repeat the analysis on the simulated trajectories, so that we might be able to see more clearly which contributions (e.g. trap/bead and handle compliances) lead to the differences we observed with our experimental data.

In particular to the divergences of the round trip time diffusion method, there were dramatic differences between the two reference starting points (from unfolded to folded and back, and vice-versa, e.g. Figure 7.6 and Figure 7.7). They seem to likely be from instrumental and compliance effects, as we expect that the DNA hairpins should exhibit Markovian behaviour. These difficulties notwithstanding, many opportunities have opened up to test basic physics of folding through transition path measurements.

Bibliography

- Abels, J.A., Moreno-Herrero, F., van der Heijden, T., Dekker, C., and Dekker, N.H. (2005). Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys J* *88*, 2737-2744.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2008). *Molecular Biology of the Cell*, 5th Edition edn (Garland Science).
- Andronescu, M., Aguirre-Hernandez, R., Condon, A., and Hoos, H.H. (2003). RNAsoft: A suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Res* *31*, 3416-3422.
- Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* *181*, 223-230.
- Anfinsen, C.B., Haber, E., Sela, M., and White, F.H., Jr. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America* *47*, 1309-1314.
- Ansari, A., Kuznetsov, S.V., and Shen, Y. (2001). Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc Natl Acad Sci USA* *98*, 7771-7776.
- Anthony, P.C., Perez, C.F., García-García, C., and Block, S.M. (2012). Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proceedings of the National Academy of Sciences* *109*, 1485-1489.
- Ashkin, A., and Dziedzic, J.M. (1971). Optical Levitation by Radiation Pressure. *Applied Physics Letters* *19*, 283-285.
- Ashkin, A., Dziedzic, J.M., Bjorkholm, J.E., and Chu, S. (1986). Observation of a single-beam gradient force optical trap for dielectric particles. *Opt Lett* *11*, 288-290.
- Baird, N.J., and Ferré-D'Amaré, A.R. (2010). Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis *RNA* *16*, 598-609
- Baker, D. (2000). A surprising simplicity to protein folding. *Nature* *405*, 39-42.
- 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*, 498-510.

- Baranov, P.V., Wills, N.M., Barriscale, K.A., Firth, A.E., Jud, M.C., Letsou, A., Manning, G., and Atkins, J.F. (2011). Programmed ribosomal frameshifting in the expression of the regulator of intestinal stem cell proliferation, adenomatous polyposis coli (APC). *RNA biology* *8*, 637-647.
- Barsegov, V., Klimov, D.K., and Thirumalai, D. (2006). Mapping the energy landscape of biomolecules using single molecule force correlation spectroscopy: theory and applications. *Biophys J* *90*, 3827-3841.
- Batey, R.T., Gilbert, S.D., and Montange, R.K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* *432*, 411-415.
- Baumann, C.G., Bloomfield, V.A., Smith, S.B., Bustamante, C., Wang, M.D., and Block, S.M. (2000). Stretching of single collapsed DNA molecules. *Biophys J* *78*, 1965-1978.
- Belew, A.T., Meskauskas, A., Musalgaonkar, S., Advani, V.M., Sulima, S.O., Kasprzak, W.K., Shapiro, B.A., and Dinman, J.D. (2014). Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nature* *512*, 265-269.
- Bell, G.I. (1978). Models for the specific adhesion of cells to cells. *Science* *200*, 618-627.
- Berg, J.M., Tymoczko, J.L., and Stryer, L. (2010). *Biochemistry* (W. H. Freeman).
- Berkovich, R., Hermans, R.I., Popa, I., Stirnemann, G., Garcia-Manyes, S., Berne, B.J., and Fernandez, J.M. (2012). Rate limit of protein elastic response is tether dependent. *Proceedings of the National Academy of Sciences of the United States of America* *109*, 14416-14421.
- Best, R.B. (2012). Atomistic molecular simulations of protein folding. *Curr Opin Struct Biol* *22*, 52-61.
- Best, R.B., and Hummer, G. (2006). Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys Rev Lett* *96*, 228104.
- Best, R.B., and Hummer, G. (2010). Coordinate-dependent diffusion in protein folding. *P Natl Acad Sci USA* *107*, 1088-1093.
- Best, R.B., and Hummer, G. (2011). Diffusion models of protein folding. *Physical chemistry chemical physics : PCCP* *13*, 16902-16911.
- Blakely, B.L., Dumelin, C.E., Trappmann, B., McGregor, L.M., Choi, C.K., Anthony, P.C., Duesterberg, V.K., Baker, B.M., Block, S.M., Liu, D.R., *et al.* (2014). A DNA-based molecular probe for optically reporting cellular traction forces. *Nat Meth* *11*, 1229-1232.

- Blount, K.F., Megyola, C., Plummer, M., Osterman, D., O'Connell, T., Aristoff, P., Quinn, C., Chrusciel, R.A., Poel, T.J., Schostarez, H.J., *et al.* (2015). A Novel Riboswitch-Binding Flavin Analog that Protects Mice Against *Clostridium difficile* Infection without Inhibiting Cecal Flora. *Antimicrobial agents and chemotherapy*.
- Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews Genetics* *13*, 770-780.
- Boerneke, M.A., and Hermann, T. (2015). Ligand-responsive RNA mechanical switches. *RNA biology* *12*, 780-786.
- Bolhuis, P.G., Chandler, D., Dellago, C., and Geissler, P.L. (2002). Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem* *53*, 291-318.
- Borgia, A., Wensley, B.G., Soranno, A., Nettels, D., Borgia, M.B., Hoffmann, A., Pfeil, S.H., Lipman, E.A., Clarke, J., and Schuler, B. (2012). Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nature communications* *3*, 1195.
- Borgia, A., Williams, P.M., and Clarke, J. (2008). Single-molecule studies of protein folding. *Annual Review of Biochemistry* *77*, 101-125.
- Bouchiat, C., Wang, M.D., Allemand, J., Strick, T., Block, S.M., and Croquette, V. (1999). Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys J* *76*, 409-413.
- Breaker, R.R. (2012). Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology* *4*.
- Brierley, I., Gilbert, R., and Pennell, S. (2010). Pseudoknot-dependent programmed -1 ribosomal frameshifting: structures, mechanisms and models. *Nucleic Acids and Molecular Biology* *24*, 149-174.
- Brierley, I., Jenner, A.J., and Inglis, S.C. (1992). Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* *227*, 463-479.
- Brierley, I., Pennell, S., and Gilbert, R.J. (2007a). Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nature reviews Microbiology* *5*, 598-610.

- Brierley, I., Pennell, S., and Gilbert, R.J.C. (2007b). Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat Rev Microbiol* *5*, 598-610.
- Bryngelson, J.D., and Wolynes, P.G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* *84*, 7524-7528.
- Buchner, J., and Kiefhaber, T. (2005). *Protein Folding Handbook* (Wiley-VCH).
- Buck, J., Fürtig, B., Noeske, J., Wöhnert, J., and Schwalbe, H. (2007). Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proc Natl Acad Sci USA* *104*, 15699-16704.
- Buck, J., Noeske, J., Wöhnert, J., and Schwalbe, H. (2010). Dissecting the influence of Mg²⁺ on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res* *38*, 4143-4153.
- Bustamante, C., Marko, J.F., Siggia, E.D., and Smith, S. (1994). Entropic elasticity of lambda-phage DNA. *Science* *265*, 1599-1600.
- Byun, Y., and Han, K. (2009). PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics (Oxford, England)* *25*, 1435-1437.
- Cao, Q., Teng, Y., Yang, X., Wang, J., and Wang, E. (2015). A label-free fluorescent molecular beacon based on DNA-Ag nanoclusters for the construction of versatile Biosensors. *Biosensors & bioelectronics* *74*, 318-321.
- Cao, S., and Chen, S.J. (2008). Predicting ribosomal frameshifting efficiency. *Physical biology* *5*, 016002.
- Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., and Lafontaine, D.A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences* *109*, E3444-E3453.
- Cellmer, T., Henry, E.R., Hofrichter, J., and Eaton, W.A. (2008). Measuring internal friction of an ultrafast-folding protein. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 18320-18325.
- Chahine, J., Oliveira, R.J., Leite, V.B., and Wang, J. (2007). Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 14646-14651.

- Chamorro, M., Parkin, N., and Varmus, H.E. (1992). An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proceedings of the National Academy of Sciences of the United States of America* *89*, 713-717.
- Chaudhury, S., and Makarov, D.E. (2010). A harmonic transition state approximation for the duration of reactive events in complex molecular rearrangements. *The Journal of chemical physics* *133*, 034118.
- 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. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 12706-12711.
- Chen, G., Wen, J.D., and Tinoco, I., Jr. (2007). Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *RNA* *13*, 2175-2188.
- Chen, X., Chamorro, M., Lee, S.I., Shen, L.X., Hines, J.V., Tinoco, I., Jr., and Varmus, H.E. (1995). Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting. *Embo J* *14*, 842-852.
- Chen, X., Kang, H., Shen, L.X., Chamorro, M., Varmus, H.E., and Tinoco, I., Jr. (1996). A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA. *J Mol Biol* *260*, 479-483.
- Chung, H.S., and Eaton, W.A. (2013). Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* *502*, 685-688.
- Chung, H.S., Louis, J.M., and Eaton, W.A. (2009). Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 11837-11844.
- Chung, H.S., McHale, K., Louis, J.M., and Eaton, W.A. (2012). Single-molecule fluorescence experiments determine protein folding transition path times. *Science* *335*, 981-984.

- Chung, H.S., Piana-Agostinetti, S., Shaw, D.E., and Eaton, W.A. (2015). Structural origin of slow diffusion in protein folding. *Science* *349*, 1504-1510.
- Cocco, S., Marko, J.F., and Monasson, R. (2003). Slow nucleic acid unzipping kinetics from sequence-defined barriers. *Eur Phys J E Soft Matter* *10*, 153-161.
- Coppins, R.L., Hall, K.B., and Groisman, E.A. (2007a). The intricate world of riboswitches. *Curr Opin Microbiol* *10*, 176-181.
- Coppins, R.L., Hall, K.B., and Groisman, E.A. (2007b). The intricate world of riboswitches. *Current opinion in microbiology* *10*, 176-181.
- Cornish, P.V., Hennig, M., and Giedroc, D.P. (2005). A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 12694-12699.
- Cornish, P.V., Stammler, S.N., and Giedroc, D.P. (2006). The global structures of a wild-type and poorly functional plant luteoviral mRNA pseudoknot are essentially identical. *RNA* *12*, 1959-1969.
- Crooks, G.E. (1999). Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E* *60*, 2721-2726.
- Dam, E., Pleij, K., and Draper, D. (1992). Structural and functional aspects of RNA pseudoknots. *Biochemistry* *31*, 11665-11676.
- Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* *14*, 796-806.
- de Messieres, M., Chang, J.C., Belew, A.T., Meskauskas, A., Dinman, J.D., and La Porta, A. (2014). Single-molecule measurements of the CCR5 mRNA unfolding pathways. *Biophys J* *106*, 244-252.
- Dill, K.A., and MacCallum, J.L. (2012). The Protein-Folding Problem, 50 Years On. *Science* *338*, 1042-1046.
- Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikl, T.R. (2008). The protein folding problem. *Annu Rev Biophys* *37*, 289-316.
- Dinman, J.D. (2006). Programmed Ribosomal Frameshifting Goes Beyond Viruses: Organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift. *Microbe (Washington, DC)* *1*, 521-527.

- Dinman, J.D., Icho, T., and Wickner, R.B. (1991). A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proceedings of the National Academy of Sciences of the United States of America* *88*, 174-178.
- 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*, 3669-3676.
- Dolgin, E. (2015). A cellular puzzle: The weird and wonderful architecture of RNA. *Nature* *523*, 398-399.
- Dong, J., Castro, C.E., Boyce, M.C., Lang, M.J., and Lindquist, S. (2010). Optical trapping with high forces reveals unexpected behaviors of prion fibrils. *Nature structural & molecular biology* *17*, 1422-1430.
- Dudko, O.K., Graham, T.G.W., and Best, R.B. (2011). Locating the Barrier for Folding of Single Molecules under an External Force. *Physical Review Letters* *107*, 208301.
- Dudko, O.K., Hummer, G., and Szabo, A. (2006). Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical Review Letters* *96*, 108101-108104.
- Dudko, O.K., Hummer, G., and Szabo, A. (2008). Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 15755-15760.
- 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*, 127-136.
- Elms, P.J., Chodera, J.D., Bustamante, C.J., and Marqusee, S. (2012). Limitations of constant-force-feedback experiments. *Biophys J* *103*, 1490-1499.
- Engel, M.C., Ritchie, D.B., Foster, D.A.N., Beach, K.S.D., and Woodside, M.T. (2014). Reconstructing Folding Energy Landscape Profiles from Nonequilibrium Pulling Curves with an Inverse Weierstrass Integral Transform. *Physical Review Letters* *113*, 238104.
- Englander, S.W., and Mayne, L. (2014). The nature of protein folding pathways. *Proceedings of the National Academy of Sciences of the United States of America* *111*, 15873-15880.
- Eschbach, S.H., St-Pierre, P., Penedo, J.C., and Lafontaine, D.A. (2012). Folding of the SAM-I riboswitch: a tale with a twist. *RNA biology* *9*, 535-541.

- Evans, E., and Ritchie, K. (1997). Dynamic strength of molecular adhesion bonds. *Biophysical Journal* *72*, 1541-1555.
- Evans, E., Ritchie, K., and Merkel, R. (1995). Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces. *Biophysical Journal* *68*, 2580-2587.
- Farabaugh, P.J. (1996). Programmed translational frameshifting. *Microbiol Rev* *60*, 103-134.
- Farabaugh, P.J. (2012). *Programmed Alternative Reading of the Genetic Code: Molecular Biology Intelligence Unit* (Springer US).
- Foster, D.A.N. (2010). High resolution optical tweezers for single molecule studies of hierarchical folding in the *pbuE* riboswitch aptamer (Edmonton: University of Alberta).
- Frederickx, R., in't Veld, T., and Carlon, E. (2014). Anomalous Dynamics of DNA Hairpin Folding. *Physical Review Letters* *112*, 198102.
- Freedman, L.P., Cockburn, I.M., and Simcoe, T.S. (2015). The Economics of Reproducibility in Preclinical Research. *PLoS Biol* *13*, e1002165.
- Frieda, K.L., and Block, S.M. (2012). Direct observation of cotranscriptional folding in an adenine riboswitch. *Science* *338*, 397-400.
- Garai, A., Zhang, Y., and Dudko, O.K. (2014). Conformational dynamics through an intermediate. *The Journal of chemical physics* *140*, 135101.
- Gebhardt, J.C.M., Bornschlöggl, T., and Rief, M. (2010). Full distance-resolved folding energy landscape of one single protein molecule. *Proceedings of the National Academy of Science USA* *107*, 2013-2018.
- Giedroc, D.P., and Cornish, P.V. (2009). Frameshifting RNA pseudoknots: structure and mechanism. *Virus research* *139*, 193-208.
- Giedroc, D.P., Theimer, C.A., and Nixon, P.L. (2000). Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol* *298*, 167-185.
- Gilbert, S.D., Montange, R.K., Stoddard, C.D., and Batey, R.T. (2006a). Structural studies of the purine and SAM binding riboswitches. *Cold Spring Harbor symposia on quantitative biology* *71*, 259-268.
- Gilbert, S.D., Stoddard, C.D., Wise, S.J., and Batey, R.T. (2006b). Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol* *359*, 754-768.

- Gopich, I.V., Nettels, D., Schuler, B., and Szabo, A. (2009). Protein dynamics from single-molecule fluorescence intensity correlation functions. *The Journal of chemical physics* *131*, 095102.
- 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.
- Greenleaf, W.J., Frieda, K.L., Foster, D.A.N., Woodside, M.T., and Block, S.M. (2008). Direct observation of hierarchical folding in single riboswitch aptamers. *Science* *319*, 630-633.
- Greenleaf, W.J., Woodside, M.T., Abbondanzieri, E.A., and Block, S.M. (2005). Passive all-optical force clamp for high-resolution laser trapping. *Phys Rev Lett* *95*, 208102.
- Greenleaf, W.J., Woodside, M.T., and Block, S.M. (2007). High-resolution, single-molecule measurements of biomolecular motion. *Annu Rev Biophys Biomol Struct* *36*, 171-190.
- Grigg, J.C., and Ke, A. (2015). Chapter Eight - Structures of Large RNAs and RNA-Protein Complexes: Toward Structure Determination of Riboswitches. In *Methods in enzymology*, A.W. Sarah, and H.T.A. Frédéric, eds. (Academic Press), pp. 213-232.
- Grønbech-Jensen, N., and Farago, O. (2013). A simple and effective Verlet-type algorithm for simulating Langevin dynamics. *Molecular Physics* *111*, 983-991.
- 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. *Nat Phys* *7*, 731-736.
- Guillaume, B., and Tom, C. (2004). On the reconstruction of diffusions from first-exit time distributions. *Inverse Problems* *20*, 1053.
- Guliaev, A.B., Cheng, S., and Hang, B. (2012). Protein dynamics via computational microscope. *World journal of methodology* *2*, 42-49.
- Gupta, A.N., Vincent, A., Neupane, K., Yu, H., Wang, F., and Woodside, M.T. (2011). Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat Phys* *7*, 631-634.
- Hagen, S.J. (2010). Solvent viscosity and friction in protein folding dynamics. *Current protein & peptide science* *11*, 385-395.
- Hänggi, P., Talkner, P., and Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Rev Mod Phys* *62*, 251-341.

- Hansen, T.M., Reihani, S., Oddershede, L.B., and Sorensen, M.A. (2007). Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 5830–5835.
- Harris, N.C., Song, Y., and Kiang, C.-H. (2007). Experimental Free Energy Surface Reconstruction from Single-Molecule Force Spectroscopy using Jarzynski's Equality. *Physical Review Letters* *99*, 068101.
- Heidarsson, P.O., Naqvi, M.M., Otazo, M.R., Mossa, A., Kragelund, B.B., and Cecconi, C. (2014). Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1. *Proceedings of the National Academy of Sciences* *111*, 13069-13074.
- Heppell, B., Blouin, S., Dussault, A.M., Mulhbachter, J., Ennifar, E., Penedo, J.C., and Lafontaine, D.A. (2011). Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nature chemical biology* *7*, 384-392.
- Heus, H.A., Puchner, E.M., van Vugt-Jonker, A.J., Zimmermann, J.L., and Gaub, H.E. (2011). Atomic force microscope-based single-molecule force spectroscopy of RNA unfolding. *Analytical biochemistry* *414*, 1-6.
- Hinczewski, M., Gebhardt, J.C., Rief, M., and Thirumalai, D. (2013). From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proceedings of the National Academy of Sciences of the United States of America* *110*, 4500-4505.
- Hinczewski, M., von Hansen, Y., Dzubiella, J., and Netz, R.R. (2010a). How the diffusivity profile reduces the arbitrariness of protein folding free energies. *The Journal of chemical physics* *132*, 245103.
- Hinczewski, M., von Hansen, Y., and Netz, R.R. (2010b). Deconvolution of dynamic mechanical networks. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 21493-21498.
- Hoffmann, A., and Woodside, M.T. (2011). Signal-pair correlation analysis of single-molecule trajectories. *Angewandte Chemie International Edition* *50*, 12643-12646.
- Holland, J.A., Hansen, M.R., Du, Z., and Hoffman, D.W. (1999). An examination of coaxial stacking of helical stems in a pseudoknot motif: the gene 32 messenger RNA pseudoknot of bacteriophage T2. *RNA* *5*, 257-271.

- Houck-Loomis, B., Durney, M.A., Salguero, C., Shankar, N., Nagle, J.M., Goff, S.P., and D'Souza, V.M. (2011). An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* *480*, 561-564.
- Howard, M.T., Gesteland, R.F., and Atkins, J.F. (2004). Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. *RNA* *10*, 1653-1661.
- Hummer, G. (2004). From transition paths to transition states and rate coefficients. *The Journal of chemical physics* *120*, 516-523.
- Hummer, G., and Szabo, A. (2001). Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Sciences of the United States of America* *98*, 3658-3661.
- Hummer, G., and Szabo, A. (2010). Free energy profiles from single-molecule pulling experiments. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 21441-21446.
- Huttenhofer, A., and Noller, H.F. (1994). Footprinting mRNA-ribosome complexes with chemical probes. *Embo J* *13*, 3892-3901.
- Hyeon, C., Hinczewski, M., and Thirumalai, D. (2014). Evidence of Disorder in Biological Molecules from Single Molecule Pulling Experiments. *Physical Review Letters* *112*, 138101.
- Hyeon, C., and Thirumalai, D. (2006). Forced-unfolding and force-quench refolding of RNA hairpins. *Biophys J* *90*, 3410-3427.
- Hyeon, C., and Thirumalai, D. (2008). Multiple Probes are Required to Explore and Control the Rugged Energy Landscape of RNA Hairpins. *Journal of the American Chemical Society* *130*, 1538-1539.
- Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* *55*, 447-458.
- Jacks, T., and Varmus, H.E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* *230*, 1237-1242.
- Jagannathan, B., Elms, P.J., Bustamante, C., and Marqusee, S. (2012). Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *P Natl Acad Sci USA* *109*, 17820-17825.

- Jagannathan, B., and Marqusee, S. (2013). Protein Folding and Unfolding Under Force. *Biopolymers* *99*, 860-869.
- Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters* *78*, 2690-2693.
- Kaji, T., Ito, S., Iwai, S., and Miyasaka, H. (2009). Nanosecond to Submillisecond Dynamics in Dye-Labeled Single-Stranded DNA, As Revealed by Ensemble Measurements and Photon Statistics at Single-Molecule Level. *J Phys Chem B* *113*, 13917-13925.
- Kang, H., Hines, J.V., and Tinoco, I., Jr. (1996). Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus. *J Mol Biol* *259*, 135-147.
- Ketteler, R. (2012). On programmed ribosomal frameshifting: the alternative proteomes. *Frontiers in Genetics* *3*, 242.
- Kim, J.N., and Breaker, R.R. (2008). Purine sensing by riboswitches. *Biology of the cell / under the auspices of the European Cell Biology Organization* *100*, 1-11.
- Kim, Y.G., Su, L., Maas, S., O'Neill, A., and Rich, A. (1999). Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency. *Proceedings of the National Academy of Sciences of the United States of America* *96*, 14234-14239.
- Kirithi, 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* *12*, 2080-2091.
- Kontos, H., Naphine, 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
- Kramers, H.A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* *7*, 284-304.
- Kreuzer, H.J., Payne, S.H., and Livadaru, L. (2001). Stretching a macromolecule in an atomic force microscope: statistical mechanical analysis. *Biophys J* *80*, 2505-2514.
- Kubelka, J., Hofrichter, J., and Eaton, W.A. (2004). The protein folding 'speed limit'. *Curr Opin Struct Biol* *14*, 76-88.

- Kuznetsov, Serguei V., and Ansari, A. (2012). A Kinetic Zipper Model with Intrachain Interactions Applied to Nucleic Acid Hairpin Folding Kinetics. *Biophys j* *102*, 101-111.
- Lannon, H., Haghpanah, J.S., Montclare, J.K., Vanden-Eijnden, E., and Brujic, J. (2013). Force-clamp experiments reveal the free-energy profile and diffusion coefficient of the collapse of protein molecules. *Phys Rev Lett* *110*, 128301.
- Lapidus, L.J. (2013). Exploring the top of the protein folding funnel by experiment. *Curr Opin Struct Biol* *23*, 30-35.
- Larsen, B., Gesteland, R.F., and Atkins, J.F. (1997). Structural probing and mutagenic analysis of the stem-loop required for Escherichia coli dnaX ribosomal frameshifting: programmed efficiency of 50%. *J Mol Biol* *271*, 47-60.
- Lee, T.H., Lapidus, L.J., Zhao, W., Travers, K.J., Herschlag, D., and Chu, S. (2007). Measuring the folding transition time of single RNA molecules. *Biophys J* *92*, 3275-3283.
- Lemay, J.F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Masse, E., and Lafontaine, D.A. (2011). Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS genetics* *7*, e1001278.
- Lemay, J.F., Penedo, J.C., Tremblay, R., Lilley, D.M., and Lafontaine, D.A. (2006). Folding of the adenine riboswitch. *Chem Biol* *13*, 857-868.
- Levinthal, C. (1969). How to Fold Graciously. Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois, 22-24.
- Li, M.S., Hu, C.K., Klimov, D.K., and Thirumalai, D. (2006). Multiple stepwise refolding of immunoglobulin domain I27 upon force quench depends on initial conditions. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 93-98.
- Li, P.T.X., Bustamante, C., and Tinoco Jr, I. (2007). Real-time control of the energy landscape by force directs the folding of RNA molecules. *Proc Natl Acad Sci USA* *104*, 7039-7044.
- Lin, J.-C., and Thirumalai, D. (2008). Relative stability of helices determines the folding landscape of adenine riboswitch aptamers. *J Am Chem Soc* *130*, 14080-14081.

- Lindorff-Larsen, K., Piana, S., Dror, R.O., and Shaw, D.E. (2011). How fast-folding proteins fold. *Science* *334*, 517-520.
- Liphardt, J., Dumont, S., Smith, S.B., Tinoco, I., Jr., and Bustamante, C. (2002). Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* *296*, 1832-1835.
- Liphardt, J., Naphthine, 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.
- Liphardt, J., Onoa, B., Smith, S.B., Tinoco, I., Jr., and Bustamante, C. (2001). Reversible unfolding of single RNA molecules by mechanical force. *Science* *292*, 733-737.
- Lunse, C.E., Schuller, A., and Mayer, G. (2014). The promise of riboswitches as potential antibacterial drug targets. *International journal of medical microbiology : IJMM* *304*, 79-92.
- Makarov, D.E. (2013). Interplay of non-Markov and internal friction effects in the barrier crossing kinetics of biopolymers: insights from an analytically solvable model. *The Journal of chemical physics* *138*, 014102.
- Makarov, D.E. (2014). Communication: Does force spectroscopy of biomolecules probe their intrinsic dynamic properties? *The Journal of chemical physics* *141*, 241103.
- Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., and Breaker, R.R. (2003). Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* *113*, 577-586.
- Mandal, M., and Breaker, R.R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* *11*, 29-35.
- Manktelow, E., Shigemoto, K., and Brierley, I. (2005). Characterization of the frameshift signal of Edr, a mammalian example of programmed -1 ribosomal frameshifting. *Nucleic Acids Res* *33*, 1553-1563.
- Manosas, M., and Ritort, F. (2005). Thermodynamic and kinetic aspects of RNA pulling experiments. *Biophys J* *88*, 3224-3242.
- Manuel, A.P., Lambert, J., and Woodside, M.T. (2015). Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories. *Proceedings of the National Academy of Sciences of the United States of America* *112*, 7183-7188.
- Marko, J., and Siggia, E.D. (1995). Stretching DNA. *Macromolecules* *28*, 8759-8770.

- Matzner, D., and Mayer, G. (2015). (Dis)similar Analogues of Riboswitch Metabolites as Antibacterial Lead Compounds. *Journal of medicinal chemistry* *58*, 3275-3286.
- Mazauric, M.H., Leroy, J.L., Visscher, K., Yoshizawa, S., and Fourmy, D. (2009). Footprinting analysis of BWYV pseudoknot-ribosome complexes. *RNA* *15*, 1775-1786.
- McDonald, J.F. (2012). *Transposable Elements and Genome Evolution* (Springer Netherlands).
- McQuarrie, D.A. (2000). *Statistical Mechanics* (University Science Books).
- 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. *J Mol Biol* *310*, 1109-1123.
- Milanesi, L., Waltho, J.P., Hunter, C.A., Shaw, D.J., Beddard, G.S., Reid, G.D., Dev, S., and Volk, M. (2012). Measurement of energy landscape roughness of folded and unfolded proteins. *P Natl Acad Sci USA* *109*, 19563-19568.
- Montange, R.K., and Batey, R.T. (2006). Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* *441*, 1172-1175.
- Mooney, R.A., Artsimovitch, I., and Landick, R. (1998). Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J Bacteriol* *180*, 3265-3275.
- Morrison, G., Hyeon, C., Hinczewski, M., and Thirumalai, D. (2011). Compaction and Tensile Forces Determine the Accuracy of Folding Landscape Parameters from Single Molecule Pulling Experiments. *Physical Review Letters* *106*, 138102.
- Nam, G.M., and Makarov, D.E. (2015). Extracting intrinsic dynamic parameters of biomolecular folding from single-molecule force spectroscopy experiments. *Protein science : a publication of the Protein Society*.
- 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.
- 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.
- Neidle, S. (1999). *Oxford Handbook of Nucleic Acid Structure* (Oxford University Press).

- Neuert, G., Albrecht, C., Pamir, E., and Gaub, H.E. (2006). Dynamic force spectroscopy of the digoxigenin-antibody complex. *FEBS Lett* *580*, 505-509.
- Neuman, K.C., and Block, S.M. (2004). Optical trapping. *Rev Sci Instrum* *75*, 2787-2809.
- Neupane, K., Manuel, A.P., Lambert, J., and Woodside, M.T. (2015). Transition-Path Probability as a Test of Reaction-Coordinate Quality Reveals DNA Hairpin Folding Is a One-Dimensional Diffusive Process. *The Journal of Physical Chemistry Letters* *6*, 1005-1010.
- Neupane, K., Ritchie, D.B., Yu, H., Foster, D.A.N., Wang, F., and Woodside, M.T. (2012). Transition Path Times for Nucleic Acid Folding Determined from Energy-Landscape Analysis of Single-Molecule Trajectories. *Physical Review Letters* *109*, 068102.
- Neupane, K., Yu, H., Foster, D.A.N., Wang, F., and Woodside, M.T. (2011). Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res* *39*, 7677-7687.
- Nixon, P.L., Rangan, A., Kim, Y.G., Rich, A., Hoffman, D.W., Hennig, M., and Giedroc, D.P. (2002). Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J Mol Biol* *322*, 621-633.
- Noeske, J., Buck, J., Fürtig, B., Nasiri, H.R., Schwalbe, H., and Wöhnert, J. (2007). Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res* *35*, 572-583.
- Noeske, J., Richter, C., Grundl, M.A., Nasiri, H.R., Schwalbe, H., and Wöhnert, J. (2005). An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. *Proc Natl Acad Sci USA* *102*, 1372-1377.
- Oliveberg, M., and Wolynes, P.G. (2005). The experimental survey of protein-folding energy landscapes. *Quat Rev Biophys* *38*, 245-288.
- Oliveira, R.J., Whitford, P.C., Chahine, J., Leite, V.B., and Wang, J. (2010a). Coordinate and time-dependent diffusion dynamics in protein folding. *Methods* *52*, 91-98.
- Oliveira, R.J., Whitford, P.C., Chahine, J., Wang, J., Onuchic, J.N., and Leite, V.B. (2010b). The origin of nonmonotonic complex behavior and the effects of nonnative interactions on the diffusive properties of protein folding. *Biophys J* *99*, 600-608.

- Olsthoorn, R.C., Reumerman, R., Hilbers, C.W., Pleij, C.W., and Heus, H.A. (2010). Functional analysis of the SRV-1 RNA frameshifting pseudoknot. *Nucleic Acids Res* *38*, 7665-7672.
- Onuchic, J.N., and Wolynes, P.G. (2004). Theory of protein folding. *Curr Opin Struc Biol* *14*, 70-75.
- Ottink, O.M., Rampersad, S.M., Tessari, M., Zaman, G.J.R., Heus, H.A., and Wijmenga, S.S. (2007). Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined-induced fit mechanism *RNA* *13*, 2202-2212.
- Pallan, P.S., Marshall, W.S., Harp, J., Jewett, F.C., 3rd, Wawrzak, Z., Brown, B.A., 2nd, 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.
- Park, S.J., Kim, Y.G., and Park, H.J. (2011). Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J Am Chem Soc* *133*, 10094-10100.
- Pennell, S., Manktelow, E., Flatt, A., Kelly, G., Smerdon, S.J., and Brierley, I. (2008). The stimulatory RNA of the Visna-Maedi retrovirus ribosomal frameshifting signal is an unusual pseudoknot with an interstem element. *RNA* *14*, 1366-1377.
- Perkins, T.T. (2014). Angstrom-precision optical traps and applications. *Annu Rev Biophys* *43*, 279-302.
- Pierse, C.A., and Dudko, O.K. (2013). Kinetics and energetics of biomolecular folding and binding. *Biophys J* *105*, L19-22.
- Pincet, F., and Husson, J. (2005). The solution to the streptavidin-biotin paradox: the influence of history on the strength of single molecular bonds. *Biophys J* *89*, 4374-4381.
- Pincus, D.L., Cho, S.S., Hyeon, C., and Thirumalai, D. (2008). Minimal Models for Protein and RNA: From Folding to Function. In *Progress in Molecular Biology and Translational Science* M. Conn, ed. (Academic Press).
- Plant, E.P., and Dinman, J.D. (2005). Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic Acids Res* *33*, 1825-1833.
- Plant, E.P., Jacobs, K.L., Harger, J.W., Meskauskas, A., Jacobs, J.L., Baxter, J.L., Petrov, A.N., and Dinman, J.D. (2003). The 9-Å solution: How mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting.

- Plant, E.P., Perez-Alvarado, G.C., Jacobs, J.L., Mukhopadhyay, B., Hennig, M., and Dinman, J.D. (2005). A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol* *3*, e172.
- Plotkin, S.S., and Wolynes, P.G. (1998). Non-Markovian Configurational Diffusion and Reaction Coordinates for Protein Folding. *Physical Review Letters* *80*, 5015-5018.
- Ponty, Y., and Leclerc, F. (2015). Drawing and editing the secondary structure(s) of RNA. *Methods in molecular biology (Clifton, NJ)* *1269*, 63-100.
- Popa, I., Fernández, J.M., and Garcia-Manyes, S. (2011). Direct Quantification of the Attempt Frequency Determining the Mechanical Unfolding of Ubiquitin Protein. *Journal of Biological Chemistry* *286*, 31072-31079.
- Pörschke, D. (1974). A direct measurement of the unzipping rate of a nucleic acid double helix. *Biophys Chem* *2*, 97-101.
- Qu, X., Wen, J.D., Lancaster, L., Noller, H.F., Bustamante, C., and Tinoco, I., Jr. (2011). The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* *475*, 118-121.
- Reining, A., Nozinovic, S., Schlepckow, K., Buhr, F., Furtig, B., and Schwalbe, H. (2013). Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* *499*, 355-359.
- Ren, J., Rastegari, B., Condon, A., and Hoos, H.H. (2005). HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *Rna* *11*, 1494-1504.
- Rieder, R., Lang, K., Graber, D., and Micura, R. (2007). Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *ChemBioChem* *8*, 896-902.
- Rieder, U., Kreutz, C., and Micura, R. (2010). Folding of a transcriptionally acting PreQ1 riboswitch. *Proc Natl Acad Sci USA* *107*, 10804-10809.
- Ritchie, D.B., Soong, J., Sikkema, W.K., and Woodside, M.T. (2014). Anti-frameshifting ligand reduces the conformational plasticity of the SARS virus pseudoknot. *J Am Chem Soc* *136*, 2196-2199.
- Ritchie, D.B., and Woodside, M.T. (2015). Probing the structural dynamics of proteins and nucleic acids with optical tweezers. *Curr Opin Struct Biol* *34*, 43-51.

- Roth, A., and Breaker, R.R. (2009a). The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* 78, 305-334.
- Roth, A., and Breaker, R.R. (2009b). The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* 78, 305-334.
- Russell, R. (2008). RNA misfolding and the action of chaperones. *Front Biosci* 13, 1-20.
- Russell, R., Zhuang, X., Babcock, H.P., Millett, I.S., Doniach, S., Chu, S., and Herschlag, D. (2002). Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci USA* 99, 155-160.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*, 4 edn (New York: Springer-Verlag).
- Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324, 163-166.
- Savinov, A., Perez, C.F., and Block, S.M. (2014). Single-molecule studies of riboswitch folding. *Biochim Biophys Acta* 1839, 1030-1045.
- Schellenberg, M.J., Wu, T., Ritchie, D.B., Fica, S., Staley, J.P., Atta, K.A., LaPointe, P., and MacMillan, A.M. (2013). A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nature structural & molecular biology* 20, 728-734.
- Schuler, B., and Hofmann, H. (2013). Single-molecule spectroscopy of protein folding dynamics--expanding scope and timescales. *Curr Opin Struct Biol* 23, 36-47.
- Schuler, B., Lipman, E.A., and Eaton, W.A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419, 743-747.
- Seol, Y., and Perkins, T. (2009). Sensitivity Of Dna-hairpins Dynamics To The Mechanism Of Force Feedback Probed Using A Surface-coupled Passive Force Clamp. *Biophysical Journal* 96, 291.
- Seol, Y., Skinner, G.M., and Visscher, K. (2004). Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys Rev Lett* 93, 118102.
- Seol, Y., Skinner, G.M., Visscher, K., Buhot, A., and Halperin, A. (2007). Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys Rev Lett* 98, 158103.

- Serganov, A., Yuan, Y.R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T., Hobartner, C., Micura, R., Breaker, R.R., and Patel, D.J. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* *11*, 1729-1741.
- Shaw, D.E., Deneroff, M.M., Dror, R.O., Kuskin, J.S., Larson, R.H., Salmon, J.K., Young, C., Batson, B., Bowers, K.J., Chao, J.C., *et al.* (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Commun ACM* *51*, 91-97.
- Shen, L.X., and Tinoco, I., Jr. (1995). The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J Mol Biol* *247*, 963-978.
- 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* *271*, 795-799.
- Solem, A.C., Halvorsen, M., Ramos, S.B., and Laederach, A. (2015). The potential of the riboSNitch in personalized medicine. *Wiley interdisciplinary reviews RNA*.
- Soong, J. (2014). Investigating the Conformational Dynamics of Biomolecules using Optical Tweezers. In *Physics* (Edmonton, Alberta, Canada: University of Alberta).
- Stigler, J., and Rief, M. (2012). Calcium-dependent folding of single calmodulin molecules. *Proceedings of the National Academy of Sciences of the United States of America* *109*, 17814-17819.
- Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J.C.M., and Rief, M. (2011). The complex folding network of single calmodulin molecules. *Science* *334*, 512.
- Stoddard, C.D., Gilbert, S.D., and Batey, R.T. (2008). Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* *14*, 675-684.
- Streisinger, G., and Owen, J.E. (1985). Mechanisms of Spontaneous and Induced Frameshift Mutation in Bacteriophage T4. *Genetics* *109*, 633-659.
- Stuart, A.M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica* *19*, 451-559.
- 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. *Nat Struct Biol* *6*, 285-292.
- Sullan, R.M., Churnside, A.B., Nguyen, D.M., Bull, M.S., and Perkins, T.T. (2013). Atomic force microscopy with sub-picoNewton force stability for biological applications. *Methods* *60*, 131-141.

- Svoboda, K., and Block, S.M. (1994). Biological applications of optical forces. *Annual Review of Biophysics and Biomolecular Structure* 23, 247-285.
- ten Dam, E.B., Pleij, C.W., and Bosch, L. (1990). RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus genes* 4, 121-136.
- Thirumalai, D., and Hyeon, C. (2005). RNA and Protein Folding: Common Themes and Variations. *Biochemistry* 44, 4957-4970.
- Truex, K., Chung, H.S., Louis, J.M., and Eaton, W.A. (2015). Testing Landscape Theory for Biomolecular Processes with Single Molecule Fluorescence Spectroscopy. *Physical Review Letters* 115, 018101.
- van Oijen, A.M. (2008). Cutting the forest to see a single tree? *Nature chemical biology* 4, 440-443.
- Wang, J.X., and Breaker, R.R. (2008). Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 86, 157-168.
- Wang, M.D., Yin, H., Landick, R., Gelles, J., and Block, S.M. (1997). Stretching DNA with optical tweezers. *Biophysical Journal* 72, 1335-1346.
- Wang, X., and Nau, W.M. (2003). Kinetics of End-to-End Collision in Short Single-Stranded Nucleic Acids. *J Am Chem Soc* 126, 808-813.
- Wang, Y., Wills, N.M., Du, Z., Rangan, A., Atkins, J.F., Gesteland, R.F., and Hoffman, D.W. (2002). Comparative studies of frameshifting and nonframeshifting RNA pseudoknots: a mutational and NMR investigation of pseudoknots derived from the bacteriophage T2 gene 32 mRNA and the retroviral gag-pro frameshift site. *RNA* 8, 981-996.
- Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737-738.
- 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, 598-603.
- Wensley, B.G., Batey, S., Bone, F.A.C., Chan, Z.M., Tumelty, N.R., Steward, A., Kwa, L.G., Borgia, A., and Clarke, J. (2010). Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* 463, 685-688.

- White, K.H., Orzechowski, M., Fourmy, D., and Visscher, K. (2011). Mechanical unfolding of the beet western yellow virus -1 frameshift signal. *J Am Chem Soc* *133*, 9775-9782.
- Wickiser, J.K., Cheah, M.T., Breaker, R.B., and Crothers, D.M. (2005a). The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* *44*, 13404-13414.
- Wickiser, J.K., Winkler, W.C., Breaker, R.R., and Crothers, D.M. (2005b). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* *18*, 49-60.
- Wied, D., and Weißbach, R. (2012). Consistency of the kernel density estimator: a survey. *Stat Papers* *53*, 1-21.
- Wills, N.M., Moore, B., Hammer, A., Gesteland, R.F., and Atkins, J.F. (2006). A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. *The Journal of biological chemistry* *281*, 7082-7088.
- Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vornrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* *407*, 327-339.
- Winkler, W.C., and Breaker, R.R. (2005). Regulation of bacterial gene expression by riboswitches. *Annual review of microbiology* *59*, 487-517.
- Wolynes, P.G. (2014). Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie*.
- Woodside, M., and Valentine, M. (2009). Single-Molecule Manipulation Using Optical Traps. In *Handbook of Single-Molecule Biophysics*, P. Hinterdorfer, and A. Oijen, eds. (Springer US), pp. 341-370.
- Woodside, M.T., Anthony, P.C., Behnke-Parks, W.M., Larizadeh, K., Herschlag, D., and Block, S.M. (2006a). Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* *314*, 1001-1004.
- Woodside, M.T., Behnke-Parks, W.M., Larizadeh, K., Travers, K., Herschlag, D., and Block, S.M. (2006b). Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 6190-6195.

- Woodside, M.T., and Block, S.M. (2014). Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu Rev Biophys* *43*, 19-39.
- Woodside, M.T., Garcia-Garcia, C., and Block, S.M. (2008). Folding and unfolding single RNA molecules under tension. *Current opinion in chemical biology* *12*, 640-646.
- Woodside, M.T., Lambert, J., and Beach, K.S. (2014). Determining intrachain diffusion coefficients for biopolymer dynamics from single-molecule force spectroscopy measurements. *Biophys J* *107*, 1647-1653.
- Xu, W., Lai, Z., Oliveira, R.J., Leite, V.B.P., and Wang, J. (2012). Configuration-Dependent Diffusion Dynamics of Downhill and Two-State Protein Folding. *The Journal of Physical Chemistry B* *116*, 5152-5159.
- Yang, S., Onuchic, J.N., and Levine, H. (2006). Effective stochastic dynamics on a protein folding energy landscape. *The Journal of chemical physics* *125*, 054910.
- Yu, C.H., Noteborn, M.H., Pleij, C.W., and Olsthoorn, R.C. (2011). Stem-loop structures can effectively substitute for an RNA pseudoknot in -1 ribosomal frameshifting. *Nucleic Acids Res* *39*, 8952-8959.
- Yu, C.H., and Olsthoorn, R.C. (2015). Monitoring ribosomal frameshifting as a platform to screen anti-riboswitch drug candidates. *Methods in enzymology* *550*, 385-393.
- Yu, H. (2013). Single-molecule studies of prion protein folding and misfolding (Edmonton: University of Alberta).
- Yu, H., Gupta, A.N., Liu, X., Neupane, K., Brigley, A.M., Sosova, I., and Woodside, M.T. (2012a). Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *P Natl Acad Sci USA* *109*, 14452-14457.
- Yu, H., Liu, X., Neupane, K., Gupta, A.N., Brigley, A.M., Solanki, A., Iveta, S., and Woodside, M.T. (2012b). Direct observation of multiple misfolding pathways in a single prion protein molecule. *P Natl Acad Sci USA* *109*, 5283-5288.
- Yu, Z., Koirala, D., Cui, Y., Easterling, L.F., Zhao, Y., and Mao, H. (2012c). Click chemistry assisted single-molecule fingerprinting reveals a 3D biomolecular folding funnel. *J Am Chem Soc* *134*, 12338-12341.

- Yusupova, G.Z., Yusupov, M.M., Cate, J.H., and Noller, H.F. (2001). The path of messenger RNA through the ribosome. *Cell* *106*, 233-241.
- Zhang, J., Lau, M.W., and Ferré-D'Amaré, A.R. (2010). Ribozymes and Riboswitches: Modulation of RNA Function by Small Molecules. *Biochemistry* *49*, 9123-9131.
- Zhang, Q., Brujić, J., and Vanden-Eijnden, E. (2011). Reconstructing Free Energy Profiles from Nonequilibrium Relaxation Trajectories. *J Stat Phys* *144*, 344-366.
- Zhang, Y., and Dudko, O.K. (2013). A transformation for the mechanical fingerprints of complex biomolecular interactions. *Proceedings of the National Academy of Sciences* *110*, 16432-16437.
- Zhuang, X. (2005). Single-molecule RNA science. *Annu Rev Biophys Biomol Struct* *34*, 399-414.
- Žoldák, G., Stigler, J., Pelz, B., Li, H., and Rief, M. (2013). Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. *P Natl Acad Sci USA* *110*, 18156-18161.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* *31*, 3406-3415.
- Zwanzig, R. (1988). Diffusion in a Rough Potential. *P Natl Acad Sci USA* *85*, 2029-2030.
- Zwanzig, R., Szabo, A., and Bagchi, B. (1992). Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America* *89*, 20-22.

Appendices

Appendix A Force Jump technical implementation

On our instrument, a force jump in the linear regions of the traps has been implemented in two ways: by jumping the applied voltage to the EOD or by jumping the frequency signal to the AOD. Their implementation is hardware specific. For our AODs, they are controlled purely by a limited software layer so timing is approximate and limited to a resolution of 1ms at best. To optimise the execution, the abstraction in Labview is removed, and the bottom-level subVIs are placed in their own parallel loop. An alternative implementation to this would be to use an FPGA, or a frequency generator that can be controlled by voltages or hardware timings.

A tricky part about doing force jumps with the EODs is to read back the output voltage as an input, which we can do with the PXI-6251 acquisition card. Then the trap positions can be precisely included into the calculations for the bead displacements from the trap centres ($Dy1$, $Dy0$). There is a complication that affects the corresponding EOD on the OT: a voltage divider was put in place to step down the output voltage range from $\pm 10V$ to $\mp 1V$, and some high frequency noise (possible impedance mismatch or a ground loop) on the upper trap ($T1$) is removed by the use of a Sallen-Key low-pass filter (see Fig A.1). For our force jump we move our traps apart at a fast ramp, e.g. 1000 nm/ms or in a discrete step, to a constant trap separation where extension changes are accompanied by force changes. Both the extension and the force on the molecule are recorded in separate traces.

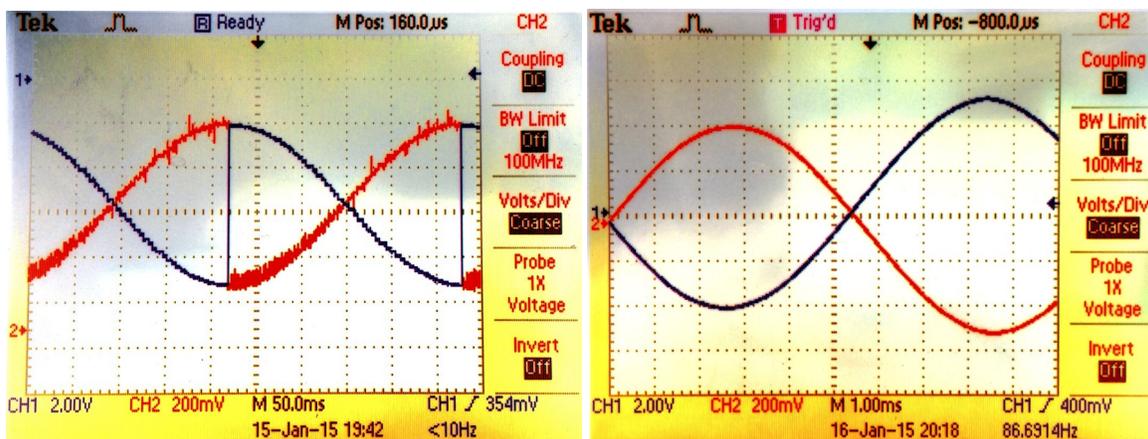


Fig A.1 EOD prefilter repair.

A: high frequency ringing on output trace (red) B: after repair with faster signal, a sine wave comes through the divider/inverter cleanly. (15-16th Jan 2015)

Our software driver for the AODs does not permit direct recording of the instance when the frequency signal to the AODs is actually changed. The workaround for this situation is to infer the trap's movement from the bead displacement from the trap centre (DyI). In the case of direct force change, by the AOD power, while measuring in the constant force region of the lower trap ($T0$) with a fixed upper trap position and the bead in the linear region, we can read back the power change in the lower trap by the bead movement in the upper trap. (At higher force, the upper bead is moved farther from the trap centre due to the tether). For our measurements, an approximately square wave was used to toggle the AOD power, and the calculation parameters were retained in the output files to determine the timing of the change in force.

When setting up the parameters for the jumps with the AOD, the high force and low force distances from the trap centre of the lower trap ($T0$) require some consideration. For our measurements, we were interested in measuring refolding behaviour at low force, so the refolding distance should be over the constant force region ($\sim 170 \pm 25\text{nm}$ from trap centre). (Halting the initialising FEC routine at about this distance from the low trap ($T0$) centre: $-170\text{nm} + \frac{\Delta F}{k_{\text{trap1}}} + \left(\frac{\Delta L_c}{2}\right)_{F_{1/2}}$). At the same time, depending on the force change desired, the upper trap power must be adjusted accordingly: if the difference in power between the two traps is too large, the lower bead will jump into upper trap, and the experiment will have to be stopped. (i.e. dropping the T1 stiffness from 0.52 to 0.28 pN/nm [2 to 1 mW preamplified AOD power])

Appendix B Extension extraction in force jump experiments

The displacements from the trap centres (Dy 's) and the extension are both recorded outputs. The force in the linear trap is calculated from the product of its Dy with the trap stiffness. Our post processing has been done in Igor Pro 6.3X. The displacement trace for the upper trap (DyI) will have some level of noise and drift. In order to extract the temporal locations of the applied force change, the DyI trace is median-smoothed with a large window (e.g. Fig B.1A), then one of two algorithms is used: a level crossing of the midpoint in the overall histogram of DyI , or the peaks from a scrolling variance window of the smoothed DyI trace (Fig B.1B). There is an underlying assumption, for consistency, that the DyI

trace will start in the high force level, and should end at the low force level. (The nature of the subVI execution in Labview is such that the timing varies, and the sample's endpoints may or may not have different levels of clipping).

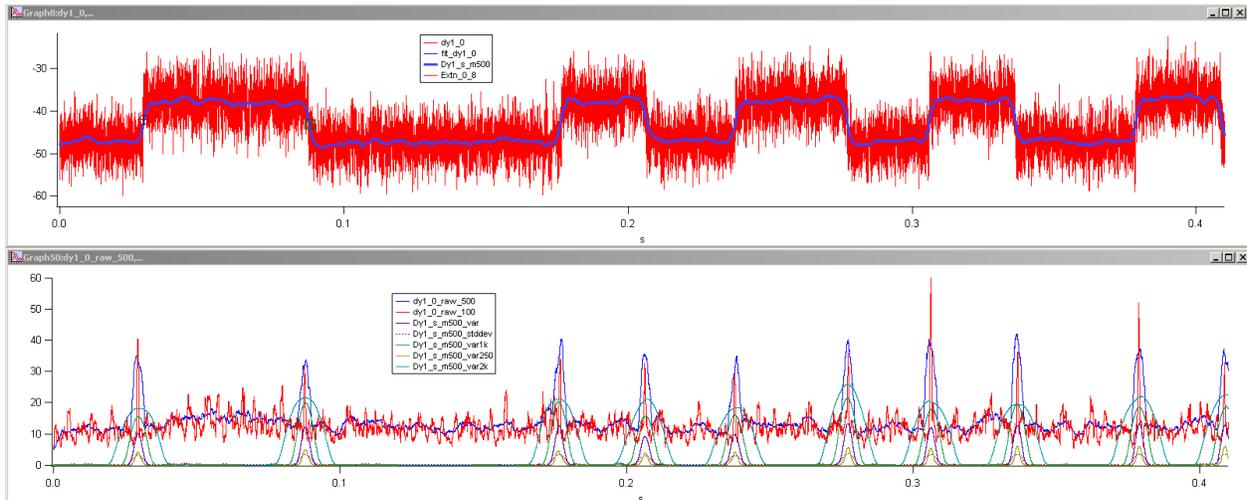


Fig B.1 Force-jump finding in Dy1 trace

A: Example force jump Dy1 trace, top, with the 500 point median smoothing laid in overtop. **B:** Bottom, corresponding variance window traces, over raw Dy1 signal in red and blue, and different widths of the variance over the smoothed trace. In both traces, the variability of execution time of a square wave can be seen.

The force stepping points in the *Dy1* are then refined by fitting a Heaviside function over the unfiltered *Dy1* trace, as the smoothing tends to shift the step forward. (As a control step, the *Dy1* trace sections can be compared to the extension sections, see Fig B.2) The time points are then applied to excise the low force extension traces from their precise beginning to their approximate end points. The endpoint of the refolding is then determined by taking a histogram for this excised extension, and starting from the beginning of the trace, finding the first arrival to the peak of that histogram. This makes it independent of long term drift in the extension trace. Further discussion of the data analysis can be found in chapter 8 with the motivation and theory.

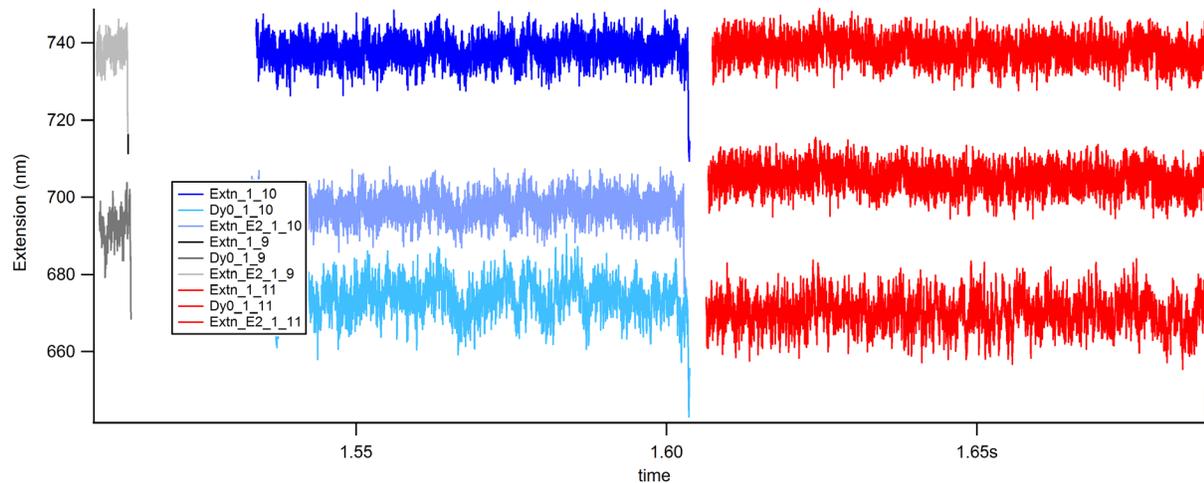


Fig B.2 Excision example: (from mol 0_3 nov 7)

Example traces of HP 30R50/T4 (sampled at 100kHz) of a 5pN force drop (0.9 to 0.7mW AOD power in $T1$) with corresponding $Dy0$ traces, showing the calculated extension trace change matches the bead displacement change in $T0$ while the bead in $T1$ is stationary (at constant force).

Appendix C High stiffness calibration

There are some cases in which it would be desirable to have stiffer traps, such as to probe brittle (stiff) molecules with tertiary structure, or to increase the time response of the force probe (see chapter 7). One way to do this is consolidate all the laser power in to a single trap, and use a micropipette (∞ stiffness) or functionalised slide. As an example, a stiffness of 1–2 pN/nm was achieved by Dong et al. in their work studying amyloid fibrils (Dong et al., 2010). AFM probes offer high stiffness, but these have presented inherent issues, such as inadequate sampling, hindering landscape reconstruction with weighted histogram analysis (Harris et al., 2007). Without deconstructing our OT, there are two approaches which we could use to increase the stiffness of our trap, which are relatively quick to implement. (Another would be the use of smaller beads and materials with a higher index of refraction than polystyrene.)

The first is to increase the amplitude of the frequency signal sent to the AOD, thereby increasing the strength of the diffraction grating formed by the device. The response from the AOD is non-linear, and after curving up to a plateau, the response will decrease as the AOD is over-driven, at which point the device will likely fail. (As we have experienced on another OT instrument in our lab, an electrical

contact to the AOD exploded when overdriven at too high a power.) For our primary AODs, they were tested out to 9mW unamplified signal beyond the usual 5mW, by measuring the power at the last mirror that gathers all the laser beams before the last telescope into the microscope objective, and at the position sensitive diode (PSD) at the condenser assembly exit (Fig C.1). A maximum value of 8mW was chosen as the operational ceiling for our AODs.

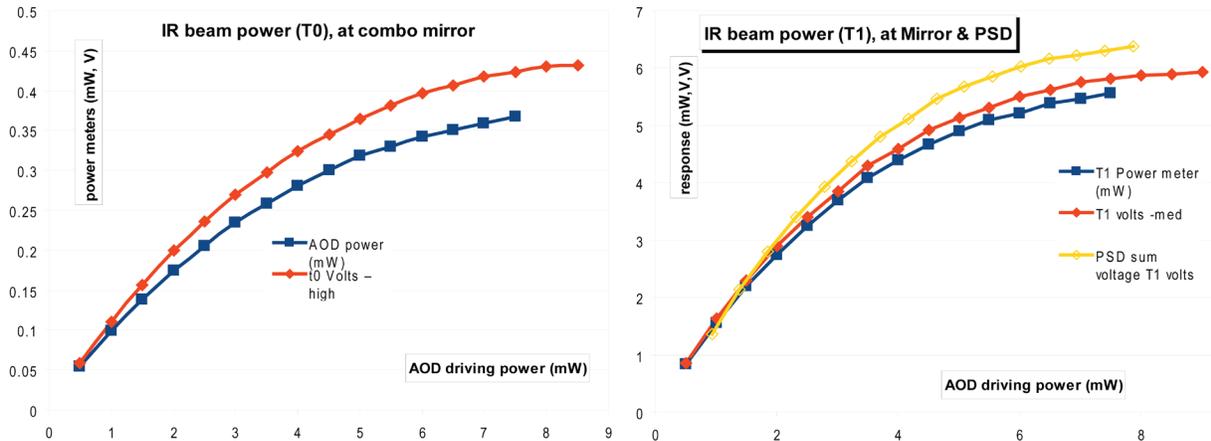


Fig C.1 AOD power throughput profiles of the OT instrument.

They show that the laser beam power throughput plateaus after 8mW. Measured with two different power meters in the optics box (blue, red), and with the PSD sum signal (yellow).

The second approach is to increase the current to the pumping diodes for the trapping laser. In our experiments we have always used 15A, but the laser control allows us to increase it up to 18A. However, when doing calibrations at this increased current it was found that stiffnesses actually decreased (especially in T1) compared to the values we measured at 15A. To investigate the difference, a beam profiler (Thorlabs) was placed in the optics box before the last telescope. As the diode current is increased, the laser beam itself broadens (Fig C.2). The effect is more problematic in our upper trap (T1). The beam power is lost because when designing the trap’s optics arrangement, the beam is set to overfill the window of the objective lens in order to maximise the focus and spread of the beam in the internal optics of the objective lens.

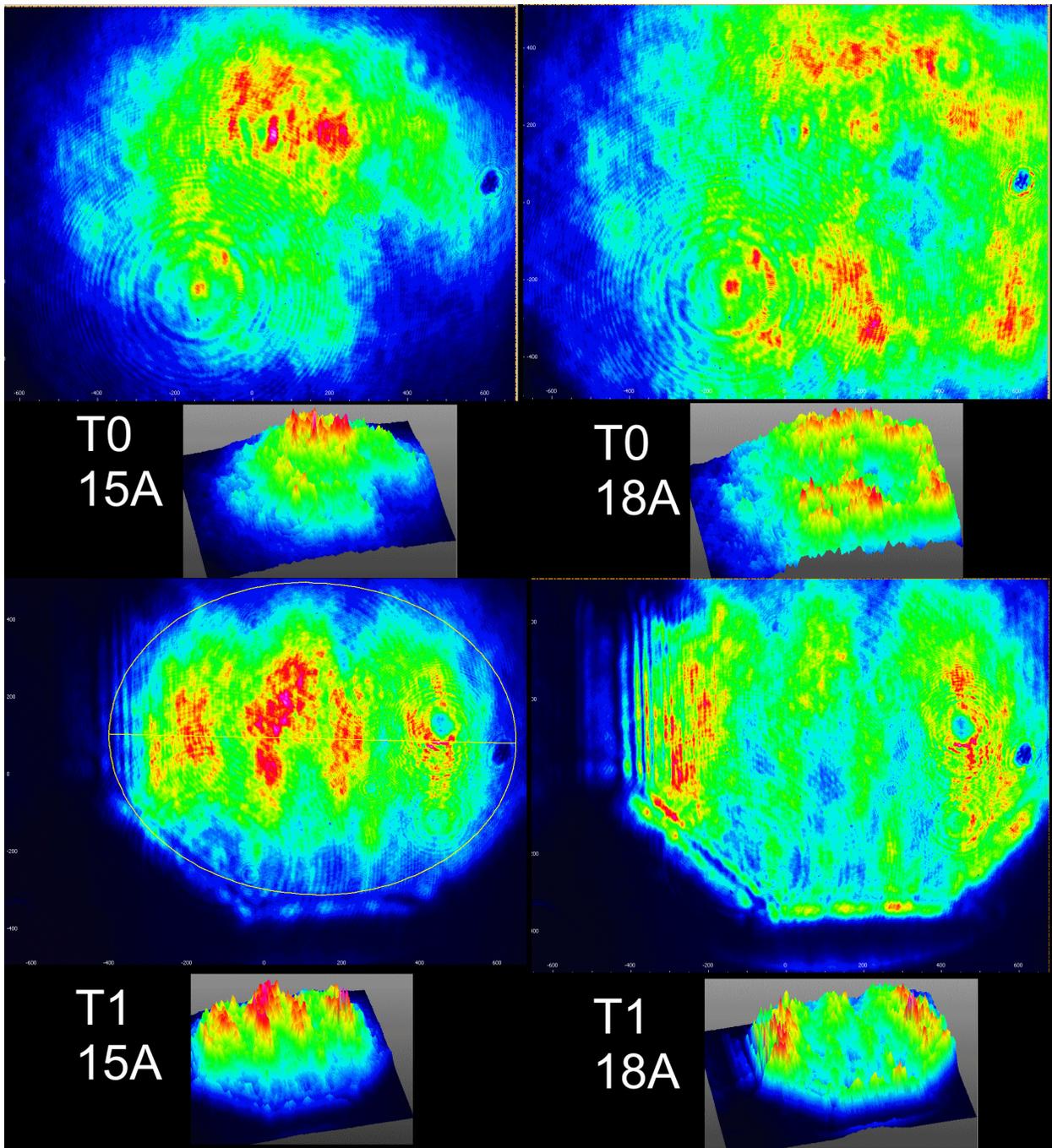


Fig C.2 Laser beam profiles measured before the last telescope before the objective lens.

Top: broadening of the beam with increased current to the pumping diodes for the lower trap (T0) (the change from 15 to 18A is qualitatively progressive.) Bottom: broadening of the beam for the upper trap (T1), meant to be stiffer than T0. The AOD deflection axis is horizontal in this image. The overfilling of the objective presumably cuts away much more power from T1 than it does from T0.

Measurements of the throughput power in 0.5A steps at the PSD (after the microscope, without laser warmup), suggested that a stiffness peak might be achievable in the range of ~ 16.3 to 16.5 A (Fig C.3). In spite of these efforts, and some gain of stiffness in the lower trap (T0), too much stiffness is lost in T1 to make the high power worthwhile: the geometry to the beam changes too much with increased pumping power. Averaging over all the measurements at higher AOD power, the best estimate for the trap stiffnesses at 8mW were: 0.626 pN/nm for T0, and 1.119 pN/nm for T1, a modest increase from those at 5mW: 0.529 pN/nm for T0 and 0.930 pN/nm for T1. (An increase of approximately 18 and 20%) These higher stiffnesses were used to measure FECs of hairpin 37 (30R50/4T) to test the inverse Wierstrass transform (Engel et al., 2014), to provide a higher temporal resolution for the transition times measurements (chapter 6) by performing measurements at constant trap separation.

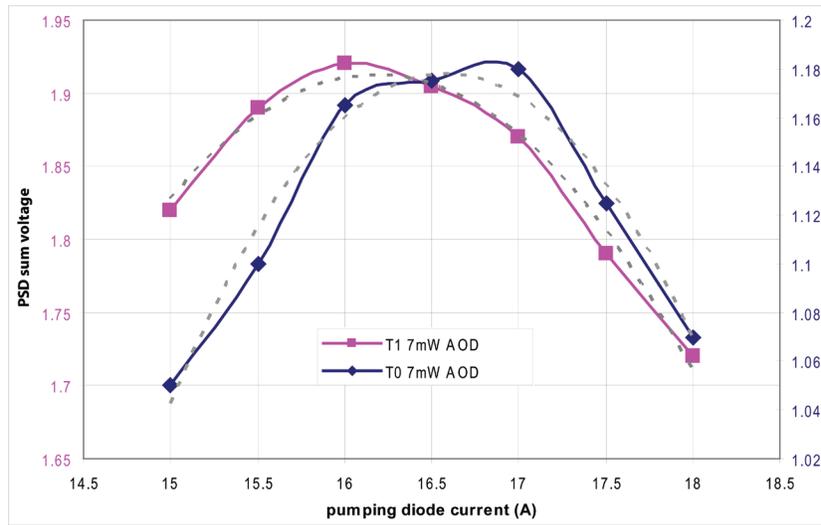


Fig C.3 Laser power measurements at the PSDs in the form of the sum voltage.

The pumping diode current midpoint offered an estimate of which point might provide concerted stiffness gains in both traps. Unfortunately, the beam geometry's contribution is a larger factor.

Appendix D Pseudoknot and riboswitch sequences

Appendix table D.1 Sequences of riboswitch RNA molecules measured *add* riboswitch aptamer, expression platform, full-length riboswitch, and oligomer sequences complementary to portions of the aptamer.

Aptamer sequence	5'- <i>CGCGGCTTCATATAATCCTAATGATATGGTTTGGGAGTTTCTAC</i> CAAGAGCCTTAAACTCTTGATTATGAAGTCT <i>TGT</i> -3' The single-stranded linkers to the duplex handles are listed in italics.
Expression platform	5'- <i>CCATTATGAAGTCTGTCGCTTTATCCGAAATTTATAAAGAGAA</i> GACTCATGAA <i>TCCC</i> -3'
Full-length sequence	5'- <i>CGCGGCTTCATATAATCCTAATGATATGGTTTGGGAGTTTCTAC</i> CAAGAGCCTTAAACTCTTGATTATGAAGTCTGTCGCTTTATCCGAA ATTTTATAAAGAGAAGACTCATAA <i>ATC</i> -3'
Oligomers complementary to aptamer	
complementary to junction J2/3 between P2 and P3 (46-53)	5'-GTAGAAAC-3'
complementary to 5' strand of P1 (13-22)	5'-ATATGAAGCC-3'
complementary to 5' strand and loop of P2 (25-38)	5'-CCATATTATTAGGA-3'
complementary to loop and 3' strand P3 (59-72)	5'-CAAGAATTTAAGGC-3'

Appendix table D.2 Pseudoknot structural properties and sequences

Pseudoknot	Total length (nt)	Stem 1 (bp)	Loop 1 (nt)	Stem 2 (bp)	Loop 2 (nt)	Unpaired nt between stems	PDB ID
PT2G32:	5'-UGACCAGCUAUGAGGUCAUACAUCGUCAUAGCAC-3'						2TPK
	32	5	1	7	7	0	
ScYLV C27A :	5'-AAGUGGCGCCGACCACUUAAAAACAACGGA -3'						2AP5
	28	5	2	3	9	1	
PEMV1:	5'-AAUUCGGUCGACUCCGGAGAAACAAAGUCA-3'						2RP1
	27	5	2	3	8	1	
ScYLV WT:	5'-AAGUGGCGCCGACCACUUAAAAACACCGGA-3'						1YG4
	28	5	2	3	9	1	
SRV1 ¹ :	5'-GCGGCCAGCUCCAGGCCGCAAACAAUAUGGAGCAC-3'						1E95
	34	6	1	6	9	0	
MMTV:	5'-GGGGCAGUCCCUAGCCCCACUCAAAAAGGGGGAU-3'						1RNK ²
	34	5	1	7	8	1	
HERV:	5'-GGGGCCAGCCUCAAGCCCCACAACAAACUGGGGCAU-3'						n.a.
	34	6	1	6	9	0	
VMV:	5'-AGGGGGCCACGUGUGGUGCCGUCCGCGCCCCUAU GUUGUAACAGAAGCACCACC -3'						n.a.
	54	7	5	7	14	7	
SARS*:	5'-GCGGUGUAAGUGCAGCCCGUCUUACACCGUGCG GCACAGGCACUAGUACUGAUGUCGUCUACAGGGCU-3'						n.a.
	68	10	3	7	9	2	

*Stem 3 shading

¹This sequence was optimized for structure analysis by NMR (Michiels et al., 2001).

²This PDB ID corresponds to an MMTV sequence modified to facilitate structural studies (Shen and Tinoco, 1995).

Appendix E Derivation the distribution of transit times in the limit of long times

Starting with equation 6.3 (Chaudhury and Makarov, 2010), we derive equation 6.4 for the distribution at long times:

$$P_{TP}(t) \approx \frac{\omega_k \sqrt{\beta \Delta G^\ddagger}}{1 - \operatorname{erf} \sqrt{\beta \Delta G^\ddagger}} \frac{\exp[-\beta \Delta G^\ddagger \coth(\omega_k t/2)]}{\sinh(\omega_k t/2) \sqrt{2\pi \sinh(\omega_k t)}}$$

The first term to approximate is

$$\frac{\omega_k \sqrt{\beta \Delta G^\ddagger}}{1 - \operatorname{erf} \sqrt{\beta \Delta G^\ddagger}}.$$

We cannot approximate with $\beta \Delta G^\ddagger$ as $\gg 1$ or $\ll 1$, since ΔG^\ddagger is similar to $k_B T$. However, there is an asymptotic series valid for large $\beta \Delta G^\ddagger$:

$$\operatorname{erf}(x) \approx 1 - \frac{e^{-x^2}}{\sqrt{\pi}} \left(\frac{1}{x} - \frac{1}{2x^3} + \frac{3}{4x^5} - + \dots \right)$$

So if we restrict ourselves to the first order of the expansion, we can now write the first term as:

$$\frac{\omega_k \sqrt{\beta \Delta G^\ddagger}}{1 - \operatorname{erf} \sqrt{\beta \Delta G^\ddagger}} \approx \omega_k \sqrt{\beta \Delta G^\ddagger} \frac{1}{1 - \left(1 - \frac{e^{-\beta \Delta G^\ddagger}}{\sqrt{\pi \beta \Delta G^\ddagger}} \right)} = \omega_k \beta \Delta G^\ddagger \sqrt{\pi} \exp \beta \Delta G^\ddagger$$

The second term to approximate is the one containing the hyperbolic trigonometric functions:

$$\frac{\exp[-\beta \Delta G^\ddagger \coth(\omega_k t/2)]}{\sinh(\omega_k t/2) \sqrt{2\pi \sinh(\omega_k t)}}$$

Using the identities:

$$\sinh y = \frac{1 - e^{-2y}}{2e^{-y}}, \quad \cosh y = \frac{1 + e^{-2y}}{2e^{-y}}, \quad \coth y \equiv \frac{\cosh y}{\sinh y}$$

in the limit of large y , we obtain

$$P_{\text{TP}}(t) \approx \omega_k \beta \Delta G^\ddagger \sqrt{\pi} \exp \beta \Delta G^\ddagger \frac{2 \exp[-\beta \Delta G^\ddagger] \exp - \omega_k t}{\sqrt{\pi}}$$

Combining the two terms, they simplify to equation 6.4, valid for long t and large $\beta \Delta G^\ddagger$ (Chaudhury and Makarov, 2010).

Appendix F Derivation of free energy as a function of non-equilibrium relaxation profiles

The following summarises the derivation of the free energy as a function of the non-equilibrium extension relaxation profiles published by Zhang and colleagues (Zhang et al., 2011).

The time dependent extension, $x(t)$, is assumed to be a one-dimensional variable, whose evolution can be described by an over-damped Langevin equation, to model the diffusion over the free energy, $G(x)$.

$$\dot{x} = -\beta DG'(x) + \sqrt{2D}\eta(t)$$

Here the \dot{x} represents dx/dt , and the prime denotes the first derivative with respect to x , while the double-prime denotes the second derivative. The equation includes the thermal β , and the diffusion coefficient is assumed to be position independent. To account for thermal fluctuations, $\eta(t)$ is a white noise term, with zero-mean and delta related covariance: $\langle \eta(t)\eta(t') \rangle = \delta(t-t')$.

The non-equilibrium probability density is defined empirically as:

$$\rho_N = \frac{1}{T_N} \sum_{j=0}^N \int_0^{t_j} \delta(x(t) - x) dt$$

Where N is the total number of observed trajectories, and T_N is the total time. As the number of trajectories goes to infinity, the sum can be replaced by an integral, and some stochastic calculus (Durrett, R. (1996). *Stochastic Calculus: A Practical Introduction*, CRC Press) will produce this Fokker-Planck equation:

$$\begin{cases} \frac{d}{dx}(\beta DG'(x)\rho(x)) + D\rho''(x) = -D\rho'(x_{folded})\delta(x - x_u), (x > x_{folded}) \\ \rho(x_{folded}) = 0 \end{cases}$$

With a constant diffusion, D divides out from both sides of the equation, and then we integrate both sides with respect to x , evaluating the integral from x_f to x . The integral of the first term:

$$\beta \int_{x_f}^x \left(\frac{d}{dx} (G'(x)\rho(x)) \right) dx = G'(x)\rho(x) - G'(x_f)\rho(x_f)$$

Here, $G'(x_f)$, the slope of the free energy, goes to zero at the folded state because we should have a potential well at that location, so we are left with only the first term. Integrating $\rho''(x)$ is straightforward, and so we are left with the integral of the Dirac-delta function on the RHS:

$$-\rho'(x_f) \int_{x_f}^x \delta(x - x_u) dx = -\rho'(x_f) H(x - x_u)$$

We only get a single term with these limits of integration, as the integral of the delta function is a Heaviside step function, such that $H(x_f - x_u)$ is zero. Grouping all the integration results, we now have:

$$\beta G'(x)\rho(x) + \rho'(x) - \rho'(x_f) = -\rho'(x_f)H(x - x_u),$$

and rearranging:

$$\beta G'(x)\rho(x) + \rho'(x) = \rho'(x_f)(1 - H(x - x_u))$$

Now we can make the deduction that $\beta G'(x)$ would be produced by taking the derivative of $\exp(\beta G(x))$, so if we multiply the equation on both sides by the exponential of the unitless energy, we can write:

$$\frac{d}{dx} \left(e^{\beta G(x)} \rho(x) \right) = \rho'(x_f)(1 - H(x - x_u)) e^{\beta G(x)}$$

Where the LHS has been grouped into the derivative, collapsing the two terms that arise from the product rule for differentiation. This equation can be integrated to produce the non-equilibrium $\rho(x)$, as shown in their paper (Zhang et al., 2011). To get the free energy, instead of multiplying by the exponential term, we instead divide by $\beta\rho(x)$:

$$G'(x) + \frac{\rho'(x)}{\beta\rho(x)} = \frac{\rho'(x_f)(1 - H(x - x_u))}{\beta\rho(x)}$$

We can make the observation that: $\frac{\rho'(x)}{\rho(x)} = \frac{d}{dx} \ln(\rho(x))$, so that if we integrate the equation, with the limits x_f to x , since we have an absorbing boundary at x_s and $\rho(x)$ is normalised between x_s and ∞ , we get:

$$\int_{x_f}^x G'(x) dx + \int_{x_f}^x \frac{\rho'(x)}{\beta\rho(x)} dx = \frac{\rho'(x_f)}{\beta} \int_{x_f}^x \frac{(1 - H(x - x_u))}{\rho(x)} dx$$

Now the LHS becomes:

$$G(x) - G(x_f) + \frac{\ln(\rho(x))}{\beta} - \frac{\ln(\rho(x_f))}{\beta}$$

$G(x)$, the energy of the folded state is our zero level in our energy landscape.

Before further integrating the RHS, we note that in this case: $\int_{x_f}^{x_u} F(x) dx = \int_{x_f}^x F(x) dx + \int_x^{x_u} F(x) dx$, for a point x between the range of x_f and x_u (folded and unfolded). So we can rewrite the RHS as:

$$\frac{\rho'(x_f)}{\beta} \left\{ \int_{x_f}^{x_u} \frac{(1-H(x-x_u))}{\rho(x)} dx - \int_x^{x_u} \frac{(1-H(x-x_u))}{\rho(x)} dx \right\}$$

Uniting all the terms:

$$G(x) = -\frac{\ln(\rho(x))}{\beta} + \frac{\ln(\rho(x_f))}{\beta} + \frac{\rho'(x_f)}{\beta} \left\{ \int_{x_f}^{x_u} \frac{(1-H(x-x_u))}{\rho(x)} dx - \int_x^{x_u} \frac{(1-H(x-x_u))}{\rho(x)} dx \right\}$$

Similarly to the argument for dropping $G(x_f)$, the middle terms on the LHS are constants that do not vary with x , so the function form for the free energy can be reduced to:

$$G(x) = -\frac{\ln \rho(x)}{\beta} - \frac{\rho'(x_f)}{\beta} \int_x^{x_u} \frac{(1-H(x-x_u))}{\rho(x)} dx$$

The step function is zero when less than x_u , and is +1 after x_u , so for lengths greater than the unfolded length x_u , we have the Boltzmann inversion for the free energy that we have seen previously.