

**University of Alberta**

High resolution optical tweezers for single molecule studies of hierarchical folding in the *pbuE* riboswitch aptamer

by

Daniel Foster

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

Master of Science

Department of Physics

© Daniel André Norman Foster  
Spring 2010  
Edmonton, Alberta

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

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

## **Examining Committee**

Michael T. Woodside, Physics

Mark Freeman, Physics

Jack Tuszynski, Physics and Oncology

Marek Duszyk, Physiology

## **Dedication**

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

## Abstract

Riboswitches are gene regulatory elements found in messenger RNA that function by changing structure upon the binding of a ligand to an aptamer domain. Single adenine-binding pbuE riboswitch aptamer RNAs were unfolded and refolded co-transcriptionally using optical tweezers for single molecule force spectroscopy. The kinetic and energetic properties of distinct folding intermediates were characterised with and without the binding of adenine. These observed intermediates were related to structural elements of the aptamer, which were found to fold sequentially, in a transcriptionally independent manner. The mechanical switch underlying the regulatory action of the riboswitch was observed directly (adenine stabilisation of the weakest helix), and the energy landscape for the folding was reconstructed.

The construction of a dual-beam optical trap with separate detection and trapping laser beams manipulated and focused into a rigid, modified inverted microscope is also described. This instrument aims to achieve ångström-level resolution through careful design to reduce noise.

## **Acknowledgements**

The publication included in this thesis was done in collaboration with members of Steven Block's lab at Stanford University. I would like to thank our collaborators, who made the riboswitch study possible, in particular William Greenleaf (now at Harvard University), Kirsten Frieda, and Professor Block. They kindly allowed us to join in collaboration with them to work on the riboswitch pulling assay at Stanford when we were unable to use the lab at NINT due to construction delays. Their efforts were central, essential, and greatly appreciated.

My supervisor/advisor, Prof. Michael Woodside, who has always been open and approachable, and has always kept a sense of humour about things. After all, scientific writing is not an easy task, and Michael helped to remove the 'Cubist-Picasso' and 'Jackson Pollock' aspects of the document.

I would like to thank Cuauhtémoc Garcia-Garcia (Dr. Garcia<sup>2</sup>), who was a post-doc in our group for a short time, for his insightful review of early documents.

Much of our work in the lab depends upon parts produced by the talented machinists in the physics department: Gilbert Lachat, Paul Zimmerman, Travis Peters, Boris Tomasevic, and Tony Paget, who have helped us design and improve many parts, including our optics enclosure and our microscope reconstruction. Tony Walford, who runs the student machine shop in CCIS, has also made many parts that we use with our optics.

## Table of Contents

Dedication.....	
Abstract.....	
Acknowledgements.....	
Table of Contents.....	
List of Tables.....	
List of Figures and Illustrations.....	
List of Symbols, Abbreviations and Nomenclature.....	
CHAPTER 1: INTRODUCTION.....	1
1.1 The role of RNA in the cell.....	1
1.2 RNA structure and folding.....	5
1.3 Studying riboswitch folding, one molecule at a time.....	12
1.4 Thesis outline.....	17
1.5 References.....	17
CHAPTER 2: OPTICAL TWEEZERS.....	20
2.1 Theory of optical trapping.....	21
2.1.1 Optical trapping in the Rayleigh limit.....	23
2.1.2 Optical trapping in the Mie limit.....	24
2.2 Construction of the optical trap.....	27
2.3 Layout.....	29
2.3.1 Trapping beam.....	29
2.3.2 Layout - Detection beam.....	35
2.3.3 Beam steering considerations.....	36
2.3.3.1 Acousto-optic deflectors (AODs).....	36
2.3.3.2 Electro-optic deflectors (EODs).....	38
2.3.4 Position Detection.....	39
2.3.5 Microscope.....	41
2.4 Stiffness calibration of the optical trap.....	44
2.4.1 Power spectrum calibration.....	44
2.4.2 Variance method.....	49
2.4.3 Viscous Drag Calibration.....	50
2.5 Summary.....	51
2.6 References.....	52
CHAPTER 3: RIBOSWITCHES: A NEW GENE REGULATORY ELEMENT.....	54
3.1 Types of riboswitches.....	57
3.2 The purine riboswitches: structure and function.....	60
3.3 The pbuE adenine riboswitch.....	68
3.4 References.....	72
CHAPTER 4: OPTICAL TRAPPING MEASUREMENTS.....	74
4.1 Dynamic force measurements.....	74
4.1.1 Elastic properties of nucleic acids.....	77

4.1.2 Analysis of FECs.....	79
4.2 Constant Force Measurements.....	83
4.2.1 Analysis of constant force measurements .....	86
4.3 References.....	89
CHAPTER 5: DIRECT OBSERVATION OF HIERARCHICAL FOLDING IN SINGLE RIBOSWITCH APTAMERS .....	91
5.1 Abstract.....	91
5.2 Single molecule study of hierarchical folding in the pbuE riboswitch.....	92
5.3 Figures .....	100
5.4 References and Notes.....	104
5.5 Supporting Online Material .....	106
5.5.1 Materials and Methods .....	106
5.5.1.1 Constructs. ....	106
5.5.1.2 Optical trap. ....	107
5.5.2 Force-extension curves.....	108
5.5.3 Force distributions.....	111
5.5.4 Equilibrium free energy.....	112
5.5.5 Kinetics from FECs.....	113
5.5.6 Hairpin unfolding predictions.....	115
5.5.7 Refolding trajectories. ....	116
5.5.8 Free energy landscapes.....	120
5.5.9 Co-transcriptional folding.....	121
5.5.10 Primary Sources of Experimental Error. ....	121
5.5.11 References .....	126
CHAPTER 6: FUTURE WORK .....	127
6.1 Single-molecule force spectroscopy of riboswitches.....	127
6.2 References.....	133

## List of Tables

Table 5.1 (S1). Measured values describing individual folding steps of <i>pbuE</i> aptamer.....	125
Table 5.2 (S2). Model calculation results for folding of the P2 and P3 hairpins.....	125

## List of Figures and Illustrations

Figure 1.1 Illustration of the flow of genetic information in a cell.....	2
Figure 1.2 Aptamer of the PreQ1 riboswitch.....	3
Figure 1.3 The chemical structure of DNA..	6
Figure 1.4 The chemical difference between DNA and RNA..	7
Figure 1.5 The B-form helix in DNA..	8
Figure 1.6 Sample RNA structures.....	10
Figure 1.7 RNA cleavage.....	10
Figure 1.8 Riboswitch action.....	13
Figure 1.9 smFS methods..	16
Figure 2.1 Rayleigh and Mie limit forces.....	22
Figure 2.2 Geometric optics ray diagram..	26
Figure 2.3 Photograph of optical layout..	28
Figure 2.4 Diagram of the layout of the infrared trapping beam..	29
Figure 2.5 Stability of the trapping laser. ....	31
Figure 2.6 Optics enclosure .....	32
Figure 2.7 Conjugate optical planes in microscopy.....	33
Figure 2.8 Detector laser layout diagram.....	35
Figure 2.9 Linearity of optical deflectors response..	38
Figure 2.10 The redesigned condenser assembly..	42
Figure 2.11 Video capture of 1.0 $\mu\text{m}$ beads.....	43
Figure 2.12 Force profile in an optical trap..	45
Figure 2.13 Trap power spectrum.....	48
Figure 3.1 Metabolism and genome of <i>B. subtilis</i> ..	57
Figure 3.2 Riboswitch classification by structure.....	59

Figure 3.3 Secondary structure of the <i>pbuE</i> riboswitch.....	60
Figure 3.4 Other purine riboswitch aptamers. ....	61
Figure 3.5 The crystal structure of the <i>xpt</i> riboswitch aptamer. ....	64
Figure 3.6 Purine structure.....	65
Figure 3.7 Structure and function of the <i>xpt</i> and <i>add</i> riboswitches.. ....	67
Figure 3.8 The crystal structure of the <i>add</i> riboswitch aptamer. ....	69
Figure 4.1 Force extension curve assay. ....	76
Figure 4.2 FEC analysis.....	80
Figure 4.3 Scaling analysis of unfolding rates.....	84
Figure 4.4 Constant force measurements.....	85
Figure 5.1 Assay design.....	100
Figure 5.2 Unfolding force distribution in FECs.....	101
Figure 5.3 Kinetics of aptamer refolding and binding.....	102
Figure 5.4 Aptamer states and energetics determined by refolding at constant force. ....	103
Figure 5.5 FECs observed when adenine is not bound to the aptamer. ....	123
Figure 5.6 Graphical representation of the work performed during stretching..	123
Figure 5.7 Refolding traces in presence and absence of anti-sense oligos. ....	124
Figure 6.1 The structure and action of the <i>pbuE</i> and <i>addA</i> -riboswitches. ....	129
Figure 6.2 Secondary and crystal structures of SAM riboswitch aptamers.....	131

## List of Symbols, Abbreviations and Nomenclature

<b>Symbol</b>	<b>Definition</b>
RNA	ribonucleic acid
DNA	deoxyribonucleic acid
mRNA	messenger RNA
miRNA	micro RNA
sRNA	small RNA
CRISPR	clustered regularly interspaced short palindromic repeats
UTR	untranslated region
ds NA	double stranded nucleic acid
ss NA	single stranded nucleic acid
A	adenine (base) adenosine (nucleoside)
G	guanine (base) guanosine (nucleoside)
C	cytosine (base) cytidine (nucleoside)
T	thymine (base) thymidine (nucleotide)
U	uracil (base) uridine (nucleoside)
ORF	open reading frame
RBS	ribosomal binding site
UUU	uridine rich tract
RNAP	RNA polymerase
tRNA	transfer RNA
FS	force spectroscopy
smFS	single molecule force spectroscopy
NMR	nuclear magnetic resonance
FRET	Förster resonance energy transfer
smFRET	single molecule FRET
RBS	ribosomal binding site
NTP	nucleotide-triphosphate
SAM	S-Adenosyl methionine
TPP	thiamine pyrophosphate
PBS (pol cube)	polarising beam splitter
$\lambda/2$ waveplate	half-wave plate
EOD	electro-optic deflector
AOD	acousto-optic deflector
PSD	duolateral position sensitive diode
NA	numerical aperture
DIC	differential interference contrast (microscopy)

## Chapter 1: Introduction

### 1.1 The role of RNA in the cell

RNA (ribonucleic acid) plays a central role in biology because it is involved in many different cellular processes. The principal role of RNA is often considered to be as an intermediate in the flow of genetic information, as illustrated by the so-called “central dogma of biology” (Crick 1970, Fig1.1), which states that genetic information flows from DNA (deoxyribonucleic acid) to RNA to protein. Genetic information is stored in the form of DNA, but in order to express the gene, first the DNA must be transcribed into RNA, and then the RNA is translated into protein by the ribosome (a molecular machine for protein biosynthesis consisting of both proteins and RNA).

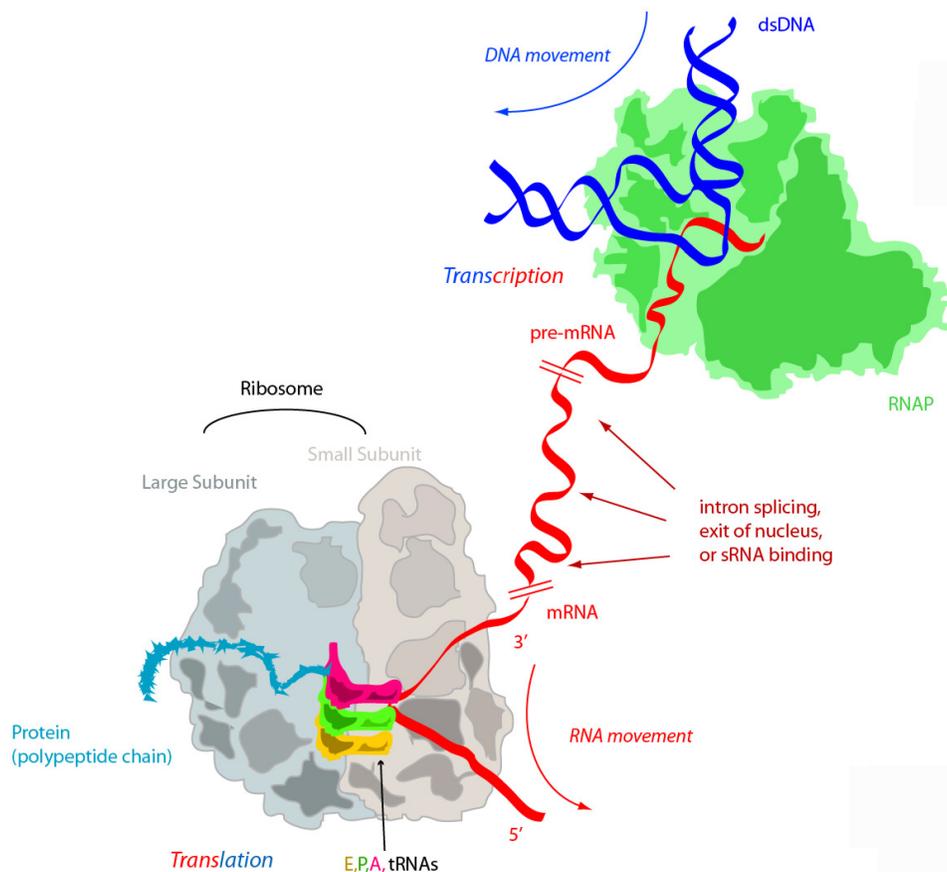
However, the role of RNA goes well beyond this simple picture. RNA can serve the same role of genetic storage as DNA, as seen in retroviruses such as HIV (Moore 2009), and it can also act as an enzymatic catalyst, typically one of the principal roles of proteins. These catalytic RNAs are known as ribozymes; examples of which include self-splicing introns<sup>1</sup> (Cech 1990), RNase P, (Altman 1990), and the ribosome itself. In eukaryotes, gene regulation is often accomplished by regulatory proteins binding to DNA and DNA methylation<sup>2</sup> (Jaenisch 2003). However, it has been discovered that RNA can also take on gene

---

<sup>1</sup> Eukaryotic genes are made up of two parts: exons and introns. During pre-mRNA processing in the eukaryotic cell nucleus, introns are removed, and the final product, mRNA, contains only the exons. Self-splicing introns utilise the self-catalytic capacity of RNA to cut themselves out of pre-mRNA without involving an external enzyme.

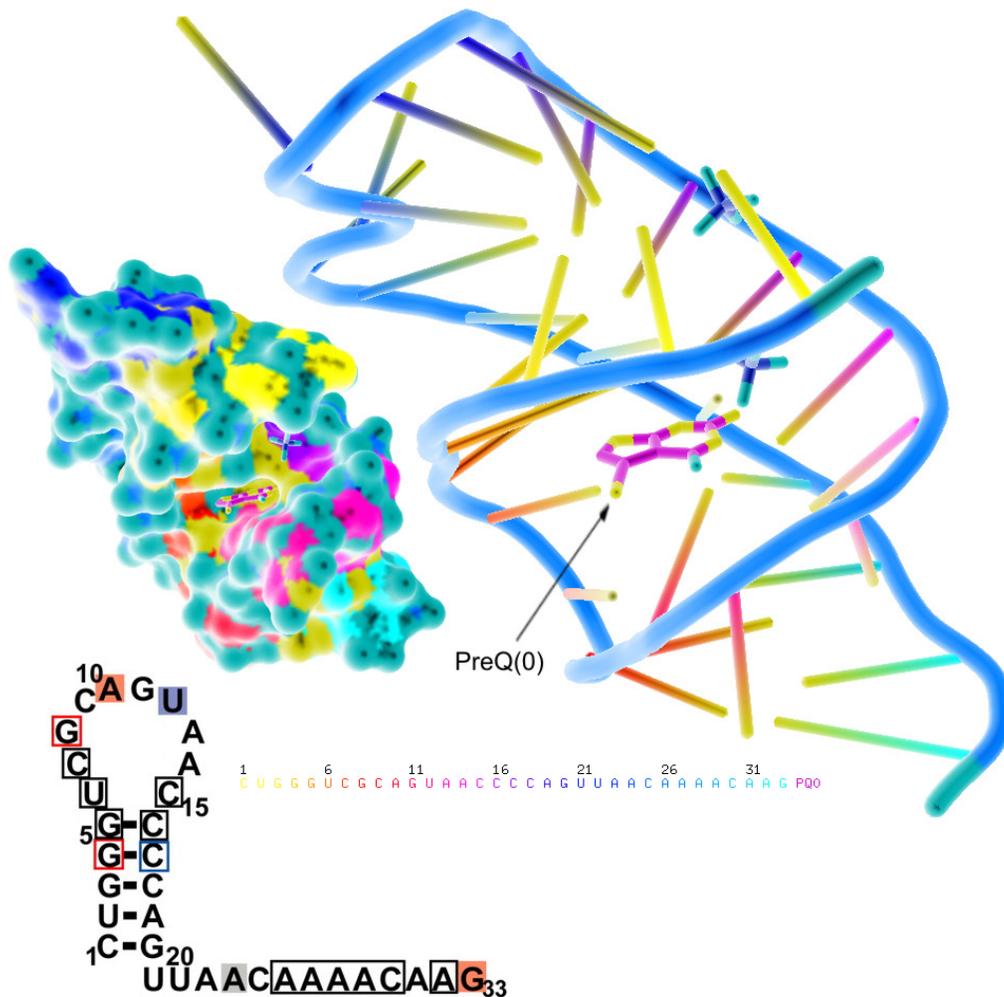
<sup>2</sup> DNA methylation involves the addition of a methyl group to the 5' carbon of the cytosines in the promoter region of a gene by a methyltransferase enzyme (Berg 2006).

regulatory roles, such as through RNA interference (Fire 1998, 2007), which has revolutionised genetic studies. Most interestingly, from the point of view of this thesis, RNA is capable of highly specific ligand binding, another protein-like role (Fig1.2). Such binding was observed originally in laboratory assays to develop so-called “aptamers” (Ellington 1990, Stoltenburg 2007), but recently a class of naturally-occurring ligand-binding RNAs was discovered in gene-regulatory RNAs known as riboswitches (Montange 2008, Lemay 2006).



**Figure 1.1** Illustration of the flow of genetic information in a cell. Information contained in the DNA is read by RNA polymerase (RNAP) and transcribed into RNA. This RNA may be further processed (e.g. splicing) before being translated into protein by the ribosome. The ribosome is a macromolecular assembly around the mRNA that translates the gene sequence into an amino acid chain, reading from the 5' to 3' direction. Gene regulation may take place at any step along the process.

All this variety has generated a huge and growing interest over the past few decades for a molecule that was previously underappreciated. RNA continues to surprise in the wide variety of roles it can take on, and it is now thought by many, according to the “RNA world” hypothesis (Federov 2004, Cech 2009), that earlier in life’s origins, RNA played the roles that DNA and protein fulfill today.



**Figure 1.2** Aptamer of the PreQ1 riboswitch. Two structural renderings (PDB ID 3GCA and 3FU2) are shown, including magnesium ions and the PreQ0 ligand (inset: secondary structure). The PreQ1 riboswitch is the smallest known to date (Spitale 2009).

If we consider RNA's roles in an evolutionary context, then this creates a more complex picture of its roles in the cell. Even if proteins have replaced RNA in many or most of its past roles, RNA's capabilities as an enzyme and a genetic regulator remain of great importance. From an evolutionary standpoint proteins provide a more specific and stable molecule for sensing and regulation than RNA can, and are thus favoured, but simply using RNA as a metabolic short-cut for equivalent regulation provides an argument for its continued use by cells to this day<sup>3</sup>.

As it turns out, RNA plays regulatory roles in different ways. Its splicing activity is important to pre-messenger-RNA (pre-mRNA) processing, permitting different regulation by cell type in metazoans, and subsequent malfunctions here are linked to human diseases (Cooper 2009). Other eukaryotic RNA regulators include micro-RNA (miRNA) (Schickel 2008) and short interfering RNA (siRNA) (Mello 2007), while in bacteria their roles are completed by small RNA regulators (sRNA), clustered regularly interspaced short palindromic<sup>4</sup> repeats (CRISPR) (Barrangou 2007) RNAs, and riboswitches (Waters 2009).

Riboswitches, the subject of this thesis, are RNA elements found in untranslated regions (UTRs) in either the 5' or 3' end of mRNA (Lemay 2006). Regulation by riboswitches typically provides the capacity to mediate cell responses to changing environmental conditions by acting on the stress response

---

<sup>3</sup> In terms of numbers, RNA regulators require the synthesis of ~100-200 bases while the average protein in *e. coli* requires 350 amino acids (Waters 2009)

<sup>4</sup> A double-stranded nucleic acid, such as DNA, is said to be palindromic if it is equal to its complementary sequence read backwards.

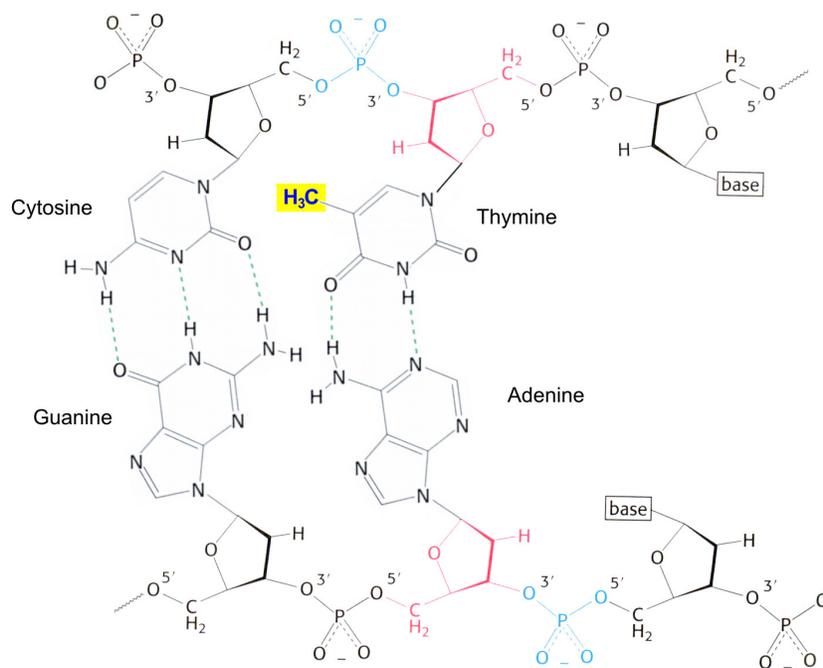
or the metabolism. The importance of riboswitch regulation is most notable in *Bacillus subtilis*, 2% of whose genes are regulated in this manner (Kim 2008). The role of riboswitches differs across species, however: whereas the *B. subtilis* *glmS* mRNA riboswitch provides negative regulation in *cis*, in *E. coli* the *glmS* mRNA is instead regulated by *trans* acting sRNAs.

Regulatory activity by riboswitches is made possible by the complex structures that RNAs can form. The folding up of this nucleic acid into a very specific geometry permits the riboswitch to form a dynamic sensor that is able both to discriminate a target ligand very sensitively, and to undergo a specific structural rearrangement upon ligand binding. It is this structural rearrangement which ends up regulating gene expression. Generally, other regulatory RNAs are not so reliant on the interplay of tertiary interactions. Riboswitches are thus an example of RNA folding that acts directly on gene regulation.

## **1.2 RNA structure and folding**

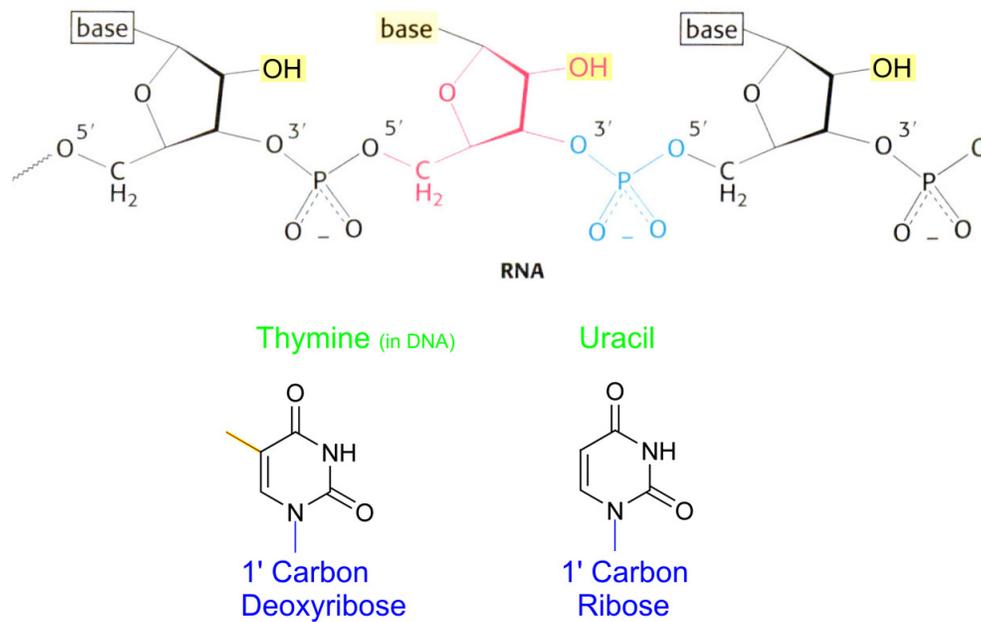
A hallmark of biological macromolecules is that their structure and function are related very tightly, and RNA is no exception. Chemically, RNA is a close cousin of the more familiar DNA (deoxyribonucleic acid). Both are polymers of nucleotides, which consist of a ribose or deoxyribose sugar, a pyrimidine or purine base, and a phosphate group, connected by phosphodiester

bonds<sup>5</sup> (Fig1.3, Fig1.4). Interactions between these nucleotides, such as hydrogen bonding and the stacking of aromatic rings between the bases result in three-dimensional structures. Generally speaking, for DNA this means a double helix in the familiar “B-form,” held together by Watson-Crick base pairing between the two strands in the helix and base stacking along each strand (Saenger 1984) (Fig 1.5). A wide variety of other forms of DNA are possible, as well as non-canonical base-pairing interactions (Neidle 1999)

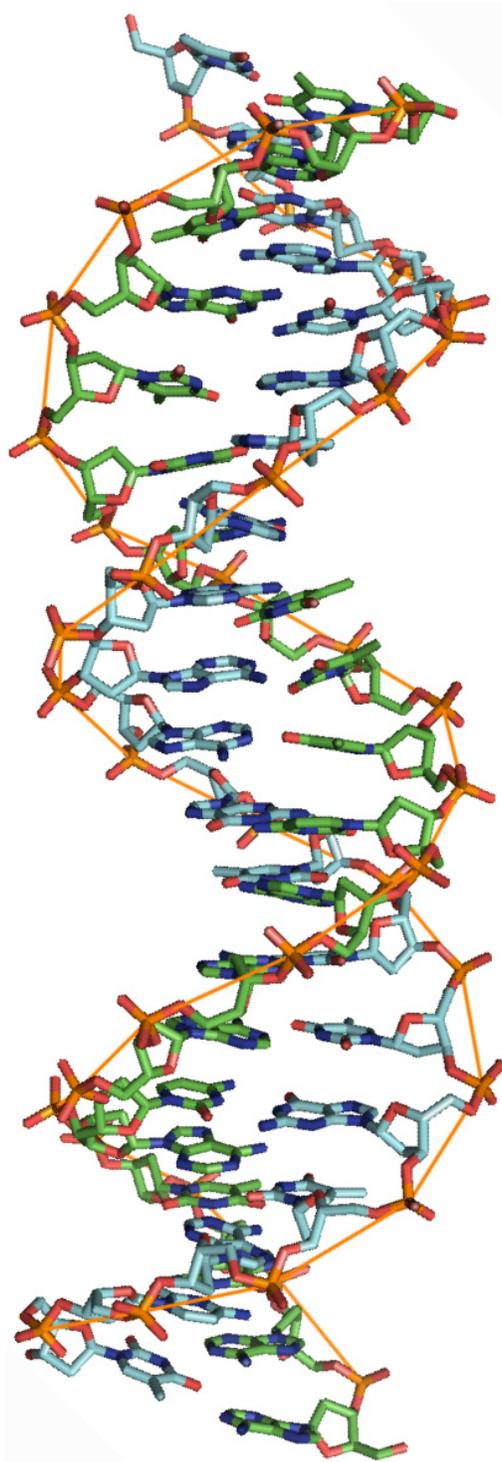


**Figure 1.3 The chemical structure of DNA. DNA uses 4 bases: adenine (A) pairs with thymine (T), guanine (G) pairs with cytosine (C). Thymine has an extra methyl group that differentiates it from uracil (U), found instead of T in RNA. The phosphate groups linking the deoxyribose sugars are charged. G:C pairs are more stable than A:T pairs due to additional hydrogen bonds between the bases (green dashed lines) (Figure adapted from Berg 2006).**

<sup>5</sup> The synthesis of the bases for DNA includes an extra step: nucleotide monophosphates (NMP's), used in RNA synthesis are dehydroxylated to 2'-deoxy-nucleotide monophosphates (dNMP's).



**Figure 1.4** The chemical difference between DNA and RNA. There are two main differences: RNA has a 2' hydroxyl group, and U instead of T. The 2' hydroxyl group (highlighted, absent in deoxyribose) permits cleavage of the backbone at the phosphate links between the bases. In the production of a DNA thymine base, a methyl group is added to a uracil base.



**Figure 1.5** The B-form helix in DNA. A rendering of DNA (PDB ID 2OR1) with a backbone ribbon as a guide. The phosphates are shown in red and orange, while the complementary bases are green:blue.

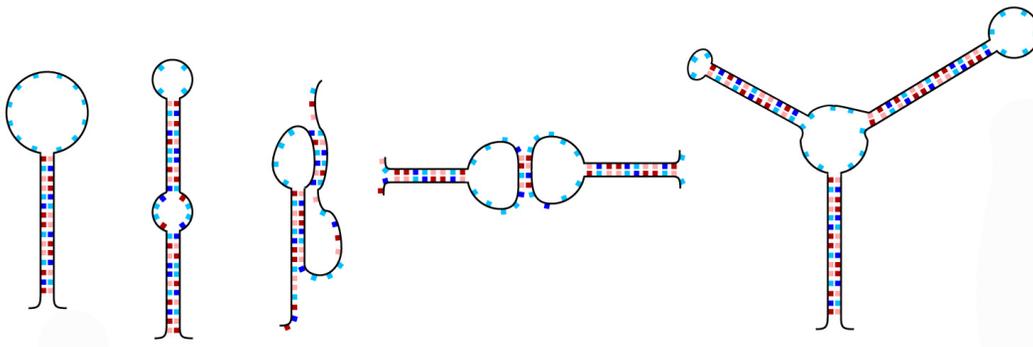
RNA structures are typically much more diverse than those of DNA. One reason for this is that RNA is generally synthesized as single-strands (ss), in contrast to DNA, which is usually replicated in semi-conserved double strands (ds). The nucleotides within a single strand can then interact with each other, forming structures other than simple helices. The extra hydroxyl group that is present in RNA also increases the capacity for hydrogen bonding<sup>6</sup>, further promoting more complex structures (as opposed to the case in DNA<sup>7</sup>). As a result, in RNA folding both secondary and tertiary interactions must be taken into account. Examples of RNA structures include ds A-form helices in siRNAs, loops in hairpins, multiple helix junctions, pseudoknots, and kissing loop complexes, among others<sup>8</sup> (Fig 1.6). The wide range of structures in RNA is linked to its wide variety of roles in the cell.

---

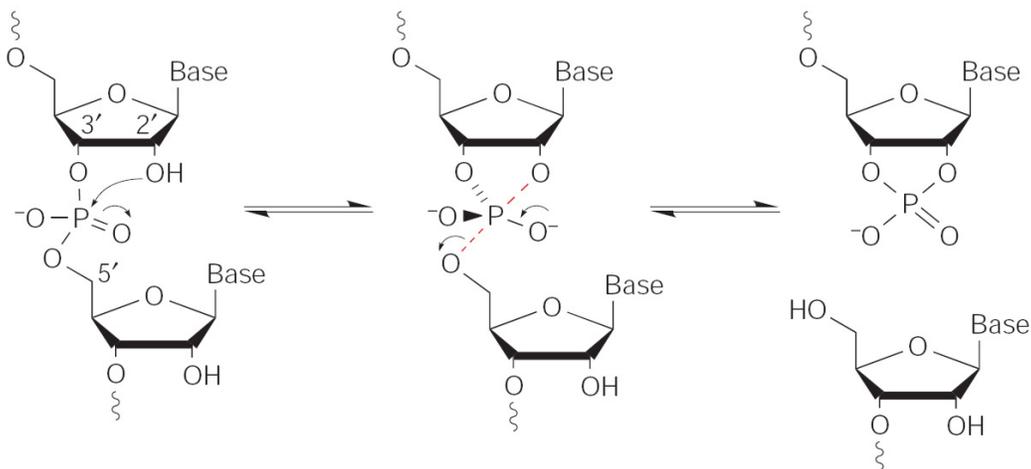
<sup>6</sup> The extra hydroxyl group in RNA gives the nucleic acid the capacity to cut itself apart: the phosphodiester bond is cleaved through a transesterification reaction (Soukup 1999). This capacity is used for enzymatic purposes in ribozymes, but limits the chemical stability of RNA (Fig 1.7). It is also exploited for in-line probing of RNA structure (Mandal 2004).

<sup>7</sup> There is a second important distinction between RNA and DNA (Fig 1.3, Fig 1.4). While they both utilize four nucleotide bases, they differ by a pyrimidine substitution: thymine (T) in a DNA molecule in the place of uracil (U) (with some exceptions, Vértessy 2009). Before the deoxynucleotide of uracil is incorporated into DNA, the pyrimidine is methylated (León 1998). This solves two problems caused by uracil. First, uracil is somewhat promiscuous in its base pairing, forming non-canonical base pairings with guanine and even itself. The methylation restricts thymine to pairing with adenine alone, and makes it more hydrophobic (helps in keeping the bases inside the B-helix). Second, the most common mutation in DNA is the spontaneous deamination of cytosine (C) into U. The exclusion of U in DNA permits a rapid repair of the error.

<sup>8</sup> In the case of ssRNAs, structure is important because of how it facilitates function (such as PDB#1Y27).



**Figure 1.6 Sample RNA structures. Basepaired helices and ss regions are both indicated schematically. From left to right: a hairpin, hairpin with a bulge, a pseudoknot, kissing loops, a multi-loop junction. The two dimensional view is a simplification, as the hairpin will actually be an A-form helix with a loop at the top, while a bulge or kink in the stem could make this helix bend or twist. The tertiary structures are also very sequence dependent (Hendrix 2005). Since secondary structure can be independently stable, the examples shown here can very well be more than mere transient states.**



**Figure 1.7 RNA cleavage. The direction by which the hydroxyl of the ribose sugar accomplishes the cleavage of an RNA strand (Mandal 2004).**

The process by which structure is formed is known as folding. RNA folding is similar to folding in proteins, but driven by different forces. Whereas in proteins folding is driven largely by hydrophobic interactions, in RNA electrostatics plays a much larger role, due to the charged phosphate group in each nucleotide (carrying 1 negative charge). Folding of RNA often requires the participation of divalent cations such as magnesium to help mediate the backbone charges. Base pairing and stacking interactions are very stable, so that secondary structure (which arises from interactions between the nucleotides) is generally more stable than tertiary structure (which comes from interactions between secondary structures). As a result, RNA folding has a hierarchy: first secondary structure forms, then tertiary (Brion 1997), in contrast with how proteins behave (Daggett 2003). In both cases, however, folding is able to avoid Levinthal's paradox (the observation that a random search through all configurations for a typical protein would take orders of magnitude longer than the age of the universe, rather than the typical  $\sim 10^{-3}$ - $10^1$  s range, Levinthal 1968), by engaging in a biased sampling of configuration space, and making use of cooperative intramolecular interactions to drive the molecule towards its native conformation (Zwanzig 1992, Gonzalez 2008).

The study of folding processes presents a substantial challenge, as it involves many different processes at many timescales. Proteins may completely fold on a  $\mu$ s timescale, or it may take them minutes or longer; nucleic acids typically fold over a time-range of  $10^{-4}$  –  $10^1$ s. Methods like X-ray

crystallography<sup>9</sup> and NMR provide a three-dimensional picture of the final, folded state, but yield only limited information about the dynamics of folding. Several other methods have been applied to study RNA folding, including small-angle X-ray scattering, fluorescence spectroscopy, footprinting, calorimetry, absorbance and circular dichroism spectroscopy, and a variety of computational approaches. Each of these has added to our picture of how RNA folds, but there are still many gaps to be filled by other methods, such as single molecule experiments, which offer new avenues of exploration.

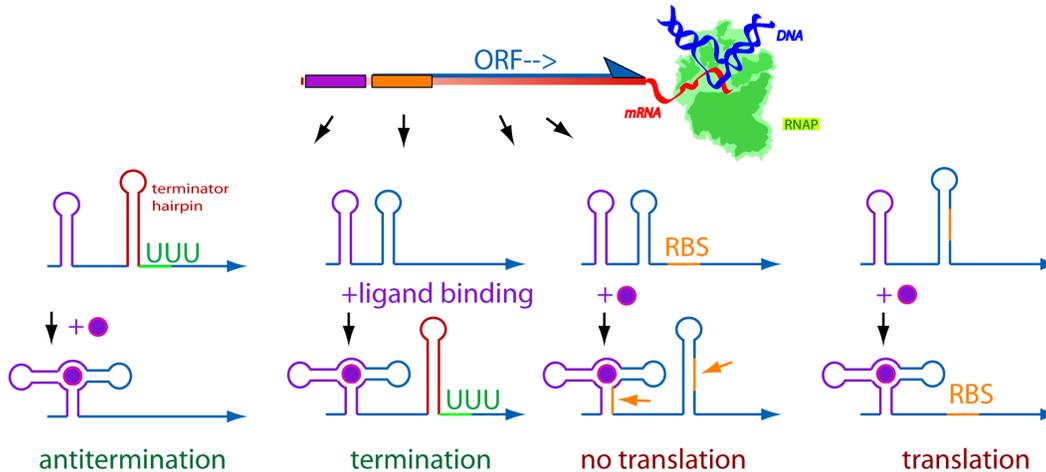
### **1.3 Studying riboswitch folding, one molecule at a time**

This thesis focuses on studying the folding of riboswitches at the level of single molecules, using optical tweezers to probe the folding dynamics. Riboswitches represent a particular opportunity for folding studies, since their folding is tied directly to their function as gene regulators. Most riboswitches are found in the 5' UTR. They are composed of two structural domains: an aptamer (ligand binding) domain and an expression platform. The aptamer typically contains or forms a pocket that is able to bind a target ligand very specifically. The ligand binding then changes the stability of the structure of the expression platform. This expression platform is the part of the riboswitch that effects the gene regulation: for example, it may be composed of a structure that terminates or anti-terminates transcription, or it may contain structure(s) that prevent translation initiation (Fig

---

<sup>9</sup> The self-cleaving capacity of RNA presents an additional obstacle to crystallisation, not to mention the widespread presence of nucleases (RNA digesting enzymes) that can easily contaminate samples!

1.8). Conformational change in the riboswitch, induced by an environmental signal, is therefore what triggers gene regulation.



**Figure 1.8 Riboswitch action.** In bacteria, riboswitches form as mRNA is synthesised by RNAP. Several patterns of regulation are possible, including modulation of transcription through formation of terminators or anti-terminators, and modulation of translation through sequestration of the ribosome binding site (RBS). The riboswitch aptamer need not necessarily be the structure that sequesters the RBS (shown by arrows); it could be in a separate hairpin. Enzymic regulation (RNA-splicing), another mode of regulation, is not shown here. ORF: open reading frame; UUU: uridine rich tract (the DNA-RNA hybrid in the polymerase is weakest when the RNA has a series of uridines, promoting transcription termination). Adapted from Waters (2009).

The riboswitch we have examined is the *pbuE* adenine riboswitch from *Bacillus subtilis*. Many riboswitches, including the *pbuE*, are relatively small molecules (containing ~100 nt), yet they display quite varied behaviour. They therefore offer a relatively simple and small system in which to explore fundamental questions about macromolecular folding. To simplify the system further and to obtain a clearer understanding of the folding, we have looked at the

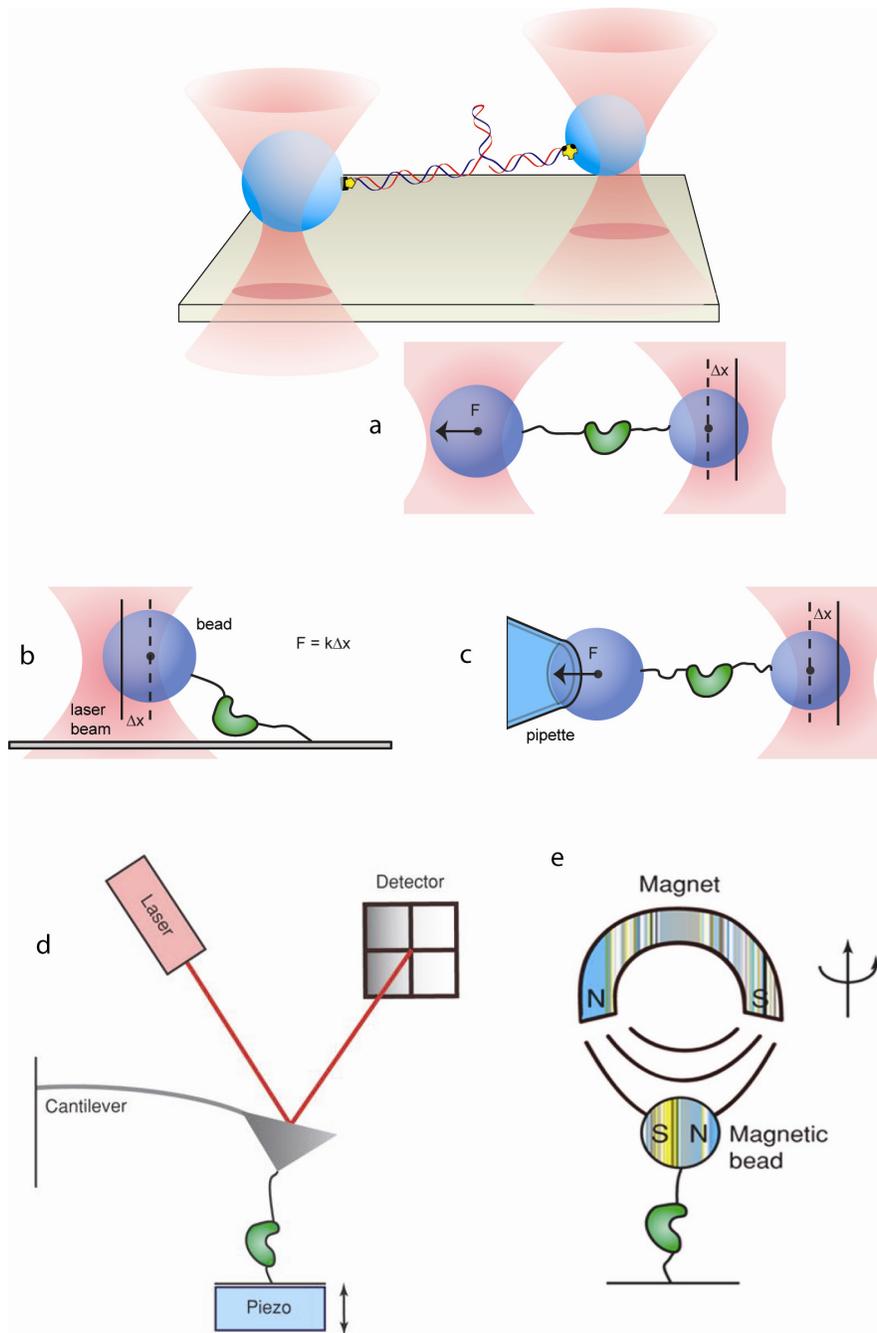
aptamer domain alone. The approach we use is single molecule force spectroscopy using optical tweezers.

The study of single molecules offers the possibility of seeing what bulk measurements cannot. States that occur rarely or transiently, as well as heterogeneous behaviour, can all be observed in single molecule experiments. This means that intermediates can be seen directly, and a single folding trajectory may be followed in its entirety. This opens up the possibility to characterise all these aspects of the folding that would be lost to the ensemble average in a bulk measurement (van Oijen 2008). The price to be paid for the extra information from single molecule measurements is a reduced throughput, because the single-molecule approach is a time-consuming serial method.

A large number of single molecule techniques have been developed. Techniques such as particle tracking, single-molecule fluorescence spectroscopy, or Förster resonance energy transfer (FRET) allow the behaviour of the molecule to be observed passively (Greenleaf 2007, Zhuang 2005). “Active” or force-based techniques, however, offer more direct control over the molecule in an experimental assay. In these approaches, a controlled force is applied to the molecule, and its response is monitored. Force-based methods include *in vitro* atomic force microscopy (AFM), viscous fluid flow, and magnetic and optical tweezers (Fig 1.9). The single-molecule force spectroscopy (smFS) provided by these active techniques offers unique capabilities for studying molecular folding. Force stretches out the unfolded structures, thereby converting folding into motion, which can be measured very precisely. Since force times distance yields

an energy, the applied force allows the energy landscape of the molecular folding to be probed and manipulated (Greenleaf 2007). By using force to unfold the molecule, abrupt temperature or chemical denaturant concentration changes (otherwise necessary to induce folding and unfolding) can be avoided. The strength (stiffness) of the force probe can also be changed to scale the force resolution to the level appropriate for the molecule being studied.

The study of RNA folding is one to which optical tweezers are particularly well suited, because they offer a gentler and more controlled means of manipulating an RNA molecule than AFMs, due to their lower probe stiffness and the simple chemistries that have been developed to control the geometry of the force application. In optical tweezers experiments, the RNA molecule is attached to “handles” (micron long polymer tethers such as dsDNA) at each end, which are in turn attached to functionalised, micron-sized plastic beads. These beads are trapped by the gradient force from the tweezers, allowing tension to be applied to the RNA through the handles. Here, we use force spectroscopy of single *pbuE* adenine riboswitch aptamer molecules to gain insight into the folding thermodynamics and kinetics, the accessible states, and the folding hierarchy.



**Figure 1.9 smFS methods. Optical tweezers: a dual trap setup in (a), a single trap for a surface assay in (b), and the setup with a pipette as the second clamp in (c). Example setups of AFM (d) and magnetic tweezers (e). Both optical and magnetic tweezers can be used to apply torsion as well as tension. Adapted from Greenleaf (2007) and Woodside (2009).**

## 1.4 Thesis outline

In the next chapter (2), I discuss optical trapping and design of the trap that I have helped build at the University of Alberta. The properties of riboswitches, with a particular focus on the *pbuE* riboswitch I have been studying, are presented in chapter 3. Chapter 4 gives an overview of how optical traps can be used to measure biomolecular folding, and what considerations are required in the data analysis. The results of our study of the riboswitch, done in collaboration with researchers at Stanford University, are presented in chapter 5 as a reprint of our paper published in the journal *Science*. In the closing chapter (6), I discuss the next steps to be taken following the results obtained in this work.

## 1.5 References

Altman S (1990), Nobel lecture. *Enzymatic cleavage of RNA by RNA*, Biosci Rep. **10**:317-37.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007), *CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes*, Science **315**:1709-12.

Berg JM, Tymoczko JL, Stryer L (2006), *Biochemistry*, W.H. Freeman 6th ed. ISBN-10: 0716787245.

Brion P, Westhof E (1997), *Hierarchy and dynamics of RNA folding*, Annu Rev Biophys Biomol Struct. **26**:113-37.

Cech TR (1990), Nobel lecture. *Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena*, Biosci Rep. **10**:239-61.

Cech TR (2009), *Crawling out of the RNA world*, Cell. **136**:599-602.

Cooper TA, Wan L, Dreyfuss G (2009), *RNA and Disease*, Cell, **136**:777-93.

- Crick F (1970), *Central dogma of molecular biology*, Nature **227**:561-3.
- Daggett V, Fersht AR (2003), *Is there a unifying mechanism for protein folding?*, Trends Biochem Sci. **28**:18-25.
- Ellington AD, Szostak JW (1990), *In vitro selection of RNA molecules that bind specific ligands*, Nature **346**:818-22.
- Fedorov A, Fedorova L (2004), *Introns: mighty elements from the RNA world*, J Mol Evol. **59**:718-21.
- Fire, A, Xu, S, Montgomery, MK, Kostas, SA, Driver, SE, and Mello, CC (1998), *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans*, Nature **391**:806-11.
- Fire AZ (2007), *Gene silencing by double-stranded RNA*, Cell Death Differ. **14**:1998-2012.
- Gonzalez RL Jr. (2008), *Navigating the RNA folding landscape*, Nat Chem Biol. **4**:451-2.
- Greenleaf WJ, Woodside MT, Block SM (2007), *High-Resolution Single-Molecule Measurements of Biomolecular Motion*, Annu. Rev. Biophys. **36**:171-90
- Hendrix DK, Brenner SE, Holbrook SR (2005), *RNA structural motifs: building blocks of a modular biomolecules*, Q Rev Biophys. **38**:221-43.
- Jaenisch R, Bird A (2003), *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals*, Nat Genet. **33** Suppl:245-54.
- Lemay JF, Lafontaine DA (2006), *Le riborégulateur adénine : Un nouveau mode de régulation génétique*, Med Sci (Paris). **22**:1053-9.
- León P (1998), *Inhibition of Ribozymes by Deoxyribonucleotides and the Origin of DNA*, J Mol Evol. **47**:122-6
- Levinthal C (1968), *Are there pathways for protein folding?*, J. Chim. Phys. Phys.- Chim. Biol. **65**: 44–5.
- Mandal M, Breaker RR (2004), *Gene regulation by riboswitches*, Nat Rev Mol Cell Biol. **5**:451-63.
- Mello CC (2007), *Return to the RNAi world: rethinking gene expression and evolution*, Cell Death Differ. **14**:2013-20.
- Montange RK, Batey RT (2008), *Riboswitches: emerging themes in RNA structure and function*, Annu Rev Biophys. **37**:117-33.

- Moore MD, Hu WS (2009), *HIV-1 RNA Dimerization: It Takes Two to Tango*, AIDS Rev. **11**:91-102.
- Neidle S (1999), *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press; 1st ed. ISBN-10: 0198500386
- Saenger W (1984), *Principles of Nucleic Acid Structure*, Springer-Verlag, ISBN-10: 0387907610
- Schickel R, Boyerinas B, Park SM, Peter ME (2008), *MicroRNAs: key players in the immune system, differentiation, tumorigenesis and cell death*, Oncogene **27**: 5959–74.
- Soukup GA, Breaker RR (1999), *Relationship between internucleotide linkage geometry and the stability of RNA*, RNA **5**:1308-25.
- Spitale RC, Torelli AT, Krucinska J, Bandarian V, Wedekind JE (2009), *The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain*, J. Biol. Chem. **284**: 11012-6
- Stoltenburg R, Reinemann C, Strehlitz B (2007), *SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands*, Biomol Eng. **24**:381-403.
- van Oijen AM (2008), *Cutting the forest to see a single tree*, Nat Chem Biol. **4**:440-443
- Vértessy BG, Tóth J (2009), *Keeping Uracil Out of DNA: Physiological Role, Structure and Catalytic Mechanism of dUTPases*, Acc Chem Res. **42**:97-106
- Waters LS, Storz G (2009), *Regulatory RNAs in bacteria*, Cell **136**:615-28.
- Woodside MT, Valentine MT (2009), *Single molecule manipulation using optical traps*, Handbook of Single-Molecule Biophysics XXIV, Chap **12**, ISBN: 978-0-387-76496-2
- Zhuang X (2005), *Single-Molecule RNA Science*, Annu. Rev. Biophys. **34**:399-414
- Zwanzig R, Szabo A, Bagchi B (1992), *Levinthal's Paradox*, Proc. Natl. Acad. Sci. USA, **89**:20-2.

## Chapter 2: Optical tweezers

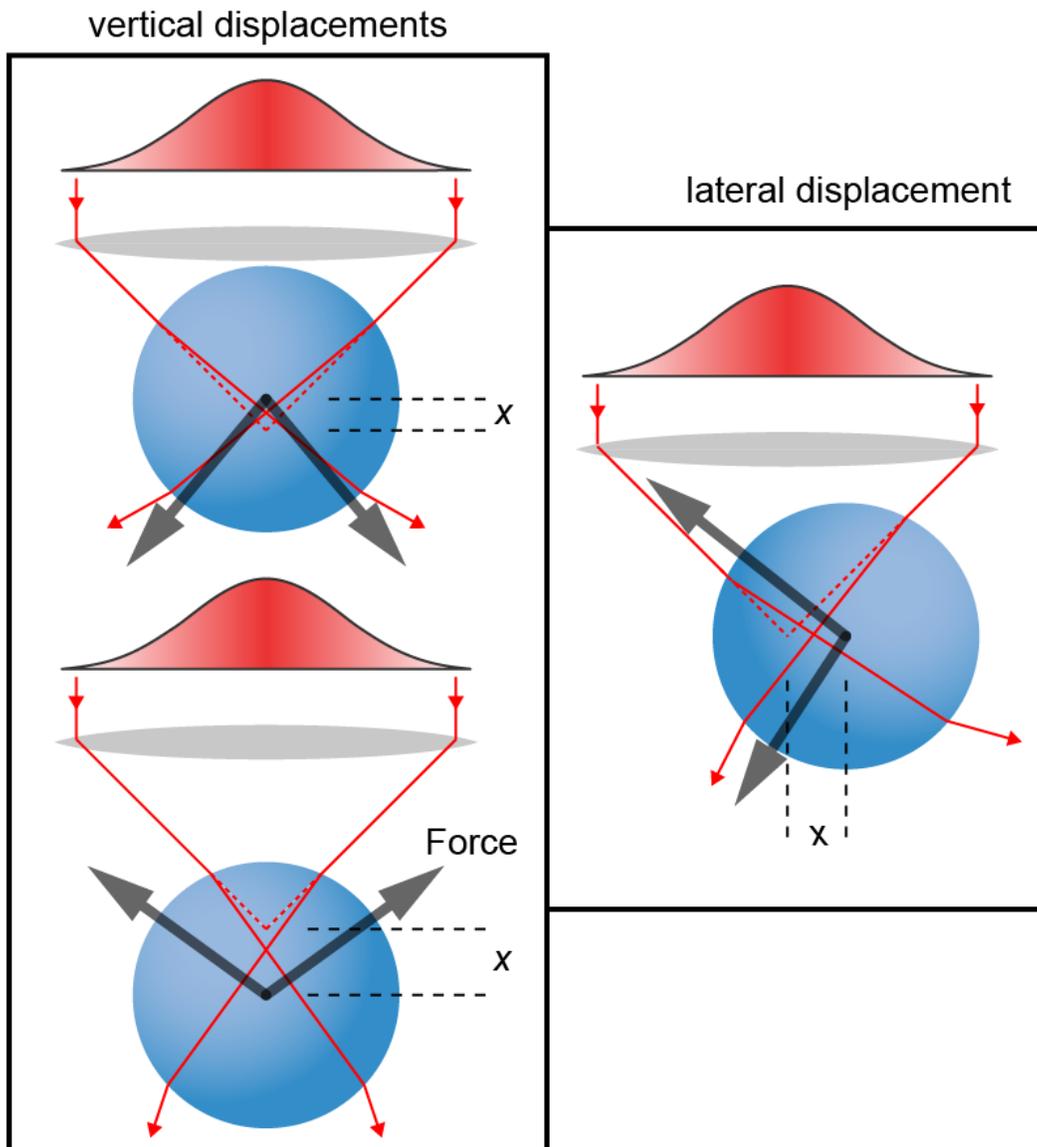
An optical trap (also known as optical tweezers) is a device that allows one to capture a small dielectric object in a focused laser beam. Once an object is trapped, it can be moved and manipulated in a controlled manner. Optical traps are often built around a microscope, because the microscope optics allow 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 objective and condenser lenses. As will become apparent in the discussion of optical trapping principles below, the lenses must have high numerical aperture (NA, defined in microscopy as the sine of 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 dielectric objects, such as glass or plastic (polystyrene) beads. Out of the electromagnetic interaction between the light and the dielectric, there exists a small region where the bead is held by a harmonic force, like a Hookean spring in three dimensions.

Optical tweezers have become well established ever since the pioneering efforts of Arthur Ashkin, Steven Chu and their colleagues at Bell Labs (Ashkin 1986, 1971), and the technology continues to be refined by the contributions of many groups, with applications in many different fields: Adding Raman spectroscopy (Snook 2009), stretching cells (Guck 2005), trapping metal nanoparticles (Dienerowitz 2008), using structured light fields (Dholakia 2008).

Some special considerations must be taken for manipulating single biological molecules. Molecules themselves are too small to trap directly, so instead they must be handled by proxy: they are attached to dielectric beads via molecular “handles.” These beads are then held in the trap. The dielectric beads are functionalised chemically, allowing them to be connected to the molecule being studied with chemical specificity. This specificity ensures that the geometry of the experiment (bead—handle—molecule) is well controlled (see Fig 1.9).

## 2.1 Theory of optical trapping

The tightly focused laser beam of an optical trap forms a double cone with a diffraction-limited spot at the apex where a dielectric particle will experience a net restoring force that effectively traps the particle. For a small range of displacements, this restoring force is linear with displacement and the trap can therefore be considered as a Hookean-spring. Two limits can be considered for modeling an optical trap: Rayleigh scattering, where the wavelength  $\lambda$  is much larger than the particle radius  $a$ ; or Mie scattering (ray optics), where the reverse holds (Fig 2.1). In the case of optical traps suitable for biological molecules, near-infrared wavelengths of  $\sim 1000$  nm are generally used in order to avoid damaging the sample (Neuman 1999). Since physical constraints from the trapping force limit the typical bead sizes for effective trapping to a range from a few 100 nm to a few 1000 nm, *i.e.*  $a \sim \lambda$ , neither limit is truly appropriate. However, these two limits provide important physical insights into how trapping works, so both descriptions will be examined briefly.



**Figure 2.1 Rayleigh and Mie limit forces. Bead displacement with respect to the beam gradient: the unbalanced force pushes the bead to the centre of the trap. In the Mie limit, ray optics defines the unbalanced force.**

### 2.1.1 Optical trapping in the Rayleigh limit

We first consider the interaction between a dielectric sphere and the incident EM radiation in the Rayleigh regime, where  $a \ll \lambda$ . A force is exerted on the particle by the electric field, and it can be broken into two components: a scattering force  $F_{scat}$ , from photons that are elastically scattered (absorption and re-radiation), and a gradient force  $F_{grad}$  arising from the interaction of dipoles induced in the dielectric with the gradient of the electric field.

The time-averaged axial scattering force is determined by the intensity of the laser beam,  $I_0$ , the scattering cross-section of the particle,  $\sigma$ , and the speed of light in the medium (Harada 1996, Ashkin 1986):  $F_{scatt} = I_0 \sigma n_m / c$ , where  $n_m$  is the index of refraction of the medium. For a sphere of radius  $a$ , the scattering

cross-section is:  $\sigma = \frac{128\pi^5 a^6}{3\lambda^4} \left( \frac{m^2 - 1}{m^2 + 2} \right)^2$ , where  $\lambda$  is the wavelength of the light

and  $m$  is the relative index of refraction, i.e.  $m = n_p/n_m$  the ratio of the index of refraction of the particle to that of the medium (Kerker 1969, Jackson 1999). The

gradient force is calculated from the time average of the Lorenz force on dipole,

$p$ , induced in the polarisable sphere:  $\vec{F}_{grad} = (\vec{p} \cdot \vec{\nabla}) \vec{E}$ , where  $\vec{p} = \alpha \vec{E}$  for a

polarisability  $\alpha$ . Taking the time average of this force, we find:

$$\langle \vec{F}_{grad} \rangle = \langle (\vec{p} \cdot \vec{\nabla}) \vec{E} \rangle = \frac{2\pi n_m}{c} \left( \frac{m^2 - 1}{m^2 + 2} \right) a^3 \vec{\nabla} I_0 \quad \mathbf{2-1}$$

When the relative index is greater than unity, the gradient force is in the same direction as the intensity gradient, or in other words it points towards the most intense part of the beam, which is the focal point. Thus when the bead moves

away from the focal point, it experiences a restoring gradient force pushing it back to the focal point. This is the component of the force that creates the trap. The scattering component, on the other hand, is always in the direction of the laser propagation, with the result that any trapped particle will be pushed away from the focal point of the laser beam.

From these equations, the scattering force clearly increases with the bead radius much faster than the gradient force. If the particle is too large, the scattering force will overcome the gradient force and no trapping will take place. However, if the particle is too small, the trapping potential created by the gradient force may not be large enough compared to the thermal energy,  $k_B T$ , to keep the particle trapped against thermal fluctuations. The Rayleigh scattering picture therefore indicates that the range of sizes for successful trapping is limited. Typically this range is  $\sim 200\text{-}5000$  nm, depending on the index of refraction. For particles that are trapped effectively, the scattering force ensures that the equilibrium position of the trapped particle will always be axially offset (in the direction of the beam propagation) from the actual focal point of the laser beam.

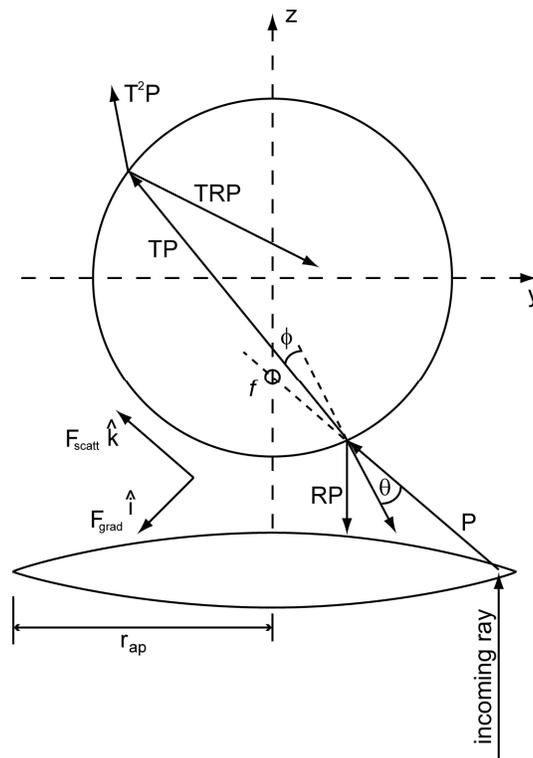
### ***2.1.2 Optical trapping in the Mie limit***

In the regime where the particle is much larger than the incident wavelength, geometric optics can be used to describe the forces while ignoring diffraction limits. The force on the sphere is given by (Svoboda 1994, Ashkin 1992):

$$\begin{aligned} \vec{F} = & \frac{n_m P}{c} \left( 1 + R \cos 2\theta - \frac{T^2 [\sin(2\theta - 2\phi) + R \cos 2\theta]}{1 + R^2 + 2R \cos 2\phi} \right) \hat{k}_{scatt} \\ & + \frac{n_m P}{c} \left( R \sin 2\theta - \frac{T^2 [\sin(2\theta - 2\phi) + R \cos 2\theta]}{1 + R^2 + 2R \cos 2\phi} \right) \hat{i}_{grad} \end{aligned} \quad \mathbf{2-2}$$

Here, the  $k^{\text{th}}$  and  $i^{\text{th}}$  components represent the scattering and gradient components of the force,  $R$  and  $T$  are the Fresnel reflection and transmission coefficients,  $P$  is the power of the ray incident at angle  $\theta$ , while  $\phi$  is the angle of refraction and  $n_m P/c$  is the momentum. This ray optics model, originally described by Ashkin, is illustrated in figure 2.2 (Ashkin 1992).

The ray optics picture does not show any dependence of the forces on bead radius and thus cannot explain the size effects in trapping. However, it does indicate the importance of the NA of the focusing system, since high-angle rays contribute the most to the trapping force (especially the axial force). One important practical result is that stronger trapping is predicted to be achieved by overfilling the aperture of the objective lens with the laser beam, despite losing beam intensity (Svoboda 1994). Another result is that traps created from a  $\text{TEM}_{01}^*$  (“doughnut”) mode, rather than the usual gaussian  $\text{TEM}_{00}$  mode, should have a higher stiffness, due to higher laser intensity on the periphery of the beam. Such traps have in fact been demonstrated in recent years (Sakai 2007, Jesachar 2004).



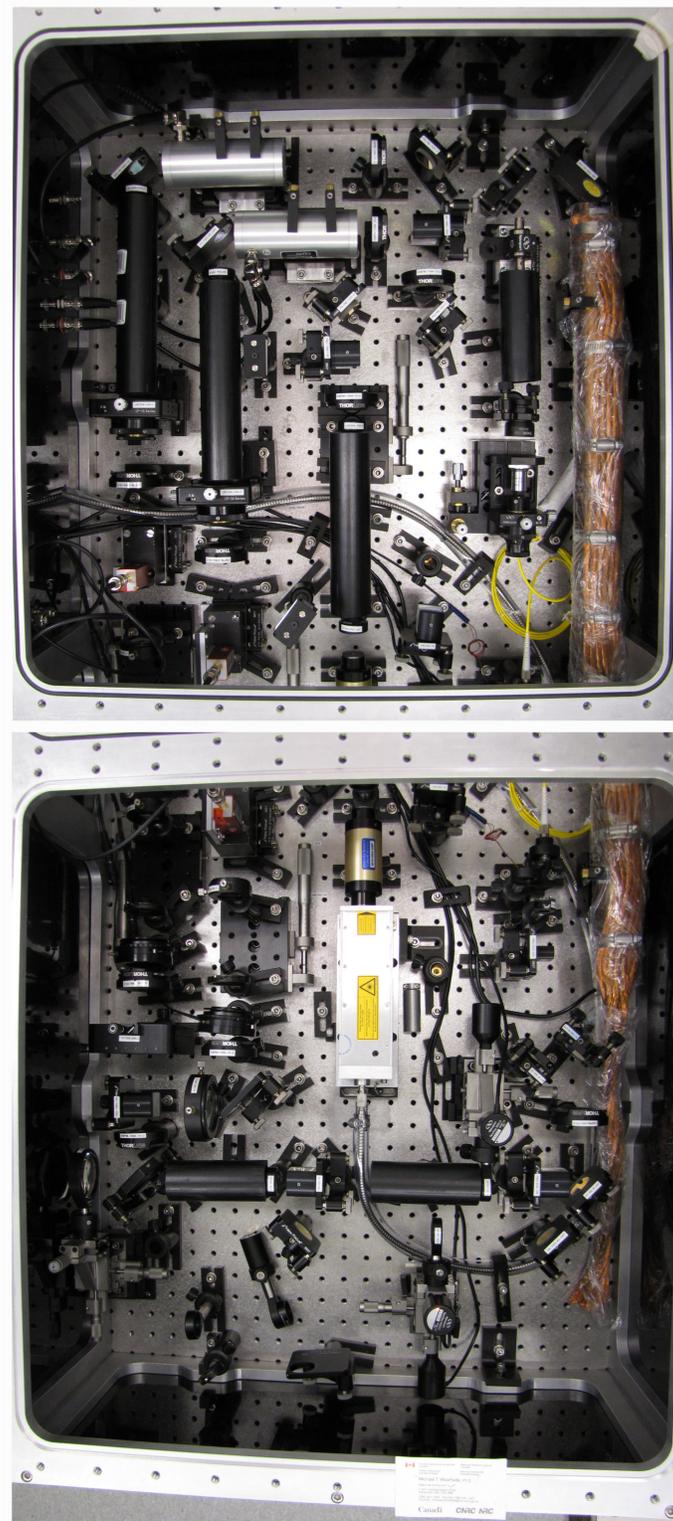
**Figure 2.2 Geometric optics ray diagram. An extremal ray enters the aperture of the objective lens and is focussed toward the focal point at  $f$ . The ray is incident on the bead at angle  $\theta$  and refracted at angle  $\phi$ , giving rise to the reflected and transmitted powers RP, TP, TRP,  $T^2P$ , etc... (after multiple reflections/transmissions).  $r_{ap}$  and  $\hat{n}$  represent the radius of the aperture and the vector normal to the surface of the bead at the intersection of the extremal ray. Adapted from (Svoboda 1994).**

As stated above, most optical traps do not operate in either of these limits, and the results of these two pictures must be accepted only qualitatively (and provisionally). For example, ray optics predicts a particle-size-independent force, contradicting the Rayleigh picture prediction that the force scales with the cube of the radius; experimental behaviour is closer to the Rayleigh picture (Perkins 2009). More generally, to obtain quantitative agreement with experiment, more

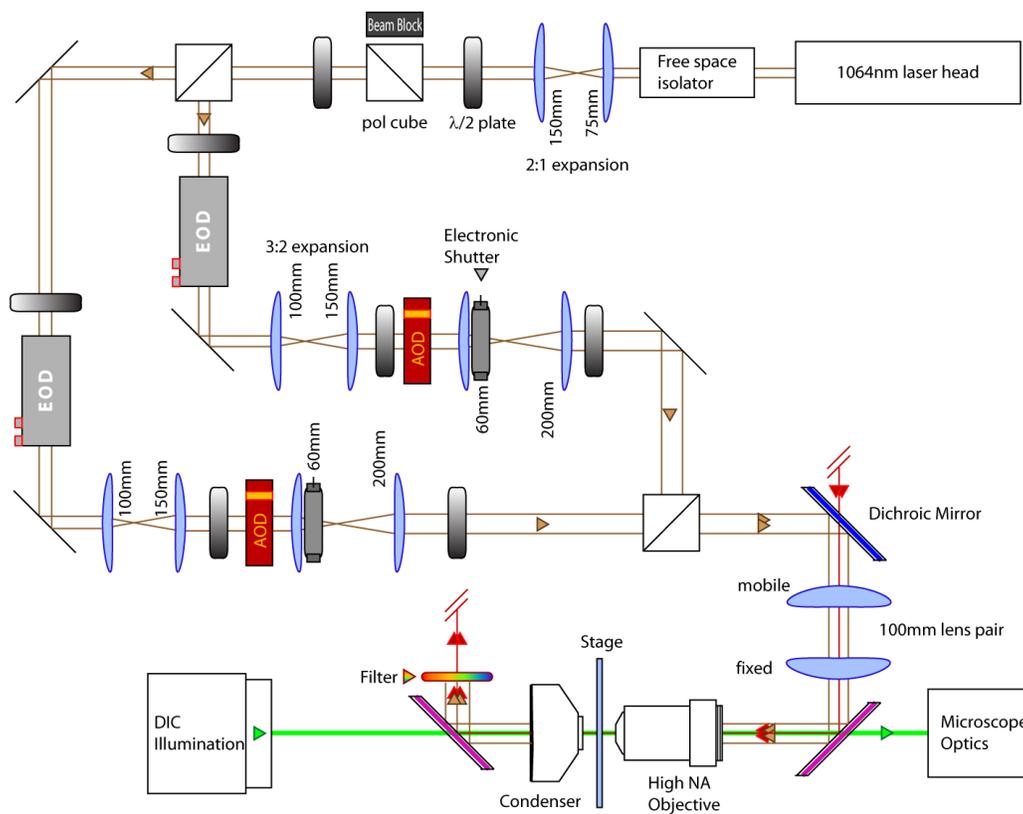
complex theories have been developed. These can be reasonably successful: for example, work by Rohrbach found that experimental results differed from theoretical predictions by ~5% for lateral trapping and ~10-30% for axial trapping (Rohrbach 2005). However, no new qualitative understanding was gained from this treatment.

## **2.2 Construction of the optical trap**

In order to make careful measurements of biomolecular folding reactions, sensitive optical tweezers capable of measuring motions of as little as a single Å are required. These are very complex instruments with many parts. All optical traps contain the same basic elements: laser light source, beam control optics, focusing optics, and detection optics. The beam control optics are used to shape the beam properly (e.g. overfilling the focusing objective lens) and to allow the position of the beam to be controlled, before the beam is inserted into the focusing optics (often a microscope objective lens) and then routed to a detector. To obtain the highest resolutions, great care must be taken in designing and constructing the trap optics, trying to minimise the many different potential sources of noise. The design of the optical trap that we have built is described below, along with the concerns that motivated this design. Our design uses two different laser sources, one to generate the beam that traps the beads, the other to generate a beam used to detect the motions of the beads. For clarity, the layout of the trapping and detecting beam optics are described separately. Schematics of the trapping beam optics and detection beam optics are shown in figures 2.4 and 2.8, respectively.



**Figure 2.3** Photograph of optical layout. This shows the optical layout inside the optics enclosure in the lab. The infrared beam path is clarified in the next figure.



**Figure 2.4** Diagram of the layout of the infrared trapping beam. The beam is isolated from the laser using a Faraday free space isolator, and split by polarisation using polarising beam splitters (PBS – pol cubes). Electronic shutters are placed away from the high intensity focal points of the telescopes.

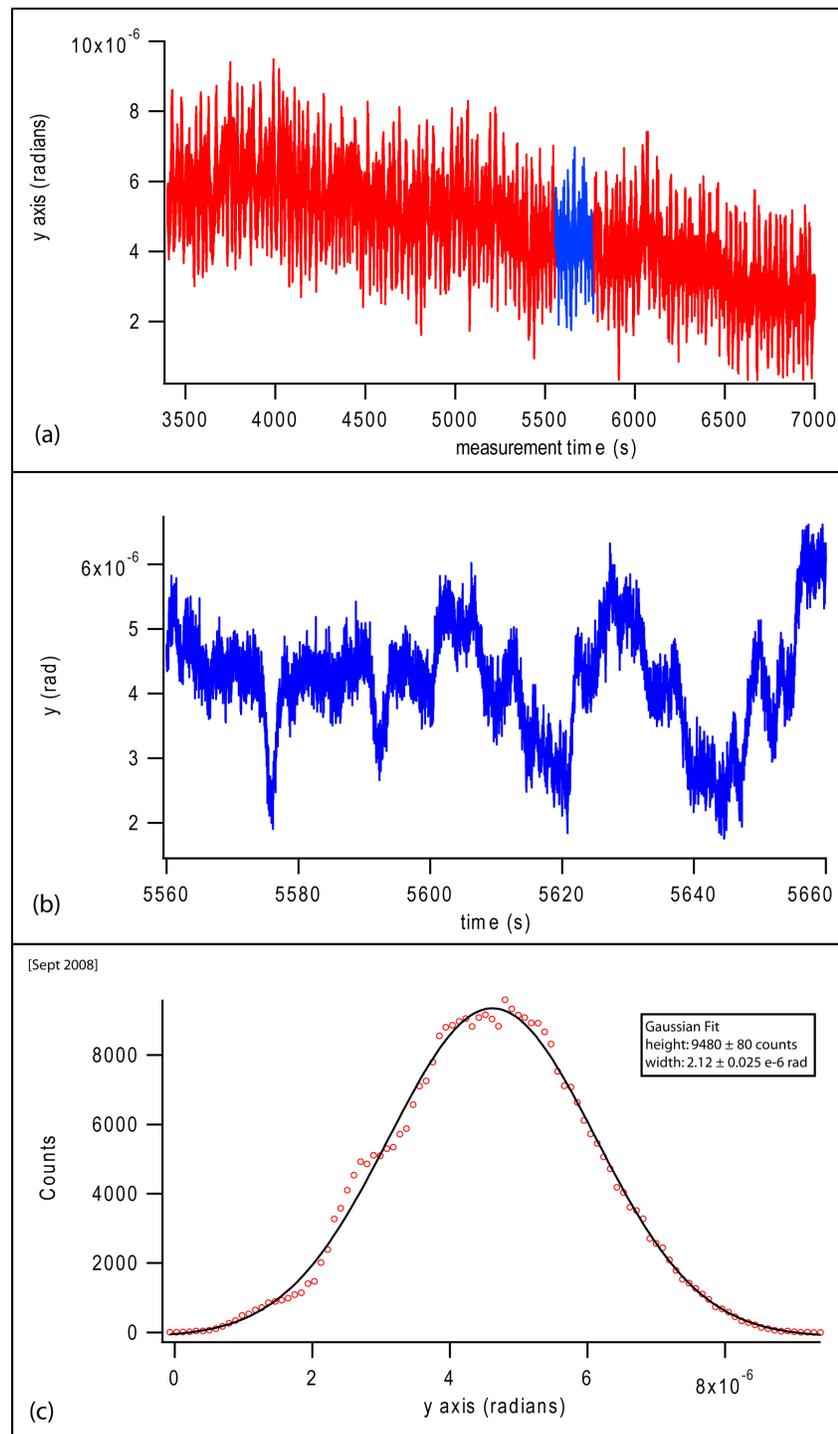
## 2.3 Layout

### 2.3.1 Trapping beam

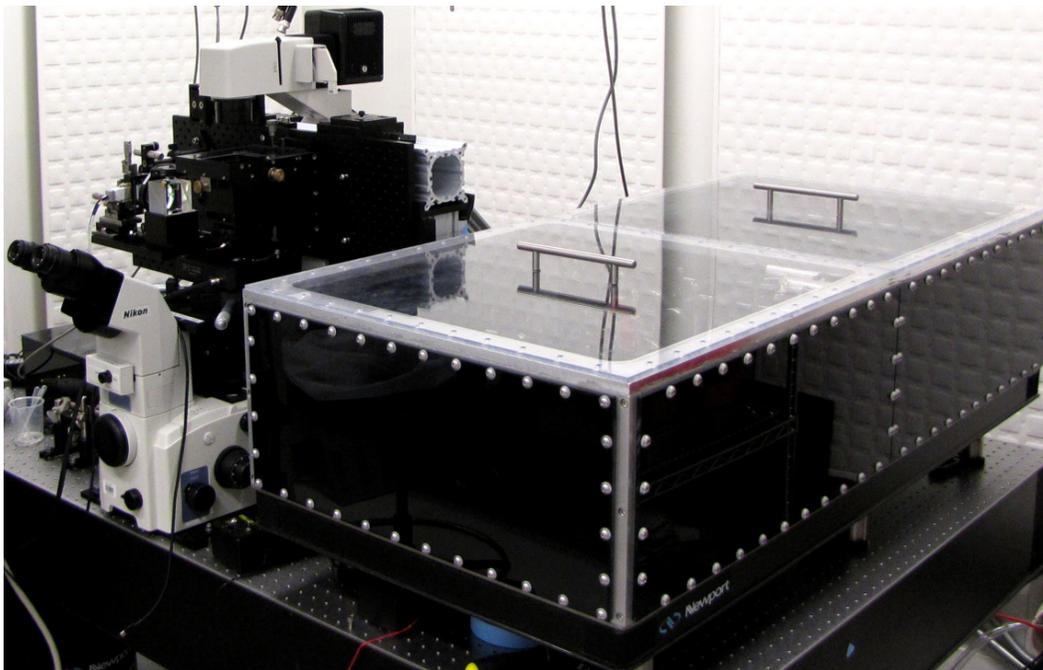
The trapping laser beam source is a 5W Nd:YVO<sub>4</sub> diode-pumped solid state (DPSS) infrared laser with a wavelength of 1064nm. The laser output is a polarised, single mode (Gaussian TEM<sub>00</sub>) beam, collimated in an external 2x Keplerian beam expander. The power stability and pointing stability (i.e., changes in the direction of propagation) of the trapping laser are crucial characteristics for high-resolution trapping, since any fluctuations in the beam will be perceived as

motions of the molecule being studied. To prevent mode hopping due to back-reflections, which can result in changes in power and pointing, the laser head is optically isolated with a Faraday isolator. The intrinsic fluctuations of the laser beam are also characterised and monitored on a regular basis. A typical measurement of the pointing stability is shown in figure 2.5a: fluctuations occur on the order of 5  $\mu$ rad over the course of an hour. Fluctuations on long ( $>1$  s) timescales dominate (quantitatively); as measurements take place over a timescale of  $\sim 10$  s to minutes, these are a particular concern.

In addition to intrinsic noise from the laser source, vibrational and acoustic noise can couple into the optical tweezers. Air currents generated from ventilation or convection can cause the beam pointing to fluctuate, as can temperature changes (due to differential thermal contraction or expansion). We therefore make great efforts to reduce all noise sources of this type. The optical tweezers apparatus is built on an optical table mounted on an isolated concrete slab, to minimise the introduction of environmental vibrations. The temperature of the lab is controlled by thermal panels, ensuring temperature drift of less than 0.1  $^{\circ}$ C per hour. Noisy equipment such as power supplies and amplifiers are stored outside the lab, to reduce acoustic noise. Finally, the effects of convective air currents are reduced through three measures: the optical path of the trapping beam is kept within a solid enclosure (Fig 2.6), within this enclosure the beam path is protected by rigid tubing (especially at the focal points), and the enclosure can be filled with He gas if needed to reduce the optical effects of any remaining air currents (since He has an index of refraction 10 times closer to unity than that of air).



**Figure 2.5 Stability of the trapping laser. The stability trace over the period of an hour, following 1 hour of warmup in (a), a 100s closeup of the fluctuations in (b), and a Gaussian fit to the histogram of that hour of measurement. (Data courtesy of Hao Yu.)**

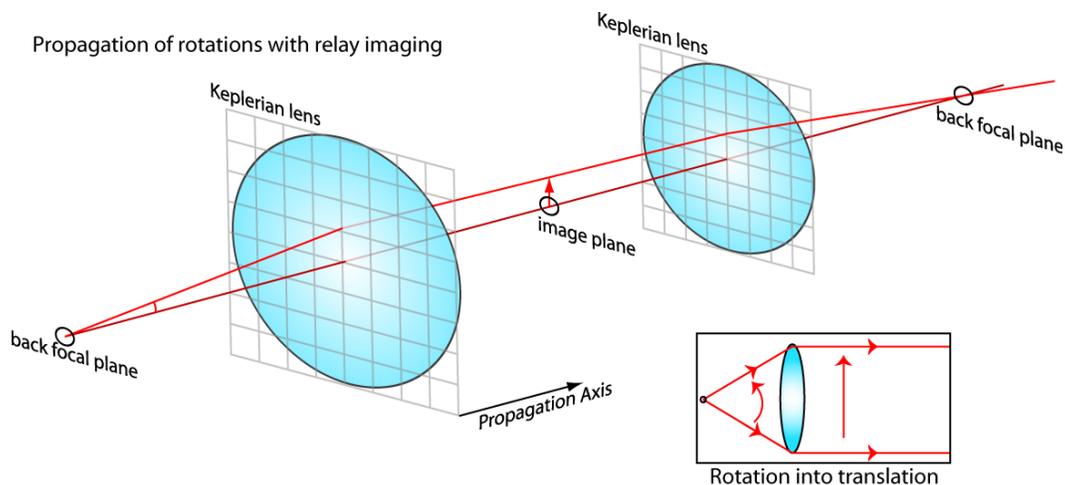


**Figure 2.6 Optics enclosure**

A crucial design aspect for reducing measurement noise is to use a differential measurement. This is done by making two trapping beams (for two traps), by splitting the output of the laser source into two orthogonal polarisations. These two polarisations share the same optical path throughout much of the instrument, and hence much of the noise to which they are subjected is the same for both traps. Thus while pointing fluctuations will change the absolute positions of the traps at the microscope slide (the specimen plane), since both traps are from the same source and are subject to the same noise these fluctuations will not change the relative positions of the traps, permitting higher-resolution measurements. The source output beam is split in two using a polarising beam splitter (PBS) after a  $\lambda/2$  (half) waveplate (to rotate the linear polarisation). The

two beam intensities are set to be unequal, one ~50% more intense than the other, for implementing a passive force clamp (described in chapter 4).

After two trapping beams have been created, they are both passed through optical elements that allow the beam to be steered in the specimen plane. (In the context of microscopy, there are two important planes: the specimen or image plane found at the focus of the objective lens, and the back focal plane. All focal planes can be related to these two as their respective conjugates.) Since translations in the specimen plane correspond to rotations in the back focal plane of the objective lens (Shaevitz 2006), as illustrated in figure 2.7, optical elements that can rotate the beam must be placed in planes conjugate to the back focal plane.



**Figure 2.7 Conjugate optical planes in microscopy.** The microscope slide is in the image plane, while our deflectors are in the planes conjugate to the back focal plane. They are placed in this plane so that the deflections of the beam they provide create translations in the image plane. Inset: Lenses turn rotations into translations. A pair of lenses set up as a Keplerian telescope is used to propagate the rotations to the next back focal plane conjugate: the back aperture of the objective (Shaevitz 2006).

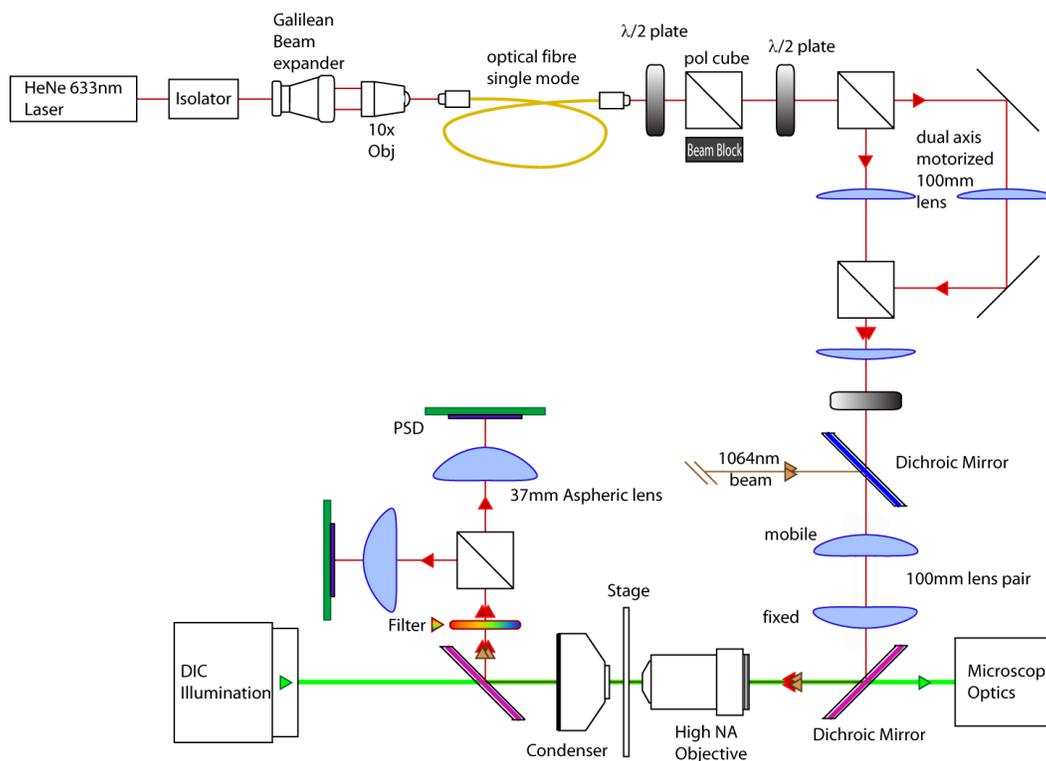
The trap we have designed uses three different steering elements: first an electro-optic deflector (EOD), then an acousto-optic deflector (AOD), and finally a lens mounted on a 3-axis translation stage. The EOD and AOD operate on different physical principles, and have different practical advantages and disadvantages as deflectors, described in more detail in section 2.33. The EOD deflects the beam in the x-axis, the AOD in the y-axis, and the lens is used to move both the trapping and detector beams together in x, y, and z (axial) directions.

Since the deflectors must be located at planes conjugate to the back focal plane, we use Keplerian telescopes to relay the back focal plane (figures 2.4, 2.7). These telescopes are also used to expand the diameter of the beam, so as to overfill the objective lens. A 1:1 telescope relays the back focal plane from the objective lens to the steering lens, a 3.3:1 telescope relays it between the steering lens and the AOD, and finally a 1.5:1 telescope relays it between the AOD and the EOD. The magnifications were chosen to ensure that the beam is wide enough to overfill the objective without losing too much beam power.

Each beam can be blocked by an electronically-controlled shutter located after the AOD. Once the two beams have passed through the EOD, AOD, and expanding telescopes, they are recombined using a PBS. The two trapping beams are then combined with the two detection beams using a dichroic mirror, and sent into the microscope via the steering lens. The objective lens used to focus the traps is a very high NA, oil immersion lens: 100x, NA 1.49 (Nikon).

### 2.3.2 Layout - Detection beam

The detection beam uses a high-power HeNe laser at 633nm (fig 2.8). Most of the experimental issues discussed for the trapping beam apply equally to the detection beam, except that the detection beam does not need to overfill the objective, and it does not need to be actively (and rapidly) steered. As a consequence, the detection beam path is simpler. The laser output is first optically isolated, expanded, and collimated, and then it is coupled into a single-mode optical fibre. The fibre reduces beam pointing fluctuations, converting them into intensity fluctuations (which are less critical for the detection beam), and also prevents the propagation of unwanted modes.



**Figure 2.8 Detector laser layout diagram. Notable differences from the trapping optics are deflection via motorised lenses, and manual fibre-coupling of the laser.**

Just as with the trapping laser, the detection laser output is divided into two orthogonally polarised beams using a  $\lambda/2$  waveplate and PBS, though here the beam intensities are equal. While the detection beams do not need to be steered actively, they do need to be aligned very precisely on top of the trapping beams. To do this, we use steering lenses placed at a conjugate back focal plane. These steering lenses are part of the 1:1 Keplerian telescopes (just like the steering lens for all of the beams described above) that relay the back focal plane to the appropriate positions. These lenses are mounted on motorised 3-axis translation stages, to align the detector beams precisely over the equilibrium positions of the two traps. After passing through the objective lens, the specimen plane, and the condenser lens, the detection beams are isolated from the trap beams by filters, separated by polarisation using a PBS, and focused onto duo-lateral position sensitive photodiodes (PSDs) located near the plane conjugate to the back focal plane of the condenser lens.

### ***2.3.3 Beam steering considerations***

The components used for beam steering need to be stable when static, but have the ability to make rapid, repeatable deflections of the laser beam. We have chosen to use a combination of different steering elements, notably EODs and AODs.

#### ***2.3.3.1 Acousto-optic deflectors (AODs)***

AODs function as a random access diffraction grating created by sound-induced index changes in a  $\text{TeO}_2$  crystal. It is sandwiched between a piezoelectric transducer, driven by a radiofrequency (RF) signal, and an acoustic absorber, that

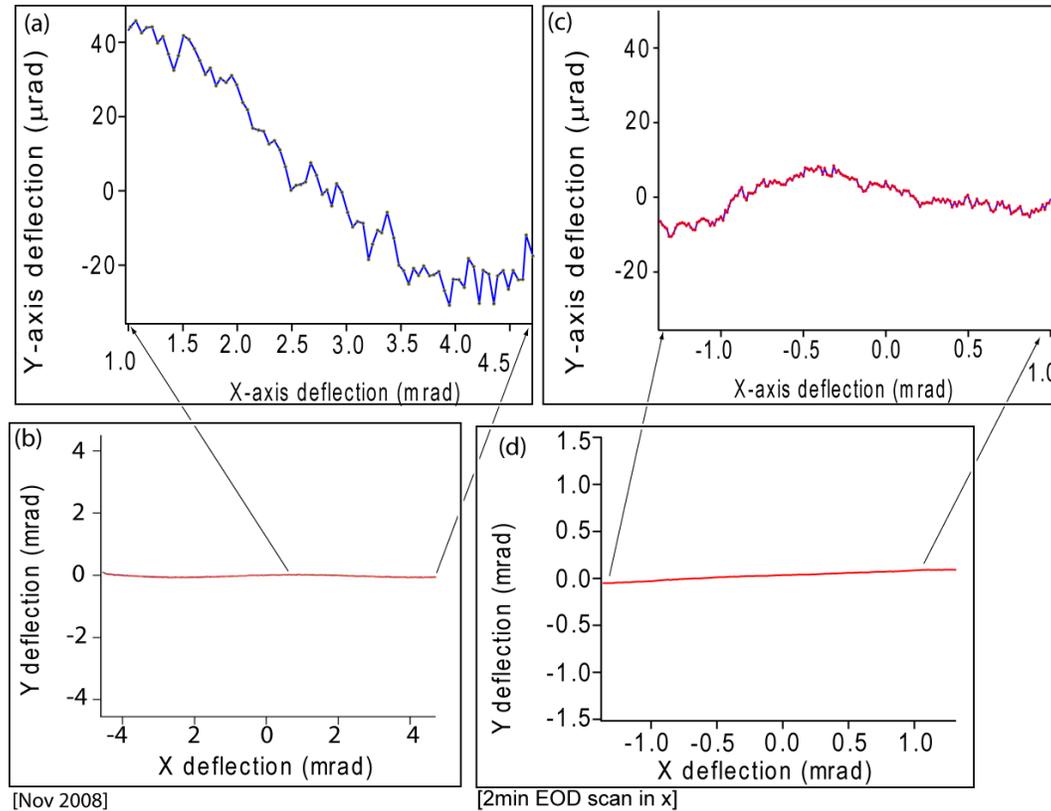
is cut at angle with respect to the transducer-crystal interface to reduce reflection. The transducer sets up a travelling sound wave in the crystal. The shift in density of the crystal due to the sound wave alters its index of refraction, thereby creating an index grating in the crystal with a spacing set by the sound wavelength. By changing the driving frequency of the transducer, the deflection angle of the diffracted beam can be changed. The deflection angle  $\Delta\theta$  of the first order diffracted beam is given by:

$$\Delta\theta \propto \lambda f / v_{sound}, \quad \mathbf{2-3}$$

where  $\lambda$  is the wavelength of light in vacuum,  $f$  is the transducer driving frequency, and  $v_{sound}$  is the speed of sound in the crystal. Deflections of up to 10's of mrad can be achieved. Changing the magnitude of the transducer driving frequency changes the amplitude of the diffracted beam. AODs thus allow both the direction and amplitude of the beam to be controlled. The timescale for deflection changes is also fast, related to the propagation time for the sound wave across the crystal (on the order of  $\mu\text{s}$ ).

While AODs have many advantageous features, there are several drawbacks to their use. First, they cause significant optical loss due to light in the unused diffraction orders, typically 20-40% per AOD. Their frequency response is not always flat, resulting in changes in light intensity as the beam is deflected. Most troubling for precision optical traps, however, the deflection using AODs has been found not to be completely linear for very small deflections ( $\mu\text{rad}$  scale), possibly due to reflections from the absorber do occur that persist as a standing

wave and degrade the performance. These non-linearities result in periodic deviations from a straight deflection (Fig 2.9a).



**Figure 2.9 Linearity of optical deflectors response. (a, b) Deflection response of beam using AOD when sweeping in the y axis. Quasi-periodic ‘wiggles’ are observed. (c, d) Deflection response from an EOD lacks such features, and also has a smaller range. The diagonal trend in (d) is due to misalignment between the EOD and PSD.**

### 2.3.3.2 Electro-optic deflectors (EODs)

EODs involve a  $\text{LiTaO}_3$  (lithium tantalite) crystal whose index of refraction can be changed by the application of an external electric field. An index gradient is established laterally across the crystal by external electrodes, deflecting the light passing through the crystal. The angle of refraction is given

by:  $\theta = \kappa l V / w^2$ , where  $l$  is the length of the crystal,  $w$  the width,  $V$  the applied voltage, and  $\kappa$  is a constant specific to the material (Neuman 2004). EODs overcome many of the drawbacks of AODs (Valentine 2007): transmission losses in the deflected beam are typically only a few percent, the intensity does not vary as the beam is deflected, and the deflection response is very linear (and even faster than for AODs). This linear response is illustrated in figure 2.9c. However, EODs have their own, different disadvantages. The deflection efficiency is much lower; even with very long crystals ( $\sim 12$  cm), it is difficult to generate deflections larger than 1 mrad. The long, narrow crystals also make alignment difficult, and require longer beam paths, increasing the vulnerability to noise. Most problematically, a low-intensity “ghost” beam that is not properly deflected is usually seen (Valentine 2007), which can degrade the trap performance. This ghost beam, made up of the component of the polarisation perpendicular to that deflected by the index gradient, can be minimised (but not completely removed) by very careful alignment. In our trap, the ghost beam represents a 1-2% loss in trapping power.

#### ***2.3.4 Position Detection***

The positions of the beads in the optical traps are detected by collecting the light scattered from the beads on a PSD. The detection beam is at much lower power (20 mW) than the trapping beam, so that it does not create its own trapping potential. Just as importantly, it is at a different wavelength, 633 nm, so that it can be separated from the trapping beam with filters. There are many different

detection schemes possible for optical traps; this design that we have chosen permits absolute position detection, simplified position measurements, and quicker position calibration for the trapped bead compared to other setups (Neuman 2004).

The PSDs are sensitive both to the lateral (x,y) position of the bead in the trap, from the differential voltage signal across the PSD, and also to the axial (z) position of the bead, from the total amount of light collected on the PSD (Neuman 2004). Since the lateral and axial detection sensitivity depend in opposite ways on the NA of the detector (lateral detection sensitivity is increased with higher NA, axial sensitivity is decreased), a compromise must be reached. We generally measure the distance between two beads in the same plane (one held in each of the two traps) in a differential configuration, hence we are typically most concerned about lateral sensitivity. When it is necessary to determine the axial position of the bead sensitively, however, such as during trap stiffness calibrations (see section 2.4), the detector NA can be reduced by closing an iris in the back focal plane of the condenser.

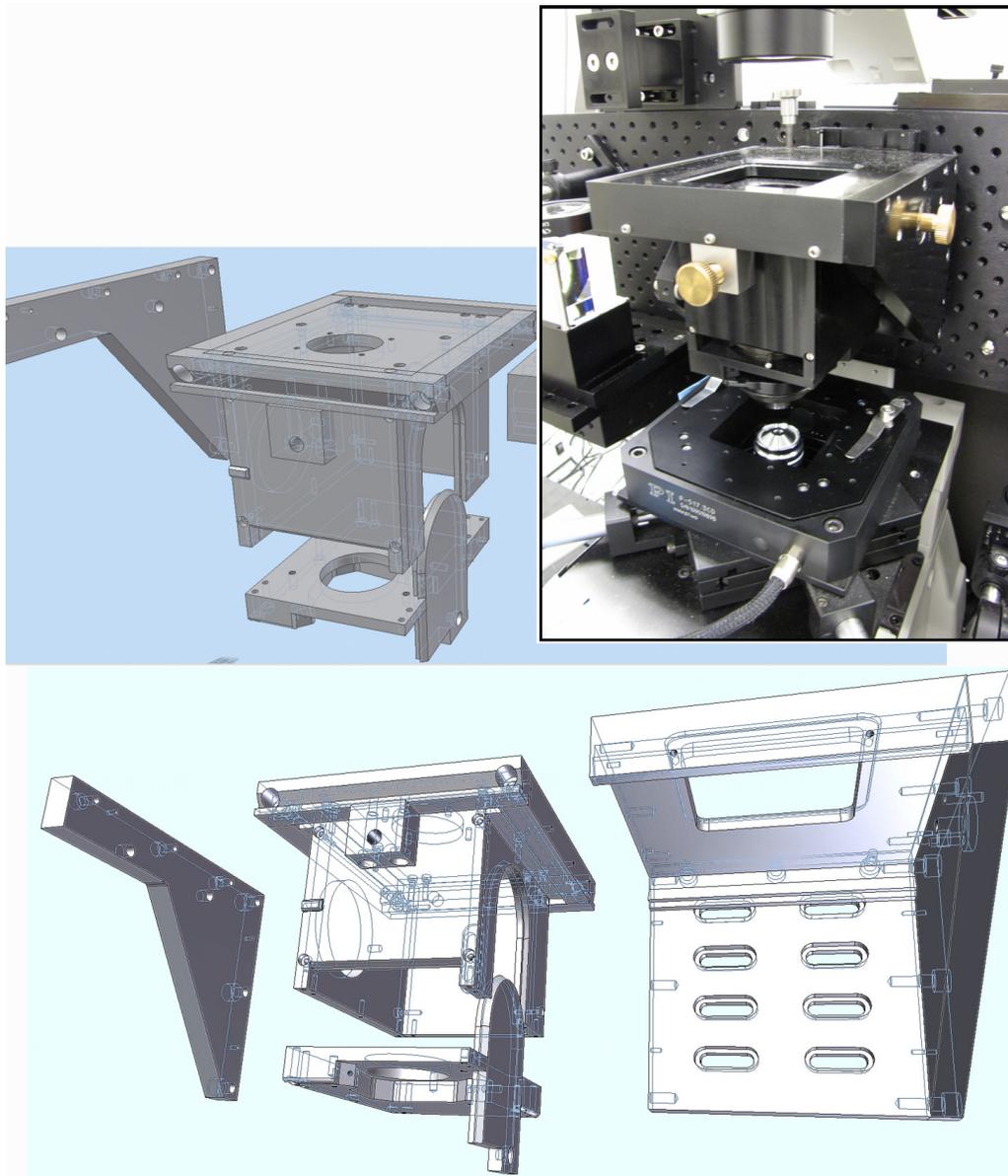
The position detector is calibrated by using the deflectors (AODs and EODs) to scan a bead held in the trap across a 2-dimensional grid of known positions, while measuring the average voltage on the PSD at each position on the grid. The measured voltages in x and y are then fit to the known positions using a polynomial function, providing a position calibration. This calibration must be performed for each trapped bead, due to variations in the bead sizes and shapes (beads may occasionally be oblate rather than spherical, and there is a known

variance in bead radius). Since the PSD voltage response varies linearly with the power of the incident light, the PSD response is normalised by the incident power, so that only the changes in voltage due to actual motions are measured.

Ultimately, the position calibration is traced back to the motion of a piezo-driven nanopositioning stage. This stage is calibrated by the vendor with sub-nm accuracy, allowing it to be used as a “ruler” against which nanoscale motions may be measured. In the laboratory, the control of optical deflectors, nanopositioning stages, and all other instruments, as well the calibration and data acquisition are done in the virtual instrumentation architecture of Labview software (National Instruments).

### ***2.3.5 Microscope***

The optical tweezers are built around an inverted optical microscope (Nikon Eclipse TE2000-U). However, the requirements for structural rigidity demanded by optical trapping are much more stringent than for simple light microscopy, hence the microscope has been extensively rebuilt. The turret at the back end of the microscope holding the condenser assembly and the lamp has been removed entirely and replaced with a framework of cross-braced 95 mm optical rails supporting an optical breadboard. The condenser assembly has also been redesigned and completely rebuilt, to increase the stability (especially reducing the drift) and to add additional optics (a mirror and filters) (Fig 2.10). The microscope stage has been rebuilt for added rigidity, and a frame added beneath to hold optics (dichroic mirror).

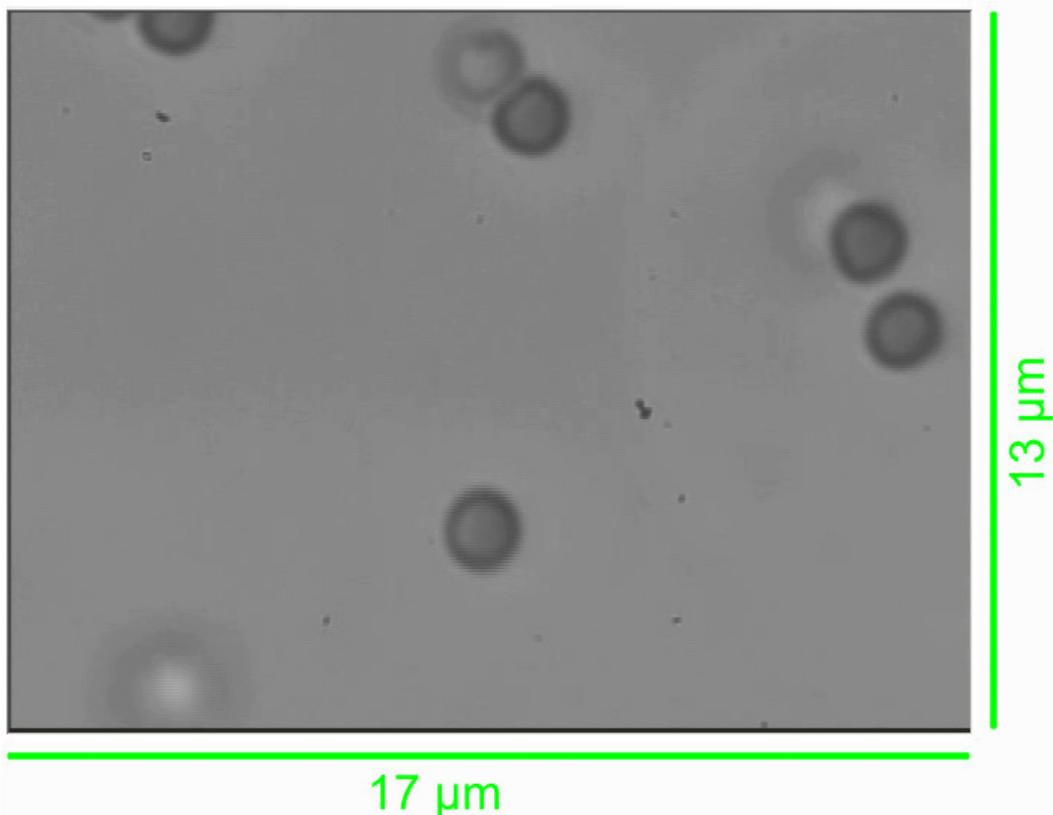


**Figure 2.10 The redesigned condenser assembly. A part of our extensively rebuilt microscope, shown in a semi-exploded view, all built for rigidity and functionality. The condenser ‘turret’ moves over brass dovetails by bronze on steel screws. All parts were designed in Autodesk Inventor, and machined in the department.**

The lens holders for the objective and condenser have been built to accommodate Nomarski-modified Wollaston prisms for differential interference contrast (DIC)

microscopy, which greatly improves the visualisation. However, these prisms have been removed, as they adversely affect the trap quality.

Samples are visualised primarily using a CCD camera attached to the camera port. This camera also plays an important (though simple) role in position calibration, through calibration of the imaging area of the camera. Video tracking of a bead with the camera is also an important tool for (somewhat crude) position detection (Neuman 2004). Since the field of view of the microscope typically contains many dozens of beads, the image in the camera is magnified by a further 3.5x, providing a field of view of  $\sim 17$  to  $19 \mu\text{m}$  by  $13 \mu\text{m}$  or  $\sim 37.5$  captured pixels per  $\mu\text{m}$  (Fig 2.11).



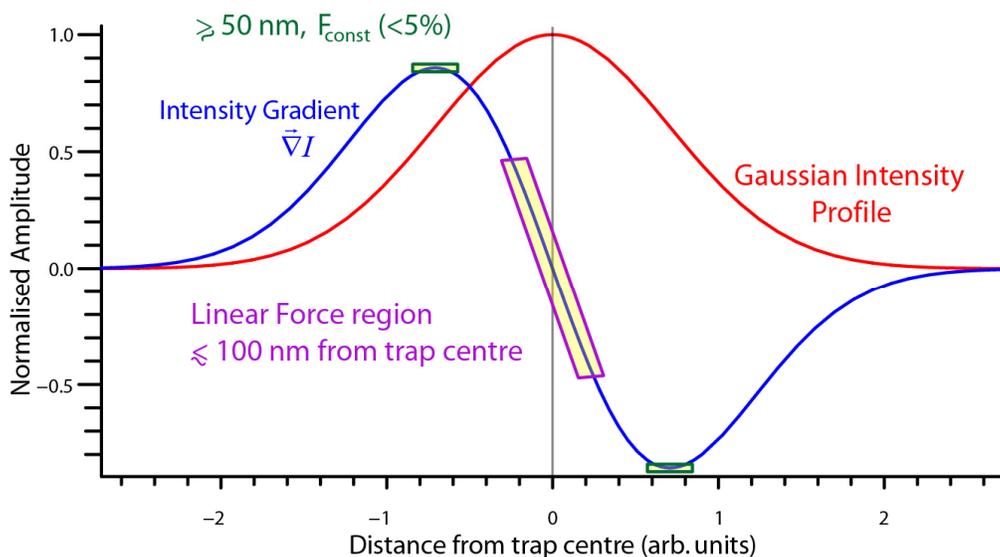
**Figure 2.11** Video capture of  $1.0 \mu\text{m}$  beads. A demonstration of the field of view under the microscope, without DIC prisms in place. (Depth of field  $\sim 0.5\mu\text{m}$ .)

## 2.4 Stiffness calibration of the optical trap

The sensitivity of optical tweezers measurements relies on the calibration of just two parameters: the position of the bead and the stiffness of the trap. Assuming that the bead remains in the harmonic region of the trap (Fig 2.12), the force is then just given by the product of the trap stiffness and the displacement of bead from the centre of the trap. The stiffness of the trap must therefore be calibrated reliably, and the size of the harmonic region (typically  $\sim 100$  nm) verified. There are three different methods of calibration commonly used: the power spectrum of the bead position fluctuations, the variance of the fluctuations, and displacement caused by viscous drag from moving fluid (Woodside 2009). All three are used to obtain a more precise estimate of the stiffness. Because the polystyrene beads vary in size by up to  $\sim 5$  %, according to the manufacturer, we calibrate the stiffness with the three methods for many beads, and average.

### 2.4.1 Power spectrum calibration

The stiffness of the trap can be determined from the power spectrum of the thermal motion of a trapped bead because the trap effectively acts as a filter on the Brownian motion of the bead. For a harmonic trap and white Brownian noise, this spectrum is a Lorentzian characterised by constant power at the low end and inverse square behaviour above a roll-off frequency,  $f_0$ :



**Figure 2.12 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 . For a Gaussian beam, a roughly linear region is observed near the trap centre (within ~50-100 nm). Farther out (~200 -300 nm), there is a constant force region ~ 50 nm wide (Greenleaf 2005). Beyond that region, this simple presentation breaks down, due to the truncation of the trapping beam at the objective lens.**

$$P(f) = \frac{k_B T}{\pi^2 \beta (f_0^2 + f^2)}, \quad P(f) = \rho^2 P_v(f) \quad 2-4$$

Here  $\beta$  is the hydrodynamic drag coefficient (for a sphere of radius  $a$ , this is the Stokes drag:  $\beta = 6\pi\eta a$ , where  $\eta$  is the viscosity of the medium),  $f_0 = \kappa/2\pi\beta$  defines the dependence of the frequency roll-off to the trap stiffness  $\kappa$ ,  $\rho$  is the linear sensitivity of the detector used to measure the spectrum, and  $P_v(f)$  is the uncalibrated power spectrum measured by the detector (Woodside 2009, Neuman 2004, Greenleaf 2007). The derivation of this spectrum follows from an analysis of the Brownian noise of a bead in the small Reynolds number limit, where

viscous drag dominates inertia (Svoboda 1994): the bead is treated as a damped massless oscillator. The equation of motion is:

$$\beta\dot{x}(t) + \kappa \cdot x(t) = F(t). \quad 2-5$$

The driving force  $F(t)$  is white, Brownian noise of zero mean, whose Fourier transform (into frequency space) is with a force-power density (with units of  $N/\sqrt{Hz}$ ) given by  $|\tilde{F}(f)|^2 = 4\beta k_B T$ . The Fourier transform of the bead position from eqn 2-5 is then:

$$2\pi\beta\left(\frac{\kappa}{2\pi\beta} - if\right)\tilde{x}(f) = \tilde{F}(f), \quad 2-6$$

from which equation 2-4 for the power spectrum is derived (Shaevitz 2006). A useful feature of this method is that the detectors do not need to be calibrated for position: the power spectrum may be observed from the raw voltage on the detectors.

As a complicating factor, trapping experiments are typically conducted close to the surface of the coverslip ( $\sim 1 \mu\text{m}$  or less). This modifies the effective viscosity of the medium, since the boundary layer of water increases the drag. The effective drag coefficient for a sphere is approximated according to Faxen's Law (Happel 1983, Svoboda 1994, Neuman 2004):

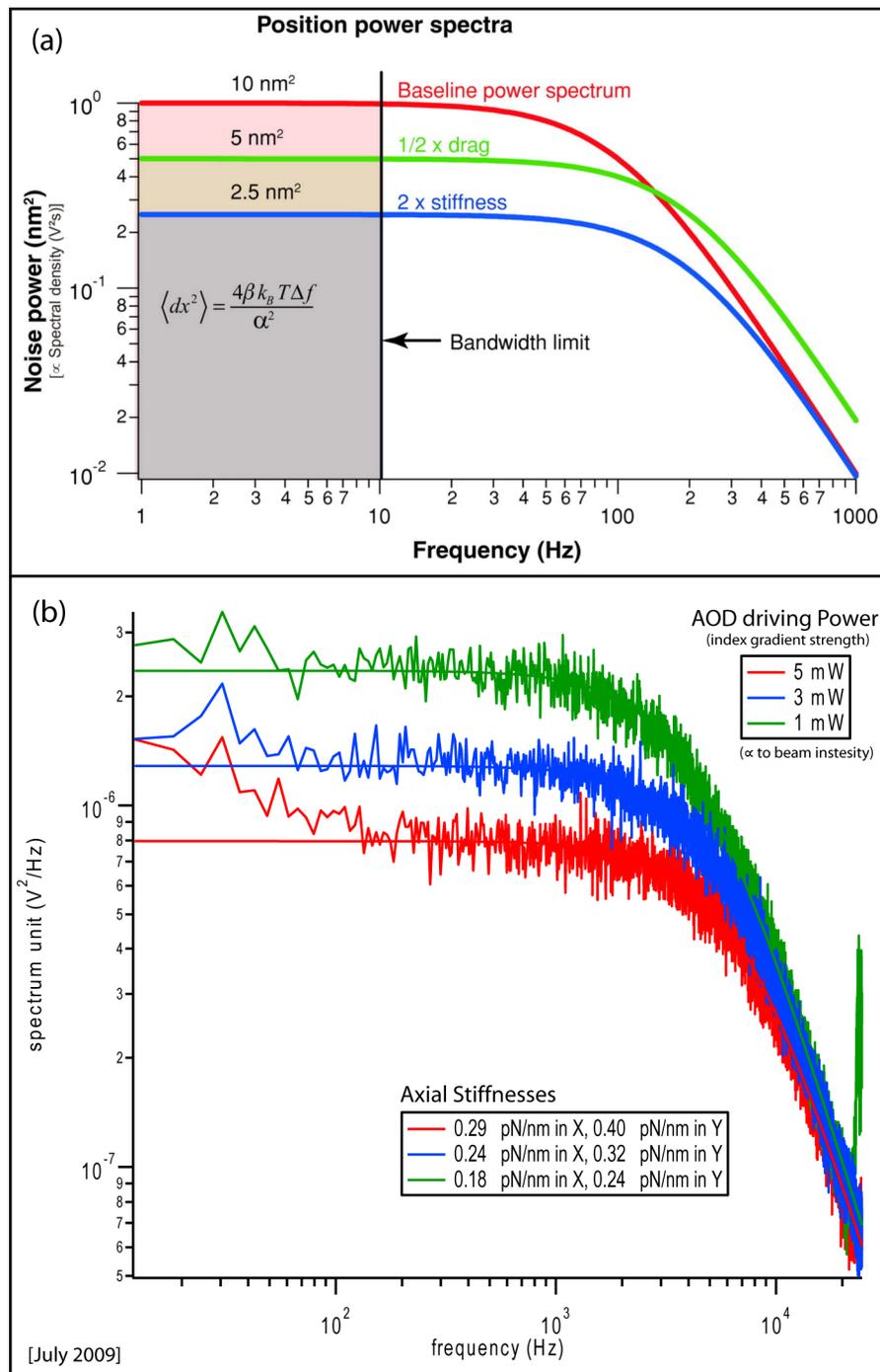
$$\beta' = \frac{6\pi\eta a}{1 - \frac{9}{16}\left(\frac{a}{h}\right) + \frac{1}{8}\left(\frac{a}{h}\right)^3 - \frac{45}{256}\left(\frac{a}{h}\right)^4 - \frac{1}{16}\left(\frac{a}{h}\right)^5} \quad 2-7$$

where  $h$  is the height above the surface, measured using the axial detection signal (which is monitored as the stage is moved up to bring the surface into contact

with the bead). A table of correction values derived from Faxen's law can be found in reference (Svoboda 1994).

The power spectrum is useful for more than simple calibration: the presence of non-Lorentzian spectral shapes indicate that the trap is not a simple harmonic oscillator. The power spectrum can therefore be used to check for beam misalignment, as well as electronic noise and mechanical vibrations (Woodside 2009). In addition, improper filtering of the data can be observed (*e.g.*, a reduction of the noise at high frequency due to unintended electronic filtering or poor photodiode response).

A typical power spectrum from our trap is shown in Fig 2.13b. The area under the curve is the square of the noise amplitude  $\langle dx^2 \rangle$  out to the bandwidth limit or filter cut-off ( $\Delta f$ ). It follows that the force resolution is given by:  $dF = \sqrt{4\beta k_B T \Delta f}$  (Greenleaf 2007). Spatial resolution of the trap can be improved by increasing the stiffness, reducing the bandwidth, or decreasing viscous drag. Note that the bead only samples the harmonic portion of the trap during such a measurement (unless  $\kappa$  is very small), because the energy required to reach the anharmonic part (over 100 nm from the trap centre) is much greater than  $k_B T$ .



**Figure 2.13** Trap power spectrum (a) The theoretical view of the position power spectrum (Greenleaf 2007). (b) Experimental measurements at various trap stiffnesses (data courtesy of Hao Yu).

The force resolution can be improved in a similar manner, but bandwidth reduction presents a dichotomy: temporal resolution is inversely proportional to filtering bandwidth (or  $f_0$  without) (Neuman 2008).

Finally, the power spectrum of a trapped bead can be used to measure sample heating due to partial absorption of laser light, due to the temperature dependence of viscosity,  $\eta(T)$  (Neuman 2004).

### **2.4.2 Variance method**

The variance method for determining the stiffness of the optical trap is based on the equipartition theorem. The thermal energy,  $\frac{1}{2}k_B T$ , drives the position fluctuations of a particle in the harmonic potential. The average energy of the particle in the potential,  $\frac{1}{2}\kappa\langle x^2 \rangle$ , is then related to the thermal energy to provide the trap stiffness in terms of the variance of the displacement of the bead from the trap centre:  $\kappa = k_B T / \langle x^2 \rangle$ . The force resolution follows from Hooke's law:  $dF = \sqrt{\kappa \cdot k_B T}$  (Neuman 2008). This method is in some ways the simplest to implement and interpret, since there is no dependence on fluid viscosity. However, it does require calibration of the position detection.

The variance  $\langle x^2 \rangle$  is equal to the integral of the position power spectrum, so that the two methods are related. However, they have different systematic errors, making it useful to measure the stiffness using both. For example, over-filtering of the data increases the calculated stiffness obtained from variance, but

decreases the stiffness obtained from the power spectrum. The variance is also a biased estimator (“extra” noise or drift contributes to an underestimate of the stiffness of the trap).

While the variance method does not provide a sensitive measure of trap alignment, it does allow a picture of the trap potential to be determined, from the probability distribution of the bead positions:  $P(x) \sim \exp[-U(x)/k_B T]$ , where  $U(x)$  is the energy of the trapped particle at position  $x$ .

### 2.4.3 Viscous Drag Calibration

The third method of determining trap stiffness, by viscous drag, is the most direct. The position of a bead held in the trap is monitored while the stage is moved to drag the bead through the medium at a known speed,  $v$ . The viscous drag force on the bead must be balanced by the restoring force from the trap. Assuming the bead remains in the harmonic part of the potential,

$$F = \beta v = \kappa \cdot x(v), \text{ hence } \kappa = \beta v / x(v). \quad \mathbf{2-8}$$

In practise this is done by driving a piezo-controlled stage laterally in a sinusoidal or triangle-wave oscillation. The characteristic response of a bead to a sinusoidal driving force of amplitude  $A_0$  and frequency  $f$  will be:

$$x(t) = \frac{A_0 f}{\sqrt{f_0^2 + f^2}} e^{-i(2\pi f t - \varphi)}, \quad \varphi = -\tan^{-1} \frac{f_0}{f}, \quad \mathbf{2-9}$$

where  $\varphi$  is a phase delay ( Neuman 2004). In the case of triangle-wave stage motion, the applied force is a square wave and within each period of motion the bead trajectory is given by:

$$x(t) = \frac{\beta A_0 f}{2\kappa} (1 - e^{-t/\beta}) \quad \mathbf{2-10}$$

The stage response has some small lag causing the exponential damping term to be convolved with that response time, suggesting that only the asymptotic value (the leading constant) should be used (Neuman 2004).

One important advantage of the viscous drag calibration is that it permits the region over which the potential can be adequately described as harmonic to be measured. By varying the speed of the stage motion, different regions of the trap may be probed (whereas the two previous calibration schemes were restricted to the proximity of the trap centre), and the linearity of force versus displacement checked. Calibration by viscous drag is however complicated by the fact that the sample stage is never perfectly perpendicular to the laser beam. The effective surface height change with stage position must therefore be measured and compensated for in the calibration (Woodside 2009, Shaevitz 2006).

## **2.5 Summary**

This chapter describes the basis for constructing and calibrating the optical tweezers in our lab. The omnipresent challenges for putting together a quiet and high resolution trap are the beam alignments and calibration of the beads used in the experiment. The fruits of that labour will produce a trap that permits ångström level resolution.

## 2.6 References

- Ashkin A (1992), *Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime*, Biophys. J. **61**:569-82.
- Ashkin A, Dziedzic JM (1971), *Optical Levitation by Radiation Pressure*, Appl. Phys. Lett. **19**:283-5.
- Ashkin A, Dziedzic JM, Bjorkholm JE, Chu S (1986), *Observation of a single-beam gradient force optical trap for dielectric particles*, Opt. Lett. **11**:288-90
- Berg-Sorensen K, Flyvbjerg H (2004), *Power spectrum analysis for optical tweezers*, Rev. Sci. Instrum. **75**:594–612.
- Carter AR, Seol Y, Perkins TT (2009), *Precision Surface-Coupled Optical-Trapping Assay with One-Basepair Resolution*, Biophys. J. **96**:2926–34.
- Dholakia K, Lee WM (2008), *Optical Trapping Takes Shape: The Use of Structured Light Fields*, Adv. at. mol. opt. phys. **56**: 261-337. ISSN: 1049-250X
- Dienerowitz, Mazilu, Dholakia (2008), *Optical manipulation of nanoparticles: a review*, J. Nanophotonics **2**:021875.
- Greenleaf WJ, Woodside MT, Abbondanzieri EA, Block SM (2005), *Passive all-optical force clamp for high-resolution laser trapping*, Phys Rev Lett. **95**:208102.
- Greenleaf WJ, Woodside MT, Block SM (2007), *High-resolution, single-molecule measurements of biomolecular motion*, Annu Rev Biophys Biomol Struct. **36**:171-90.
- Guck J, Schinkinger S, Lincoln B, Wottawah F, Ebert S, Romeyke M, Lenz D, Erickson HM, Ananthakrishnan R, Mitchell D, Käs J, Ulvick S, Bilby C (2005), *Optical Deformability as an Inherent Cell Marker for Testing Malignant Transformation and Metastatic Competence*, Biophys. J. **88** :3689–98.
- Harada Y, Asakura T (1996), *Radiation forces on a dielectric sphere in the Rayleigh scattering regime*, Opt. Commun. **124**:529-41.
- Happel J, Brenner H (1983), *Low Reynolds Number Hydrodynamics: with special applications to particulate media*, Springer, ISBN-10: 9024728770
- Jackson, JD (1999), *Classical Electrodynamics*, John Wiley & Sons, New York, 3rd Ed., Chap **9, 10**. ISBN-10: 047130932X
- Jesacher A, Fuhpater S, Bernet S, Ritsch-Marte M (2004), *Size selective trapping with optical "cogwheel" tweezers*, Opt. Express **12**:4129-35.
- Kerker, M (1969), *The Scattering of Light and other Electromagnetic Radiation*, Academic Press, New York, (p.33-7). ISBN-10: 0124045502.
- Moffitt JR, Chemla YR, Smith SB, Bustamante C (2008), *Recent advances in optical tweezers*, Annu. Rev. Biochem. **77**:205-28.

- Neuman KC, Chadd EH, Liou GF, Bergman K, Block SM (1999), *Characterization of Photodamage to Escherichia coli in Optical Traps*, Biophys. J. **77**:2856-63.
- Neuman KC, Block SM (2004), *Optical trapping*, Rev. Sci. Instrum. **75**:2787-809.
- Neuman KC, Nagy A (2008), *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy*, Nature Methods **5**:491-505.
- Noom MC, van den Broek B, van Mameren J, Wuite GJL (2007), *Visualizing single DNA-bound proteins using DNA as a scanning probe*, Nature Methods **4**:1031-6.
- Perkins TT (2009), *Optical traps for single molecule biophysics: a primer*, Laser, Photon. Rev. **3**:203-20.
- Rohrbach A (2005), *Stiffness of optical traps: Quantitative agreement between experiment and electromagnetic theory*, Phys. Rev. Lett. **95**:168102.
- Sakai K, Noda S (2007), *Optical trapping of metal particles in doughnut-shaped beam emitted by photonic-crystal laser*, Electron. Lett. **43**:107-108.
- Shaevitz JW (2006), *A Practical Guide to Optical Trapping*, unpublished.
- Snook RD, Harvey TJ, Faria EC, Gardner P (2009), *Raman tweezers and their application to the study of singly trapped eukaryotic cells*, Integr. Biol. **1**:43-52
- Svoboda K, Block SM (1994), *Biological Applications of Optical Forces*, Annu Rev Biophys Biomol Struct. **23**:247-85.
- Valentine MT, Guydosh NR, Gutiérrez-Medina B, Fehr AN, Andreasson JO, Block SM (2008), *Precision steering of an optical trap by electro-optic deflection*, Opt. Lett. **33**:599-601.
- Woodside MT, Valentine MT (2009), *Single molecule manipulation using optical traps*, Handbook of Single-Molecule Biophysics XXIV, Chap **12**, ISBN: 978-0-387-76496-2.

### Chapter 3: Riboswitches: a new gene regulatory element

As described briefly earlier, riboswitches are a gene regulatory element composed solely of RNA, in which changes in the RNA structure act to regulate gene expression. They are a non-coding part of the messenger RNA (mRNA) of the gene, typically found at the 5' end of the UTR ("upstream" of the gene), although sometimes on the 3' ("downstream") end of the gene (Wachtner 2007). The riboswitch sequence folds into alternate three-dimensional structures, where one of these structures stabilised by the binding of a target ligand (typically, a small molecule). Riboswitches are therefore an example of regulation in *cis*: the regulating molecule is the same one being regulated. With the conformational change stabilised by ligand binding, the expression of the gene is facilitated or attenuated by the *cis*-acting element (Tucker 2005).

Riboswitches were discovered in the first few years of the century, as a result of unsuccessful attempts to find "missing" regulatory proteins (Johansen 2003, Mandal 2004, Serganov 2007). Indeed, the first experimental evidence for this regulatory RNA was not obtained until 2002, independently by the research groups of Breaker and Nudler (Johansen 2003, Mandal 2003). A classic example of regulation via protein is the lac repressor, a protein that binds directly to DNA of the gene, in order to inhibit expression (Berg 2007, Wilson 2007). Although differing in mechanism from the protein regulation model, the riboswitch presents an effective way to regulate a gene pathway without DNA binding.

Since the existence of riboswitches was first confirmed, the number identified has expanded rapidly, driven largely by bioinformatic genomic searches

(Sudarsan 2006). Although the first riboswitches were discovered in prokaryotes, riboswitch candidates have been discovered in all three domains of the phylogenetic tree (Barrick 2007). Examples of riboswitches found in eukaryotes include thiamine-pyrophosphate (TPP) sensing riboswitches in fungi (e.g. genus *Aspergillus* - 3' UTR region, *Neurospora* - 5') (Mandal 2003, Cheah 2007) and plants (*Arabidopsis*, *Oryza*, *Poa*, in the *THIC* gene) (Wachter 2007, Bocobza 2007). Most known riboswitches are prokaryotic, however, and research has therefore been focused on bacterial riboswitch systems. These have often been found to regulate genes for the synthesis or transport of metabolites (Fig3.1a), and as such they may be a good target for designer antibiotics (Serganov 2009, Blount 2006, 2007).

Riboswitches can be described as “nanobiosensors:” the structure of most can be divided neatly into two parts, a sensor domain (the aptamer) that detects the signal and an effector domain (the expression platform) that acts on it. Upon ligand binding, the aptamer changes structure in a way that causes the expression platform to change structure also, turning the gene on or off. Interestingly, riboswitch aptamers are considerably more specific than the artificial binding elements created through *in vitro* RNA evolution (SELEX) (Stoltenburg 2007) after which they were named (Kim 2008, Johansen 2003). Aptamer domain structures are highly conserved in both sequence and structure; in fact the most genetically conserved part of riboswitch sequences is the aptamer domain.

Riboswitch aptamers typically involve a network of essential tertiary interactions (unlike artificial aptamers), which help them achieve high sensitivity and

specificity. In contrast, the expression platform has much greater variability in sequence, structure and mechanism. It may affect *transcription* by forming a terminator or anti-terminator hairpin loop, or affect *translation* by sequestering or releasing the Shine-Dalgarno sequence onto which the ribosomal binds in bacteria (Kim 2008). Other mechanisms by which the expression platform may modulate gene expression are self-splicing of the mRNA and (in eukaryotes) modified pre-mRNA splicing (Wachter 2007).

Before classifying riboswitches by their structure, a noteworthy generalisation can be made about the riboswitches in prokaryotes: riboswitches in Gram-negative bacteria tend to regulate translation initiation, while those in Gram-positive bacteria predominantly regulate transcription termination. This is thought to be due to the fact that genes in Gram-positive bacteria cluster in larger operons, (Fig3.1b) (Nudler 2004).

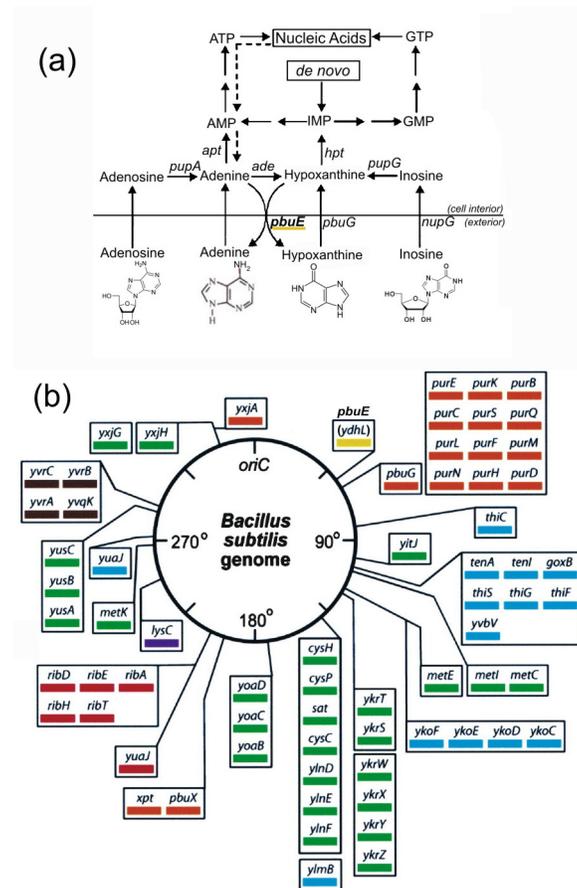


Figure 3.1 Metabolism and genome of *B. subtilis*. (a) Select purine pumping pathways . The *pbuE* pump exports adenine and hypoxanthine from the cytoplasm of the cell. (Nygaard 2005). (b) Diagram of the *B. subtilis* genome. The two efflux pumps, *pbuE* and *pbuG* are in close proximity on the *B. subtilis* genome. Overall, 2% of *B. subtilis*' genes are modulated by riboswitches (Mandal 2003).

### 3.1 Types of riboswitches

In addition to acting through diverse mechanisms, riboswitches target many different types of ligands. These range from purines, to amino acids, metabolic sugars, and vitamins or their derivatives. There is also notable variety in the structures of riboswitches: while some aptamers have structures that are only stabilised by ligand binding, others are stable independently. Some

riboswitches also involve multiple aptamers, whereas others actually act in tandem (Sudarsan 2006).

A structural classification scheme published recently divides riboswitches into two principal types, based on how ligand binding is related to aptamer structure (Montange 2008). In this scheme (Fig 3.2), riboswitches having a largely preorganised tertiary structure supporting a single, localised binding pocket are called “type I”, whereas those having a bipartite binding pocket brought and held together by ligand-induced tertiary structure are considered “type II.” Examples of type I riboswitches include the purine riboswitches (the focus of my work), the preQ<sub>1</sub> riboswitch (notable for being very small), the SAM-II riboswitch (noteworthy for its inclusion of a pseudoknot), and the *glmS* riboswitch ribozyme (in which ligand binding induces enzymatic self cleavage), (Mandal 2004). The thiamine pyrophosphate (TPP) riboswitch is a good example of the second type (Montange 2008). The global structure of the RNA is essentially formed before binding for type I aptamers, whereas binding must occur before a type II aptamer folds into its characteristic structure. For type I riboswitches, the binding pocket has access to multiple conformations in the absence of ligand binding, which may or may not be conducive to productive binding. Interestingly, this two-type classification scheme can be applied equally well to ribozymes: for example, the hammerhead ribozyme is structured much like the purine riboswitch, and the HδV ribozyme closely resembles the *glmS* riboswitch (Montange 2008).

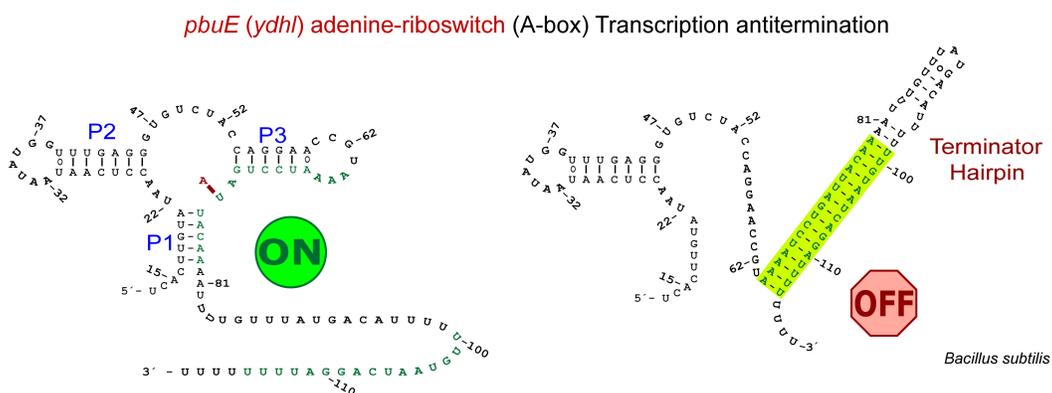
	Active Site Organisation	Core 2° structure	Peripheral 2° structure	Classes	
RNA	Type I	3-way junction	Terminal-loop(TL)+TL	Purine riboswitches (RS)	
			TL+ Internal Loop (IL)	Hammerhead ribozyme (RZ)	
		Loops	TL	Tobramycin aptamer	
			IL	Theo aptamer, ATP aptamer	
		pseudoknots	single	SAM-II RS	
			double	GImS RZ/RS	
				HδV RZ	
				CPEB RZ	
		Type II	3-way junction		S15 domain from 30S rRNA
			Helix (H) + H	5-way junction	Group-I intron
TL + H	3-way junction		M-box (magnesium) RS		
IL + H	4-way junction		SAM-I RS		
IL + IL	4-way junction		hairpin RZ		
	3-way junction		TPP RS		

**Figure 3.2 Riboswitch classification by structure (Montange 2008).** Of particular interest in this thesis are the purine riboswitches, classed as type I, with a three-way junction and two interacting terminal (hairpin) loops. The purine riboswitches and the hammerhead ribozyme are also closely related, structurally.

Looking at the different types of ligands known to be bound by riboswitches, examples include glycine, lysine, purines, vitamin B<sub>12</sub>, TPP, flavin mononucleotide (FMN), S-adenosylmethionine (SAM), preQ<sub>1</sub>, glucosamine-6-phosphate (GlcN6P), and magnesium. For several of these ligands, such as SAM, preQ<sub>1</sub>, and magnesium, multiple different classes of riboswitches with distinct behaviours are known to exist. There are also riboswitch candidates such as: S-adenosylhomocysteine (SAH), and molybdenum cofactor (Moco), as well as candidate for which the ligands are not yet known (Montange 2008).

### 3.2 The purine riboswitches: structure and function

The focus of my research in this thesis has been on the *pbuE* (formely *ydhL*) adenine riboswitch (Fig3.3) found in *Bacillus subtilis*. It is one of at least four types of purine riboswitch, all of which are closely related by their conserved sequences and secondary structures (Fig3.4), and by the fact that they are found upstream of a gene involved in purine synthesis or transport (Kim 2008). The purine riboswitches as a class are among the most studied, partly because they were discovered early on and partly due to their simple structure. They were in fact the first riboswitches to have their structure measured by x-ray crystallography, providing key insights into how structure affected function.



**Figure 3.3** Secondary structure of the *pbuE* riboswitch. Secondary structure predictions by programs such as *mfold*, show that the transcription terminator is the thermodynamically favoured structure (Serganov 2004).

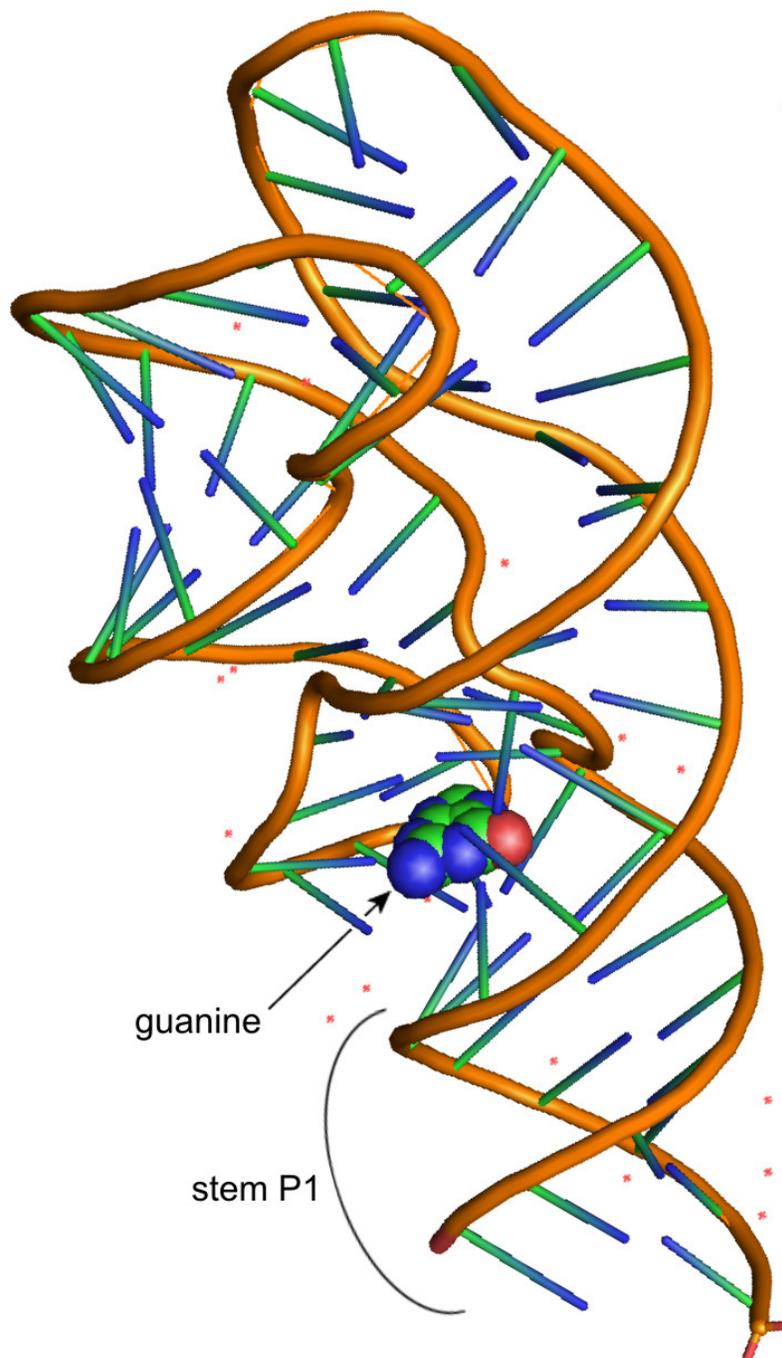


As seen in the structures of the *add* adenine riboswitch aptamer (Serganov 2004) and the *xpt-pbuX* guanine riboswitch aptamer (Serganov 2004, Batey 2004) in figures 3.4 and 3.5, the aptamer of a purine riboswitch has a shape that has been likened to a tuning fork. There are three helices, two of which (P2 and P3) are aligned side-by-side on top of the third (P1). The two parallel helices are held together by tertiary interactions. A kissing interaction between the hairpin loops L2 and L3 at the ends of helices P2 and P3 stabilises the alignment of these two helices (Batey 2004), helping to ensure that a three-helix junction is formed correctly with P1. This junction is crucial to the function of the riboswitch because it serves as the binding site for the ligand; it is also stabilised by a number of hydrogen bonding and stacking interactions, in the presence of the ligand.

With regards to the tuning-fork structure, the kissing interaction between loops L2 and L3 turns out to be essential for ligand binding, even though it is well away from the ligand binding pocket, because of the way it organises the aptamer structure globally to make binding possible. NMR and single molecule FRET (smFRET) studies have demonstrated that the three-helix junction forms independently of ligand binding, in an  $Mg^{2+}$  dependent way. However, under these circumstances it is unstructured and dynamically disordered because of a lack of stabilising hydrogen bonds in the junction, which allows it to envelop the target ligand. The interaction between the loops L2 and L3 thus “pre-organises” the binding pocket, and the formation of the loop-loop interaction is a key intermediate state occurring before ligand binding (Lemay 2006, Noeske 2007).

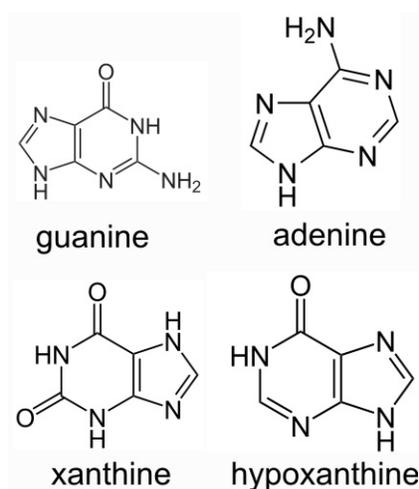
The importance of these loops is underscored by the fact that they are highly conserved phylogenetically: a mutation of any of these participating bases would disrupt the hydrogen bonding network and hence the formation of the junction. However, the aptamers seem to be insensitive to the precise details of the loop-loop interaction, as paired mutations in L2 and L3 that preserve the hydrogen bonding network do not impair the function of the aptamer (Lemay 2006). Moreover, the loop-loop interaction is not essential for the specificity of the binding pocket (Lemay 2006).

From the crystal structures and mutagenesis studies, it is clear that there is a single primary binding residue that forms a Watson-Crick basepair with the ligand in the binding pocket: cytidine for a guanine ligand, and uridine for an adenine ligand. In fact, a guanine riboswitch aptamer can be turned into an adenine riboswitch aptamer—and vice versa—simply by switching the identity of this residue. Despite the clear importance of this specific residue, however, the purine ligand also participates in many other interactions with the RNA: for example, in the *xpt* riboswitch as many as 7 hydrogen bonds are formed with the guanine ligand (Mandal 2004). The binding pocket almost completely envelopes its cognate ligand (Kim 2008, Serganov 2004), making the ligand virtually inaccessible to bulk solvent (Batey 2004). This burial of the ligand in the binding pocket helps clarify the importance of binding pocket pre-organisation: since the ligand must be able to enter the pocket but the pocket is not accessible when fully-formed, clearly the pocket must be disordered before binding, and a substantial conformation change occurs upon binding (Fig3.5).



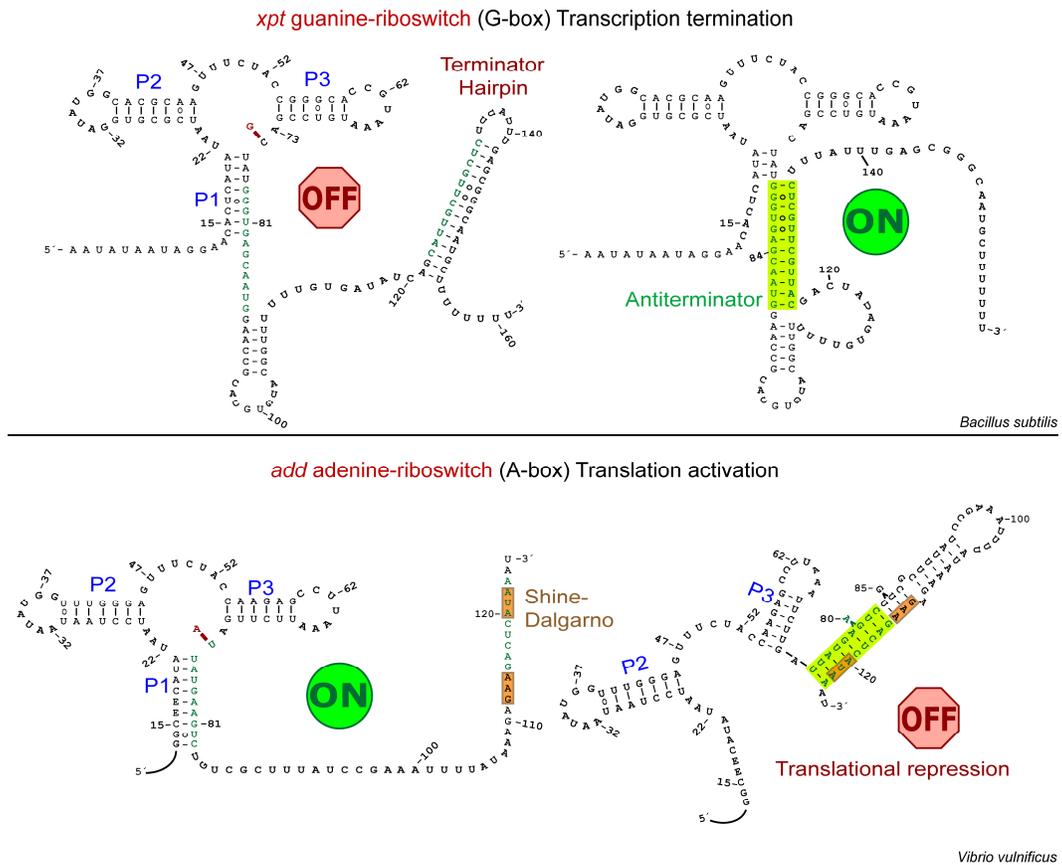
**Figure 3.5** The crystal structure of the *xpt* riboswitch aptamer (PDB ID 1Y27). The kissing loops at the top and the binding pocket surrounding the ligand can be clearly seen (compare to Fig 3.4a or 3.7).

The selectivity of the purine riboswitches is very high, permitting them to distinguish between purine nucleosides and nucleotides, purine bases, and analogues. Quantitatively in the case of the *xpt* guanine riboswitch, the  $K_D$  for binding guanine (the correct ligand) is 6 orders of magnitude lower than it is for adenine, despite the small structural difference between the two ligands (Mandal 2003). The  $K_D$  for guanine binding is 5nM (Kim 2008), while that for guanine analogues hypoxanthine and xanthine have a  $K_D$  of  $\sim 50$ nM (Serganov 2004) (Fig3.6). This selectivity is likely due to the dense network of hydrogen bonds formed between the binding pocket and the ligand: almost every recognisable feature of the ligand is involved (Serganov 2004, Mandal 2004).



**Figure 3.6 Purine structures: guanine, adenine, xanthine and hypoxanthine. Xanthine and hypoxanthine are biologically relevant analogs of guanine.**

Turning to how the purine riboswitches function, we find that the expression platform of the *xpt* G-riboswitch controls expression at the level of transcription. In the absence of the guanine ligand, an anti-terminator hairpin structure prevents transcription termination, but once the aptamer is stabilised by the ligand, a terminator hairpin is formed to stop transcription (Fig3.7) (Serganov 2004). The *pbuE* riboswitch, on the other hand, while it still controls expression at the level of transcription, functions instead as a transcription activator upon ligand binding: without its adenine ligand, a terminator hairpin is formed (Fig3.3), but with ligand bound to the aptamer the expression platform is unstructured (Mandal 2004). Yet another variation is seen with the *V. vulnificus add* adenine riboswitch, which is thought to act at the level of translation, due to the presence of unpaired Shine-Dalgarno (GAA) and initiation codon sequences to serve as a ribosomal binding site (RBS) (Fig3.7). In the absence of the ligand, the RBS is sequestered in a hairpin, preventing translation (Serganov 2004).



**Figure 3.7 Structure and function of the *xpt* and *add* riboswitches.** These are two purine riboswitches for which the crystal structures are available, and are invaluable for comparison with the *pbuE* riboswitch. The *xpt* riboswitch forms a hairpin in the presence of guanine that terminates transcription, whereas the *add* riboswitch sequesters the Shine Dalgarno (SD) sequence to prevent ribosome assembly on the mRNA in the absence of adenine (Serganov 2004).

### 3.3 The *pbuE* adenine riboswitch

The *pbuE* gene encodes for a system that pumps out adenine and its purine analogs (Fig3.1a), and thus contributes in the reduction of the uptake and utilisation of purine compounds (Nygaard 2005). This is important since some purines, such as guanine, may come out of solution at moderate concentrations (Mandal 2004). *PbuE* works in conjunction with several other genes, including *pbuG*, *nupG*, *pupA*, *pupG* (Fig3.1b), to modulate the metabolic pathways involved in nucleic acid synthesis by adjusting the cellular levels of free purine bases, and to protect against toxic purine analogs (Johansen 2003, Nygaard 2005). By turning on in the presence of adenine, the riboswitch controls the adenine level in the form of a negative feedback loop: the presence of high levels of adenine triggers the synthesis of an efflux pump to lower the cellular adenine concentration whereas the lack of sufficient adenine prevents any pumps from being made thereby allowing levels to increase.

Despite the wealth of structural information about purine riboswitches, no structure yet exists for the *pbuE* riboswitch, as experimentalists have been unable to crystallise it. However, the similarity between the secondary structure and tertiary interactions between all the purine riboswitches allows the conclusions from studies of other purine riboswitches to be carried over to the *pbuE* riboswitch, (Fig3.8).



**Figure 3.8** The crystal structure of the *add* riboswitch aptamer (PDB ID 1Y26). The tuning fork structure, kissing loops, and binding pocket at the three-stem junction are very similar to the *xpt* riboswitch (compare to Fig 3.4b or 3.7).

One very interesting structural detail about the *pbuE* riboswitch is that the terminator hairpin can be shown to be more stable than the aptamer with which it is in structural competition: the terminator (or “off”) state is therefore thermodynamically preferred, regardless of whether ligand is bound to the aptamer (Lemay 2006). Thermodynamically, then, this riboswitch should not act as a switch: once fully synthesised and under equilibrium conditions, the riboswitch would form only the ‘off’ or terminating state (Fig 3.3). However, this ignores the fact that transcription is a dynamic, non-equilibrium process in which kinetics play a key role. The aptamer is polymerised by RNAP first, being upstream of the terminator. Since the terminator sequence must be fully synthesised before the terminator hairpin can form, the aptamer has a brief window in which to bind the ligand, prevent the formation of the terminator, and permit transcription to continue. Since transcription in prokaryotes typically occurs at a rate of  $\sim 50\text{-}100$  nt/s<sup>10</sup> (Berg 2007) and the expression platform consists of  $\sim 40$  nt, this provides a window of a bit less than 1 s for binding and folding.

This feature of *pbuE* riboswitch behaviour highlights the competition between thermodynamic and kinetic considerations (the most stable structure versus the structure that forms most rapidly), and the importance of the fact that folding occurs co-transcriptionally. Due to the fact that the nucleotides are

---

<sup>10</sup> By comparison, transcription at the ribosome occurs at about 20 amino acids per second.

synthesised in a linear chain, but the interactions between nucleotides that underpin the structure often occur over large distances along the chain, co-transcriptional folding may very well permit non-equilibrium, metastable structures to form preferentially. As long as these metastable states have lifetimes long compared to the kinetics of transcription, they are stable enough to affect the gene regulation in a riboswitch.

The *pbuE* riboswitch is not the only one known to have a folding outcome determined by kinetic competition between folding and transcription: the *ribD* FMN riboswitch (in *B. subtilis*) is known to behave similarly (Wickiser 2005, 2009). In contrast, the *add* adenine riboswitch from *V. vulnificus* is believed to be dominated by thermodynamics, with adenine binding able to occur any time after transcription of the expression platform (Kim 2008). It is notable when considering kinetic control of the folding that the speed of transcription may dictate the concentration of ligand required for the riboswitch to take effect, rather than simply the binding affinity of the aptamer.

As a final point, since most folding measurements are based on pre-synthesised RNA, it has proven very challenging to study how transcription changes folding, and much remains to be learned. Riboswitches provide an excellent forum for such studies, and single-molecule measurements especially may help to probe co-transcriptional folding.

### 3.4 References

- Barrick JE, Breaker RR (2007), *The distributions, mechanisms, and structures of metabolite-binding riboswitches*, *Genome Biol.* **8**:R239.
- 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**:411–5
- Berg JM, Tymoczko JL, Stryer L (2007), *Biochemistry* New York : W. H. Freeman, 6th ed. ISBN: 0716787245.
- Blount KF, Breaker RR (2006), *Riboswitches as antibacterial drug targets*, *Nat. Biotechnol.* **24**:1558-64.
- Blount KF, Wang JX, Lim J, Sudarsan N, Breaker RR (2007), *Antibacterial lysine analogs that target lysine riboswitches*, *Nat Chem Biol.* **3**:44-9.
- Bocobza S, Adato A, Mandel T, Shapira M, Nudler E, Aharoni A (2007), *Riboswitch-dependent gene regulation and its evolution in the plant kingdom*, *Genes Dev.* **21**: 2874–9.
- Cheah MT, Wachter A, Sudarsan N, Breaker RR (2007), *Control of alternative RNA splicing and gene expression by eukaryotic riboswitches*, *Nature* **447**: 497–500.
- Hendrix DK, Brenner SE, Holbrook SR (2005), *RNA structural motifs: building blocks of a modular biomolecules*, *Q Rev Biophys.* **38**:221-43.
- Johansen LE, Nygaard P, Lassen C, Agersø Y, Saxild HH (2003), *Definition of a second Bacillus subtilis pur regulon comprising the pur and xpt-pbuX operons plus pbuG, nupG (yxjA), and pbuE (ydhL)*, *J Bacteriol.* **185**:5200-9.
- Kim JN, Breaker RR (2008), *Purine sensing by riboswitches*, *Biol Cell.* **100**:1-11.
- Lemay JF, Penedo JC, Tremblay R, Lilley DM, Lafontaine DA (2006), *Folding of the adenine riboswitch*, *Chem Biol.* **13**:857-68.
- Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR (2003), *Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria*, *Cell* **113**:577-86.
- Mandal M, Breaker RR (2004), *Adenine riboswitches and gene activation by disruption of a transcription terminator*, *Nat Struct Mol Biol.* **11**:29-35.
- Montange RK, Batey RT (2008), *Riboswitches: emerging themes in RNA structure and function*, *Annu Rev Biophys.* **37**:117-33.
- Noeske J, Buck J, Fürtig B, Nasiri HR, Schwalbe H, 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**:572-83.

- Nudler E, Mironov AS (2004), *The riboswitch control of bacterial metabolism*, Trends Biochem Sci. **29**:11-7.
- Nygaard P, Saxild HH (2005), *The purine efflux pump PbuE in Bacillus subtilis modulates expression of the PurR and G-box (XptR) regulons by adjusting the purine base pool size*, J Bacteriol. **187**:791-4.
- Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ (2004), *Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs*, Chem Biol. **11**:1729-41.
- Serganov A, Patel DJ (2007), *Ribozymes, riboswitches and beyond: regulation of gene expression without proteins*, Nat Rev Genet. **8**:776-90
- Serganov A, Huang L, Patel DJ (2009), *Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch*, Nature **458**:233-7.
- Stoltenburg R, Reinemann C, Strehlitz B (2007), *SELEX-a (r)evolutionary method to generate high-affinity nucleic acid ligands*, Biomol Eng. **24**:381-403
- Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, Breaker RR (2006), *Tandem riboswitch architectures exhibit complex gene control functions*, Science **314**:300-4.
- Tucker BJ, Breaker RR (2005), *Riboswitches as versatile gene control elements*, Curr Opin Struct Biol. **15**:342-8.
- Wachter A, Tunc-Ozdemir M, Grove BC, Green PJ, Shintani DK, Breaker RR (2007), *Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs*, Plant Cell **19**:3437-50.
- Wickiser JK, Winkler WC, Breaker RR, Crothers DM (2005), *The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch*, Mol Cell **18**:49-60.
- Wickiser JK (2009), *Kinetics of riboswitch regulation studied by in vitro transcription*, Methods Mol Biol. **540**:53-63.
- Wilson CJ, Zhan H, Swint-Kruse L, Matthews KS (2007), *The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding*, Cell. Mol. Life Sci. **64**:3-16

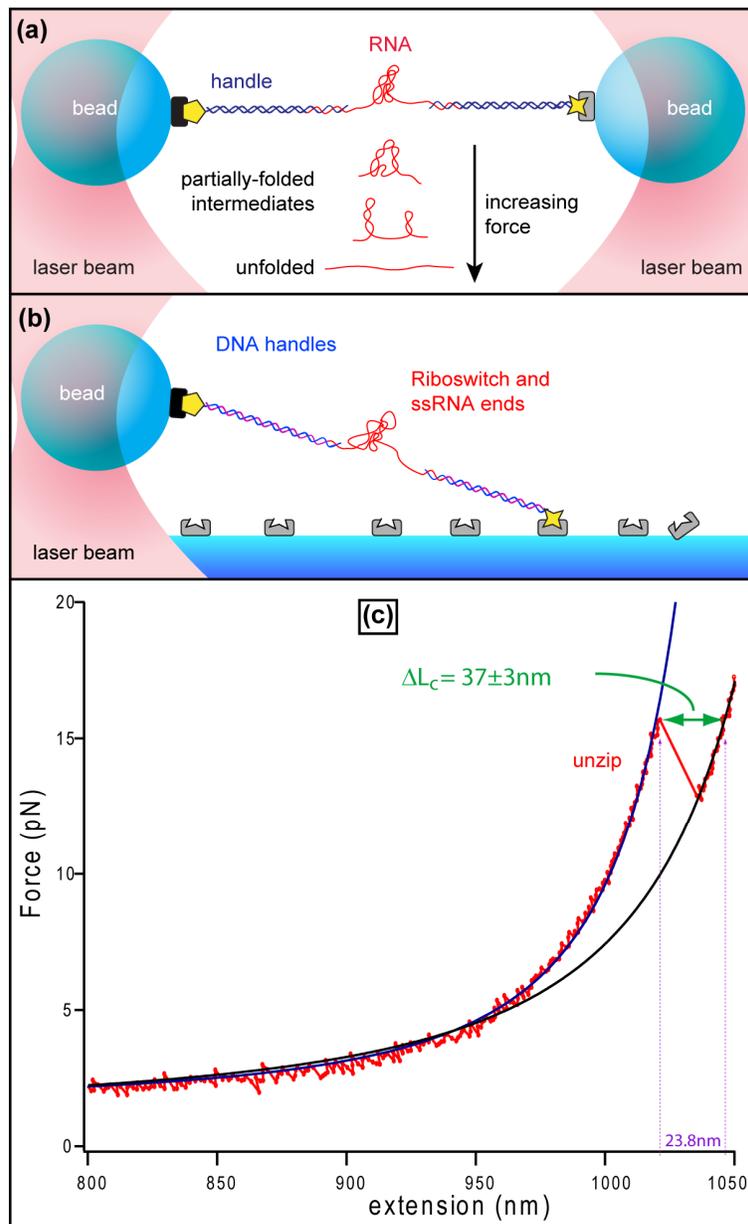
## Chapter 4: Optical trapping measurements

Single molecule force spectroscopy involves measuring the change in extension in a molecule as it folds or unfolds under the effect of a force applied to the molecule. In these measurements, force acts as the “denaturant” that causes the conformation of the molecule to change: molecules that spontaneously fold into their native structure can be unfolded by applying a large enough force, whereas molecules that have been unfolded under tension can be refolded by lowering the force. 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 two different measurement paradigms. First is the dynamic force measurement, also known as the force-extension curve, in which the molecular extension is monitored while changing the force applied to the molecule. The second is the force clamp measurement, in which the extension is measured while maintaining a constant force on the molecule. These two types of measurements formally represent different ensembles, and provide complementary information (Manosas 2005, Dudko 2006, 2008, Kreuzer 2001).

### 4.1 Dynamic force measurements

The dynamic pulling experiment typically involves pulling on the molecule with a linearly increasing applied force, and provides a plot of the force

as a function of the extension, (force-extension curve or FEC). As the force increases, structures in the molecule will be pulled apart (Fig 4.1a), often (though not always) from the least stable to the most stable, if they are accessible by tension. 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 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, as illustrated (Fig 4.1c). Note that FECs reflect all elastically compliant components in the experimental setup, including not just the molecule under study (such as the riboswitch RNA in Fig 4.1c) but also the trap stiffness and the stiffness of the dsDNA 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.



**Figure 4.1** Force extension curve assay. (a) Dual trap dumbbell assay (adapted from Woodside 2008). RNA (red) is attached to ‘handles’ of dsDNA (blue) connected by antibody-antigen links to trapped beads. As tension is increased, the RNA unfolds in steps reflecting its structure. (b) A functionalised surface may be used in the place of a second trapped bead. (c) Characteristic FEC from a *pbuE* riboswitch aptamer, measured immediately after synthesis by the RNAP, showing a single unfolding event or ‘unzip’. WLC fits for the handles alone (blue), and handles plus unfolded ssRNA (black) are overlaid on the data (red). The change in contour length found from the fits,  $\Delta L_c$ , is significantly longer than the apparent extension change at the unzip.

#### ***4.1.1 Elastic properties of nucleic acids***

Fortunately, the elastic behaviour of nucleic acids has been well characterised, so the contributions of both the dsDNA handles and unfolded ssRNA can be calculated. Indeed, this is a principal reason for using dsDNA handles in the experiment. The elasticity of dsDNA, for instance, can be modeled by a variant of the worm-like chain (WLC) (Bustamante 1994, Smith 1996, Marko 1995). WLC models typically calculate the entropic elasticity by treating 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 1997). ssRNA has also been found to be well-described by such a modified WLC model, though with quite different persistence length:  $\sim 1$  nm for ssRNA vs  $\sim 50$  nm for dsDNA (Bustamante 1994, Seol 2004). Combinations of dsDNA and ssRNA 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 (Fig 4.1c). 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). For the enthalpically modified case, this takes the form:

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

where  $L_p$  is the persistence length,  $L_c$  the contour length, and  $K$  the elastic modulus of the dsDNA (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 (Wang 1997, Marko 1995).

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  $K \sim 1000$ -1200 pN (Wang 1997, Smith 1996, Bouchiat 1999, Baumann 2000), whereas for ssRNA,  $L_p \sim 1$  nm and  $K \sim 1500$ -1600 pN (Seol 2004, 2007). For dsRNA, which takes on an A-form helix in contrast to the B-form helix of dsDNA (Saenger 1984, Neidle 1999),  $L_p \sim 60$  nm (Abels 2005); DNA-RNA hybrids, which also take on A-form helices, are expected to have a similar  $L_p$ . Interestingly, ssDNA is not as well described by a WLC model; instead, modified freely-jointed chain models or more complex models are better choices (Smith 1996, Dessinges 2002, Saleh 2009). 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).

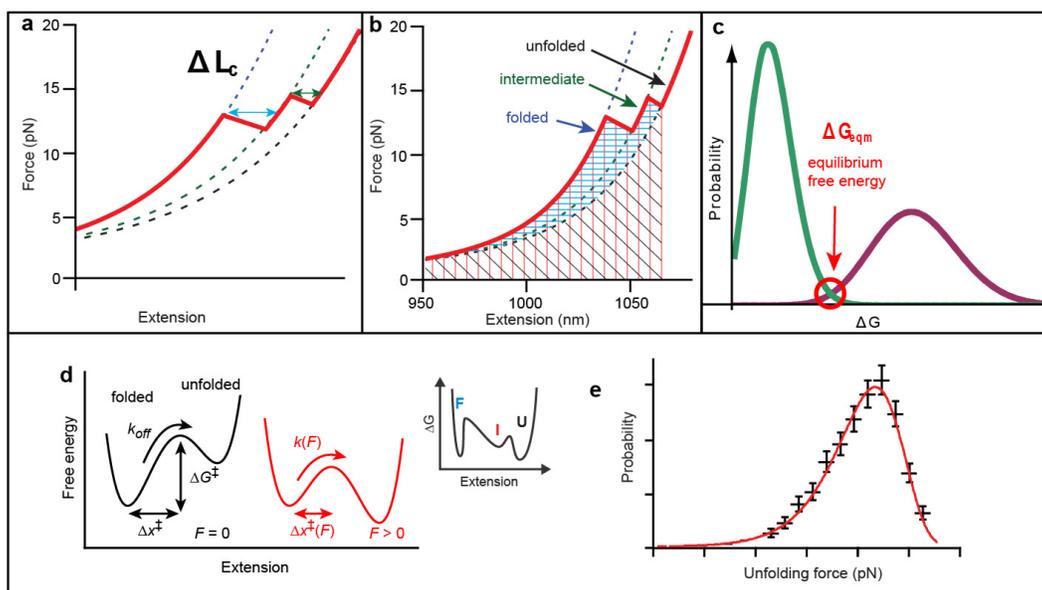
#### 4.1.2 Analysis of FECs

By fitting an FEC on either side of an unfolding event, as in Fig 4.2a, the change in contour length associated with a particular event can be determined, and hence the number of nucleotides that unfolded. However, there is much more information that can be extracted from FECs. For example, the integral of the FEC curve (the work done during the experiment) should yield the work required to unfold the molecule (Fig 4.2b). This work is related to the equilibrium folding free energy for the molecule, although not necessarily in a simple way. The force is being changed rapidly, the system is not necessarily in equilibrium, and work may therefore be dissipated. In addition, the work done on to stretch out the handles and the ssRNA liberated by the unfolding must also be accounted for (Woodside 2008).

In order to account for the work dissipated during the measurement, the Jarzynski equality (Jarzynski 1997) may be used to determine the equilibrium free energy from the distribution of non-equilibrium work measurements:

$$\Delta G_{eqm} = -k_B T \ln \left\langle \exp \left( -\frac{W}{k_B T} \right) \right\rangle \quad 4-2$$

Here the equilibrium free energy,  $\Delta G_{eqm}$ , is found from an exponentially-weighted average of the measured non-equilibrium works,  $W$  (Fig 4.2c).



**Figure 4.2 FEC analysis.** (a) FEC cartoon showing changes in the contour length  $\Delta L_c$ . (b) To determine the work done to unfold the molecule, the FEC is integrated 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  $\Delta x^\ddagger$ , the location of the barrier along the reaction coordinate, and  $\Delta G^\ddagger$ , 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.

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 (Fig 4.2c):

$$\frac{P_U(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G_{eqm}}{k_B T}\right) \quad \mathbf{4-3}$$

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 (Fig 4.2d). A molecule with two states (folded and unfolded) has a folding landscape consisting of two potential wells separated by a barrier at a distance  $\Delta x_f^\ddagger$  from the folded state (minimum of the first well) and  $\Delta x_u^\ddagger$  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 can be modified to include additional barriers and potential wells (Fig 4.2d inset).

To extract the kinetics and the reaction coordinate distances, a more sophisticated treatment has been expressed analytically by Dudko et al., who in 2006 derived an expression for the probability distribution of unfolding forces (Fig 4.2e),  $p(F)$ , given certain shapes of the energy landscape (Dudko 2006):

$$p(F) \propto \frac{k(F)}{r} \exp \left\{ \frac{k_{off}}{\Delta x^\ddagger r} - \frac{k(F)}{\Delta x^\ddagger r} \left( 1 - \frac{\Delta x^\ddagger F}{\Delta G^\ddagger} \right)^{1-1/\nu} \right\} \quad 4-4$$

where

$$k(F) = k_{off} \left( 1 - \frac{\Delta x^\ddagger F}{\Delta G^\ddagger} \right)^{1/\nu-1} \exp \left\{ \Delta G^\ddagger \left[ 1 - \left( 1 - \frac{\Delta x^\ddagger F}{\Delta G^\ddagger} \right)^{1/\nu} \right] \right\} \quad 4-5$$

$k_{off}$  is the unfolding rate at zero force,  $\Delta x^\ddagger$  is the distance to the transition state (the top of the energy barrier) from the folded state,  $\Delta G^\ddagger$  is the height of the barrier,  $r$  is the loading rate (rate of change of the force), and  $\nu$  describes the shape of the barrier:  $\nu = 1/2$  for a sharp, cusp-like barrier,  $\nu = 2/3$  for a soft, cubic potential. This expression extends a previous treatment (Evans 1997) which effectively had  $\nu = 1$ : it ignored the change in  $\Delta x^\ddagger$  with force and as a consequence systematically overestimated the unfolding rate (Dudko 2006). For a given set of FECs all at the same loading rate, these equations are used to fit the distribution  $p(F)$ , as in Fig 4.2e, and extract the parameters describing the kinetics and the shape of the energy landscape. Of particular interest in this analysis is the parameter  $\Delta x^\ddagger$ , 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.

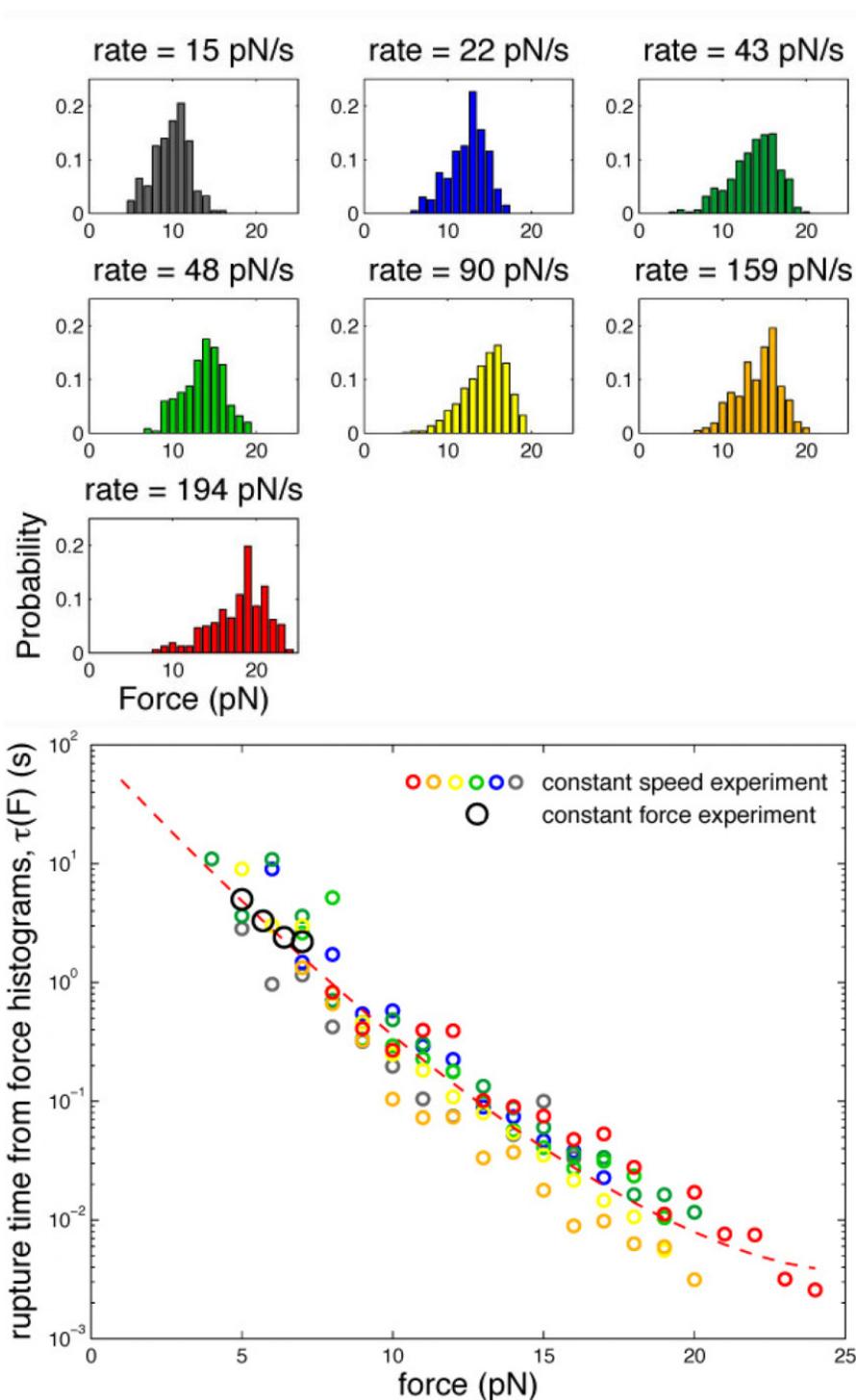
By looking the rupture force distributions at different loading rates, it is possible to extract force dependent lifetimes in a model-independent way (Dudko 2008), since the lifetime at a given force can be expressed as a weighted integral of probability distribution:

$$\tau(F) \approx \int_F^{\infty} p(f) df / [\dot{F}(F)p(F)] \quad 4-6$$

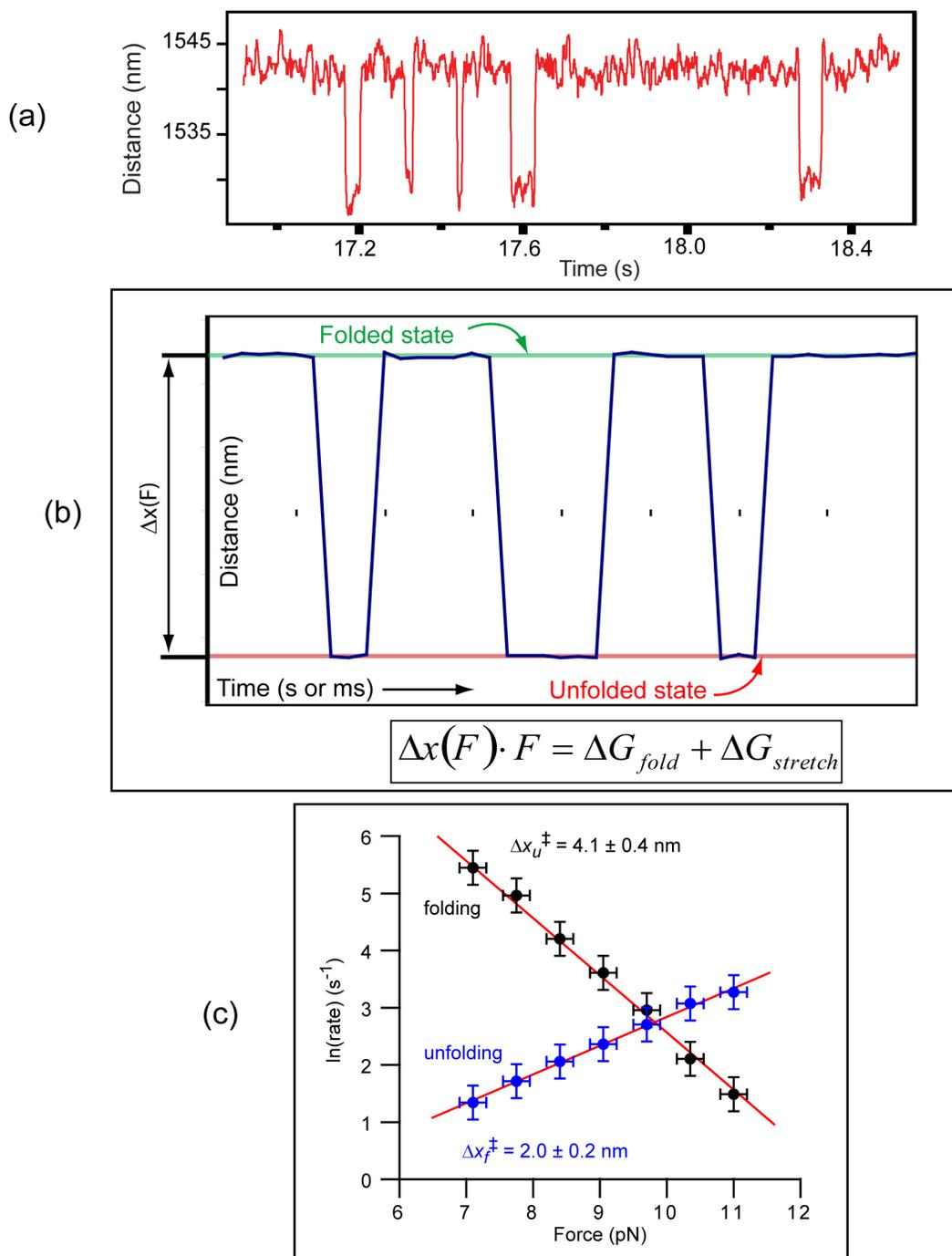
where  $\dot{F}$  is the loading rate. This implies that data from many different loading rates should all collapse onto a single master curve, independent of any assumptions about the shape of the energy landscape. The results of this method can be seen in Fig 4.3 using the data acquired for the *pbuE* aptamer (described in more detail in the following chapter): rates extracted from datasets with loading rates varying over a factor of  $\sim 20$  collapse to the same curve, and are in excellent agreement with the rates measured directly from constant force measurements.

#### 4.2 Constant Force Measurements

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 (Fig4.4a). 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 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 limited bandwidth because of finite feedback loop cycle time, which can introduce measurement artefacts as the feedback loop attempts to catch up to events that are too fast (Manosas 2005, Seol 2009). This can result in spurious rapid transitions and a change in long term dynamics.



**Figure 4.3** Scaling analysis of unfolding rates. The force-dependent unfolding rate is calculated from histograms of unfolding force distributions at different loading rates. All data collapse onto a single master curve, from which zero-force rate can be extracted. Figure courtesy of O.K. Dudko.



**Figure 4.4 Constant force measurements. (a) Extension vs time at constant force for the *pbuE* aptamer. (b) Extension changes and lifetimes can be measured directly from the record. (c) Lifetimes depend on the force applied, seen here for hairpin P2 of the aptamer.**

An alternate way to maintain a constant force is to use 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 is set by adjusting the intensity of the trapping beam (Greenleaf 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 size of this constant-force region is limited to about 50nm, meaning that the observation of larger unfolding distances will require some active feedback from the stage to stay within the desired trapping region.

#### ***4.2.1 Analysis of constant force measurements***

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 2008). Extension changes can be read directly from the data (Fig 4.4b) and converted to contour length changes using a WLC model as before. The folding free energy can be found without the use of fluctuation theorems, simply from the product of force and extension change (although the calculated energy still includes the stretching energy). Most usefully, the distribution of lifetimes at a given force can be measured directly

from the data, without any fitting. As an added benefit, certain structures that are relatively less stable may be easier to see in these equilibrium measurements.

The principal model dependence in the analysis comes from the interpretation of the force dependence of the lifetimes. 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  $\Delta x^\ddagger$ , the distance to the transition state (subscripts  $u$  and  $f$  representing unfolding and folding, respectively, while  $k^0$  is the rate at zero force):

$$\ln \frac{k_u(F)}{k_u^0} = \frac{\Delta x_u^\ddagger}{k_B T} F \quad \text{and} \quad \ln \frac{k_f(F)}{k_f^0} = -\frac{\Delta x_f^\ddagger}{k_B T} F \quad 4-7$$

This exponential force dependence can be seen for the lifetimes measured for the folding of hairpin stem P2 in the *pbuE* riboswitch aptamer (Fig 4.4c). The logarithms of the rates vary linearly with force, indicating that the transition state position is effectively constant over the force range probed here. This is in contrast to the analysis shown in Fig 4.3, where the logarithm of the rupture time (and hence unfolding rate) clearly does *not* vary linearly with force, indicating that the transition state moves significantly as the force is change over the (much larger) force range probed. As these results show, 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  $\sim \Delta F / \zeta$ , the ratio of the range of forces probed to

the curvature of the potential barrier. The curvature of a given energy landscape is difficult to measure and generally not known, but we can estimate the typical transition state shift for a molecule whose landscape has been measured (Woodside 2006a,b): the DNA hairpin 30R50/T4 has a barrier with curvature of  $\sim 2$  pN/nm. Since the folding at constant force is probed over a range of  $\sim 2$  pN for this molecule, the barrier location ought to shift by  $\sim 1$  nm, which is similar to the experimental uncertainty in the transition state location. Of course, the assumption that the transition state location can only be used when confirmed by data such as in Fig. 4.4c.

When studying the *pbuE* aptamer (or indeed any given molecule), the use of both the force-ramp and constant-force methods offers the ability to make complementary measurements on the molecule under different circumstances. FECs involve non-equilibrium unfolding, while constant force measurements measure equilibrium fluctuations. Analysis of the two different types of data allows redundant determination of the parameters characterising the folding. In addition, by looking at different kinetic regimes, the coupling between folding and transcription (for the aptamer at least) can be tested: if the folding is independent of transcription, the structures that correspond to the unfolding in FECs should be the same measured under force-clamps.

### 4.3 References

- Abelsa JA, Moreno-Herreroa F, van der Heijdena T, Dekkera C, Dekker NH (2005), *Single-Molecule Measurements of the Persistence Length of Double-Stranded RNA*, *Biophys. J.* **88**:2737-44.
- Baumann CG, Bloomfield VA, Smith SB, Bustamante C, Wang MD, Block SM (2000), *Stretching of single collapsed DNA molecules*, *Biophys J.* **78**:1965-78.
- Bell GI (1978), *Models for specific adhesion of cells to cells*, *Science* **200**:618-27.
- Bouchiat C, Wang MD, Allemand JF, Strick T, Block SM, Croquette V (1999), *Estimating the persistence length of a worm-like chain molecule from force-extension measurements*, *Biophys. J.* **76**:409-13.
- Bustamante C, Marko JF, Siggia ED, Smith S (1994), *Entropic elasticity of lambda-phage DNA*, *Science* **265**:1599-600.
- Carter AR, Seol Y, Perkins TT (2009), *Precision surface-coupled optical-trapping assay with one-basepair resolution*, *Biophys J.* **96**:2926-34.
- Crooks GE (1999), *Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences*, *Phys Rev E* **60**:2721-26.
- Dessinges MN, Maier B, Zhang Y, Peliti M, Bensimon D, Croquette V (2002), *Stretching single stranded DNA, a model polyelectrolyte*, *Phys Rev Lett.* **89**:248102.
- Dudko OK, Hummer G, Szabo A (2006), *Intrinsic rates and activation free energies from single-molecule pulling experiments*, *Phys Rev Lett* **96**:108101.
- Dudko OK, Hummer G, Szabo A (2008), *Theory, analysis, and interpretation of single-molecule force spectroscopy experiments*, *Proc Natl Acad Sci U S A.* **105**:15755-60.
- Dudko OK (2009), *Single-molecule mechanics: new insights from the escape-over-a-barrier problem*, *Proc Natl Acad Sci U S A.* **106**:8795-6.
- Evans E, Ritchie K (1997), *Dynamic strength of molecular adhesion bonds*, *Biophys J.* **72**:1541-55.
- Greenleaf WJ, Woodside MT, Abbondanzieri EA, Block SM (2005), *Passive all-optical force clamp for high-resolution laser trapping*, *Phys Rev Lett.* **95**:208102.
- Jarzynski C (1997), *Non-equilibrium equality for free energy differences*, *Phys Rev Lett* **78**:2690-2693.
- Kreuzer HJ, Payne SH, Livadaru L (2001), *Stretching a macromolecule in an atomic force microscope: statistical mechanical analysis*, *Biophys J.* **80**:2505-14.
- Manosas M, Ritort F (2005), *Thermodynamic and kinetic aspects of RNA pulling experiments*, *Biophys J.* **8**:3224-42.
- Marko JF, Siggia ED (1995), *Stretching DNA*, *Macromolecules*, **28**:8759.

Neidle S (1999), *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press; 1st ed. ISBN-10: 0198500386

Saenger W (1984), *Principles of Nucleic Acid Structure*, Springer-Verlag, ISBN-10: 0387907610

Saleh OA, McIntosh DB, Pincus P, Ribbeck N (2009), *Nonlinear low-force elasticity of single-stranded DNA molecules*, Phys Rev Lett. **102**:068301.

Seol Y, Skinner GM, Visscher K (2004), *Elastic properties of a single-stranded charged homopolymeric ribonucleotide*, Phys Rev Lett. **93**:118102

Seol Y, Skinner GM, Visscher K, Buhot A, Halperin A (2007), *Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking*, Phys Rev Lett. **98**:158103.

Seol Y, Perkins TT (2009), *Sensitivity Of Dna-hairpins Dynamics To The Mechanism Of Force Feedback Probed Using A Surface-coupled Passive Force Clamp*, Biophys J.; **96**(3) supp.1 p.291a.

Smith SB, Cui Y, Bustamante C (1996), *Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules*, Science **271**:795-9.

Steger G, Hofmann H, Förtsch J, Gross HJ, Randles JW, Sängler HL, Riesner D (1984), *Conformational transitions in viroids and virusoids: comparison of results from energy minimization algorithm and from experimental data*, J Biomol Struct Dyn. **2**:543-71

Vilar JMG, Rubi JM (2008), *Failure of the work-Hamiltonian connection for free-energy calculations*, Phys Rev Lett **100**:020601.

Wang MD, Yin H, Landick R, Gelles J, Block SM (1997), *Stretching DNA with optical tweezers*, Biophys J **72**:1335-1346.

Woodside MT, Behnke-Parks WM, Larizadeh K, Travers K, Herschlag D, Block SM (2006a), *Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins*, Proc Natl Acad Sci U S A. **103**:6190-5.

Woodside MT, Anthony PC, Behnke-Parks WM, Larizadeh K, Herschlag D, Block SM (2006b), *Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid*, Science **314**:1001-4. Erratum (2007) in: Science **315**:766.

Woodside MT, García-García C, Block SM (2008), *Folding and unfolding single RNA molecules under tension*, Curr Opin Chem Biol. **12**:640-6.

Woodside MT, Valentine MT (2009), *Single molecule manipulation using optical traps*, Handbook of Single-Molecule Biophysics XXIV, Chap **12**, ISBN: 978-0-387-76496-2

## Chapter 5: Direct observation of hierarchical folding in single riboswitch aptamers

William J. Greenleaf,<sup>1\*</sup> Kirsten L. Frieda,<sup>2</sup> Daniel A.N. Foster,<sup>4</sup> Michael T. Woodside,<sup>4,5\*†</sup>, and Steven M. Block<sup>1,3†</sup>

<sup>1</sup>Department of Applied Physics, <sup>2</sup>Biophysics Program, and <sup>3</sup>Department of Biological Sciences, Stanford University, Stanford CA, 94305 USA; <sup>4</sup>Department of Physics, University of Alberta, Edmonton AB, T6G 2G7, Canada; <sup>5</sup>National Institute for Nanotechnology, National Research Council of Canada, Edmonton AB, T6G 2M9, Canada.

\* These authors contributed equally to this work

† To whom correspondence should be addressed. Email: sblock@stanford.edu (S.M.B); michael.woodside@nrc.ca (M.T.W.)

### 5.1 Abstract

**Riboswitches regulate genes through structural changes in ligand-binding RNA aptamers. Using an optical-trapping assay based on in situ transcription by a molecule of RNA polymerase, single nascent RNAs containing *pbuE* adenine riboswitch aptamers were unfolded and refolded. Multiple folding states were characterized using both force-extension curves and folding trajectories under constant force by measuring the molecular contour length, kinetics, and energetics with and without adenine. Distinct folding steps correlated with the formation of key secondary or tertiary structures and with ligand binding. Adenine-induced stabilization of the weakest helix in the aptamer, the mechanical switch underlying regulatory action, was observed directly. These results provide an integrated view of hierarchical folding in an aptamer, demonstrating how complex folding can be resolved into constituent parts, and supply further insights into tertiary structure formation.**

## 5.2 Single molecule study of hierarchical folding in the *pbuE* riboswitch

Riboswitches are elements of mRNA that regulate gene expression through ligand-induced changes in mRNA secondary or tertiary structure (1, 2). This regulation is accomplished through the binding of a small metabolite to an aptamer in the 5'-untranslated region of the mRNA, which causes conformational changes altering the expression of downstream genes. Riboswitch-dependent regulatory processes depend crucially on the properties of aptamer folding; the kinetics and thermodynamics of folding are therefore of central importance for understanding function.

Among the simplest riboswitches are those regulating purine metabolism, which have aptamers with “tuning fork” structures (3, 4) that bind ligands at a specific residue in a pocket formed by a three-helix junction. The junction is thought to be pre-organized by numerous tertiary contacts, including interactions between two hairpin loops, but the binding pocket itself is likely stabilized only upon ligand binding (4-10). Ligand binding also stabilizes a nearby helix (3-5), sequestering residues that would otherwise participate in an alternate structure affecting gene expression (*e.g.* terminator or anti-terminator hairpins, ribosome binding sequences). Features such as ligand specificity (6, 11) and its structural basis (6, 7), the rates and energies for ligand binding and dissociation (12), the kinetics of loop-loop formation (10), and the interplay of structural pre-organization and induced fit (7-9) have recently been investigated. These studies, however, focused on isolated steps in folding, typically employing ligand analogs or investigating aptamers from different organisms. Here we obtain, from a single set of measurements, an integrated picture of secondary and tertiary structure formation, as well as ligand binding, in the aptamer of the *pbuE* adenine riboswitch from *B. subtilis*, by observing folding and unfolding trajectories of individual molecules subjected to controlled loads in a high-resolution, dual-trap optical tweezers apparatus (13).

Single-molecule force spectroscopy, which measures the extension of a molecule as it unfolds and refolds under tension, furnishes a tool for probing structural transitions: extension changes can be related to the number of nucleotides involved in folding. Furthermore, the effects of force on reaction equilibria and kinetics allow the shapes of the folding landscapes to be determined in detail (14-16). The complete folding process, starting from a fully-unfolded state (not usually probed in conventional RNA folding studies), can also be observed. This initial configuration is especially relevant to riboswitches, because aptamers fold co-transcriptionally from an initially unstructured state. Because of the tight coupling between folding and transcription, the assay was designed to measure folding of mRNA transcribed in situ (17). A single *E. coli* RNA polymerase (RNAP) molecule, transcriptionally stalled downstream of the promoter region on a DNA template (Fig. 5.1A), was attached to a bead held in one optical trap (Fig. 5.1B). The 29-nucleotide (nt) initial RNA transcript emerging from the RNAP was hybridized to the complementary cohesive end of a 3-kb dsDNA “handle” attached to a bead held in the other trap, creating a “dumbbell” geometry allowing forces to be applied between the RNAP and the 5' end of the RNA (18). Force-extension curves (FECs) showing the molecular extension measured as a function of force as the traps were moved apart at a constant rate, confirmed that this initial transcript was unstructured (Fig. 5.1C).

After constructing the dumbbells, transcription was restarted by introducing nucleoside triphosphates. The DNA template coded for the *pbuE* adenine riboswitch aptamer downstream of the initial transcript (Fig. 5.1A). Once the aptamer sequence was transcribed, RNAP was prevented from further elongation by a roadblock consisting of a streptavidin molecule bound to a 5'-terminal biotin label on the template (Fig. 5.1D). FECs measured immediately after aptamer transcription (Fig. 5.1E) revealed a characteristic series of sawtooth features that arise from contour length increases as specific structural elements

unfold (19). In the absence of adenine, two small unfolding events were typically observed (Fig. 5.1E, black). These features are produced by the unfolding of the two stable hairpins in the secondary structure, P3 and P2 (inset, Fig. 5.1E). The interactions that underpin tertiary structure by holding these hairpin loops together and structuring the binding pocket in the triple-helix junction are present only transiently in the absence of adenine (5, 8, 10). The contour length changes associated with these features,  $17 \pm 2$  nt (P3) and  $22 \pm 2$  nt (P2), are consistent with the values expected for these hairpins (19 and 21 nt, respectively) (18). In the presence of adenine, some FECs were identical to those observed in its absence, indicating in these cases that adenine was not bound to the aptamer. More commonly however, larger unfolding distances at higher forces were observed, corresponding to adenine-induced stabilization of the folded structure. In the latter case, the aptamer usually unfolded cooperatively in a single event (Fig. 5.1E, blue), but sometimes through an intermediate state (Fig. 5.1E, red).

Unfolding from the fully-folded state was analyzed in more detail by collecting multiple FECs from the same molecule. Overlaying 800 FECs shows that the aptamer unfolds over a wide distribution of forces (Fig. 5.2A), as expected for a non-equilibrium measurement (20). Three states were clearly seen: the folded and unfolded states, and an intermediate state. We fit the FECs with two worm-like chains (WLCs) in series: one for the dsDNA handle (21) and the other for the ssRNA (22), assuming a contour length of 0.59 nm/nt for RNA (23). When the aptamer unfolded fully,  $62 \pm 1$  nt were released (18), in agreement with the 63-nt aptamer length. The intermediate state is  $23 \pm 1$  nt shorter than the unfolded state, suggesting that it corresponds to a folded 21-nt P2 helix. The equilibrium free energy of the aptamer, computed using the method of Jarzynski (24, 25) from the non-equilibrium work done to unfold it (Fig. S2), is  $18 \pm 2$  kcal/mol (18). For comparison, the free energy predicted for the secondary structure in Fig. 1E is only  $\sim 12 \pm 1$  kcal/mol (10), indicating that tertiary contacts

and ligand binding stabilize the aptamer by an additional  $\sim 6 \pm 2$  kcal/mol, in reasonable agreement with earlier measurements of the binding energy of 2-aminopurine (2AP), an adenine analog (12).

The distribution of forces,  $p(F)$ , for unfolding the fully-folded aptamer (Fig. 5.2B) is well fit by an expression derived by Dudko *et al.* (20) for unfolding at fixed loading rate, parameterized by  $k_{\text{off}}$ , the unfolding rate at  $F = 0$ ,  $\Delta x^\ddagger$ , the distance to the transition state from the folded state, and  $\Delta G^\ddagger$ , the height of the energy barrier (18). Over 3,000 FECs measured for 8 molecules, at loading rates varying from  $\sim 10$ – $200$  pN/s, yielded an unfolding rate  $k_{\text{off}} \sim 0.04 \text{ s}^{-1}$  ( $\ln k_{\text{off}} = -3.5 \pm 1$ ), similar to the value of  $0.15 \text{ s}^{-1}$  measured previously by bulk kinetic methods (12). The activation energy,  $\Delta G^\ddagger$ , was  $17 \pm 4$  kcal/mol, in agreement with a previous result for the unbinding of 2AP (12). The distance to the transition state  $\Delta x^\ddagger$  was  $2.1 \pm 0.2$  nm. Given an extension of  $\sim 0.42$  nm/nt at the most probable unfolding force of  $\sim 15$  pN, this result indicates that the transition state involves the unzipping of  $\sim 2.5$  base pairs (bp) in helix P1, suggesting that the G:C basepair in P1 (Fig. 5.1E, inset) represents a structural keystone: both the binding pocket and triple-helix junction unfold once it is disrupted. Interestingly, isolated G:C basepairs located 3-4 bp from the loop of P1 are found in the other purine riboswitches, suggesting that they may be an important structural feature of this class of aptamers.

The kinetics of refolding and ligand binding were probed by observing the fraction of FECs showing the unfolding signature of the fully-folded, adenine-bound aptamer, as a function of adenine concentration and the variable time interval during which refolding could occur between successive measurements (Fig. 5.3). We fit these data to a minimal, two-step model (Fig. 5.3, inset): formation of an intermediate structure competent to bind adenine (taken to be effectively irreversible) followed by adenine binding. The complete folding process involves a hierarchy of several steps, including folding of the three

helices, formation of the loop-loop contacts and the adenine binding pocket, and binding of adenine. At  $F = 0$ , however, helix formation should be fast compared to formation of the tertiary interactions creating the binding pocket, hence we model this process using just three distinct states. A global fit of data to the time-dependent folding probabilities returned  $k_1 = 0.4 \pm 0.05 \text{ s}^{-1}$ ,  $k_{\text{off}} = 0.2 \pm 0.05 \text{ s}^{-1}$ , and  $k_{\text{on}} = 8 \pm 1 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ . The value of  $k_{\text{off}}$  is similar to that obtained above, and  $k_{\text{on}}$  is close to the value measured by bulk experiments (12). The slow folding rate implies that both aptamer folding and adenine binding occur on the same time scale as transcription itself, supporting the hypothesis that the function of this riboswitch is governed by folding and binding kinetics rather than equilibrium thermodynamics (10, 26).

The multiple steps in the overall folding reaction were studied in greater detail by unfolding the aptamer completely and monitoring its extension under constant force using a passive force clamp (27) as the force was reduced stepwise every 1-2 min. Observing the transitions in the refolding process individually, based on their different energies and time scales, four distinct steps were seen (Fig. 5.4). The first folding event, at  $\sim 9\text{-}11 \text{ pN}$  (Fig. 5.4A, red), involves length changes and force-dependent kinetics consistent with folding the 21-nt hairpin P2, as predicted by an energy landscape model for hairpin folding (16, 18). The second folding step, at  $\sim 7\text{-}8 \text{ pN}$  (Fig. 5.4A, orange), matches the properties expected for folding the 19-nt hairpin P3. The identification of these steps with the folding of P2 and P3 was confirmed by blocking the folding of each hairpin separately using antisense oligomers (Fig. S3). We speculate that P2, the first fully-transcribed element, is also the most stable in order to ensure that it can form in the presence of competing, alternative secondary structures in the upstream mRNA that might delay or prevent the formation of the proper aptamer structure.

In contrast to the adenine-independent events described above, the two folding transitions observed at lower forces were found to be adenine-dependent. For forces below  $\sim 7$  pN at saturating adenine concentrations, the aptamer spent significant time in the shortest-extension state (Fig. 5.4A, green and blue), which we identify as the folded, adenine-bound state. That identification was confirmed by measuring a contour length change of  $63.1 \pm 0.8$  nt when the 63-nt aptamer was completely unfolded from this state. The contour length change between this state and the one with only P2 and P3 folded,  $21 \pm 2$  nt, is consistent with the 23 nt not involved in P2 and P3 folding. In addition, a transient intermediate was observed between these two states,  $14 \pm 1$  nt from the folded state. Because there are 15 nucleotides in and adjacent to P1 (Fig. 5.1A), we identify this intermediate as a state where P2 and P3 are folded and the adenine binding pocket is pre-organized by tertiary contacts, sequestering the nucleotides between P2 and P3, but P1 remains unfolded. The extensions of all five states (fully unfolded, P2 folded, P2/P3 folded, P1 unfolded, and fully folded), scaled by the fractional extension per nucleotide at a given force, are evident in histograms of records (Fig. 5.4B).

Constant-force extension records in the absence of adenine (Figs. 5.4C, D) indicate very different behavior at low forces: the P1-unfolded state is strongly populated below  $\sim 6$  pN, whereas the folded state is only significantly populated below  $\sim 4$  pN. Even at such low forces, the folded state lifetime is short, with a rapid equilibrium between folded and P1-unfolded states. These differences can be understood if the P1-unfolded state is the adenine-competent state. At saturating adenine concentrations, the formation of this state leads rapidly to an adenine-bound, folded state that is long-lived even at  $\sim 7$  pN (Fig. 5.4A), and the P1-unfolded state is thus rarely occupied. In contrast, absent adenine, the P1-unfolded state is frequently occupied even at low forces. Occasionally, the transient folding of P1 was observed even with adenine present (Fig. 4A, green),

likely indicating that adenine was not bound at that instant. The single-molecule records thus directly reflect an adenine-induced stabilization of helix P1 that underpins the switching action of the riboswitch (28).

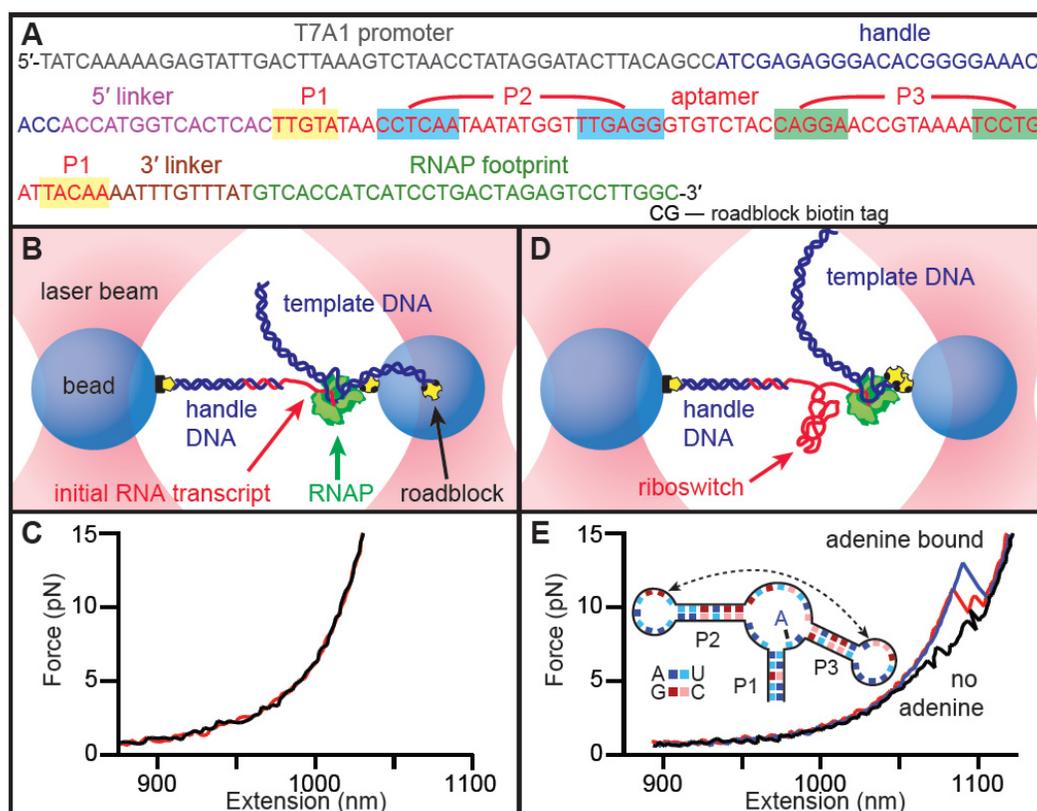
Each of the folding transitions can be analyzed individually as a two-state process, enabling a piecewise reconstruction of the energy landscape for folding, both with and without adenine (Fig. 5.4E). The relative free energies of the five observed states were determined from extension histograms, while the locations and heights of the energy barriers between states were determined from the force-dependence of the kinetics (15, 18). From these landscapes, we find that the tertiary contacts that form the adenine-competent state, which are primarily base-pair and base-quartet interactions between the hairpin loops (4, 9), stabilize the structure by an additional  $2.7 \pm 0.3$  kcal/mol (18). The transition state for breaking these interactions lies  $\sim 1$  nm from the adenine-competent state, indicative of their short range (29). We also find that adenine binding stabilizes the folded state by  $4 \pm 1$  kcal/mol and raises the energy barrier for leaving the folded state, but does not significantly affect other properties of the landscape.

These energy landscapes dramatically illustrate the sequential folding of each structural element in the RNA. Folding proceeds through a distinct hierarchy of states, but the formation of tertiary and secondary structure is interleaved, because the energetic stabilities of these structures happen to be comparable, in contrast with the standard picture of hierarchical folding. In fact, the tertiary contacts that pre-organize the adenine-competent state are considerably more stable than the least-stable helix, P1, which is the essential component governing the switching behavior of the riboswitch. In vivo, without adenine binding to stabilize P1, this last component of the aptamer to fold would be highly susceptible to disruption by terminator hairpin invasion.

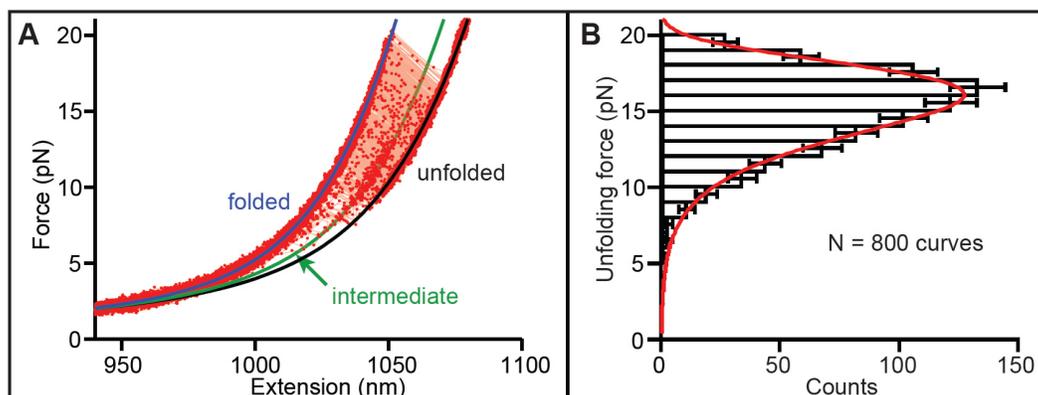
The techniques developed here point the way to a powerful method for monitoring co-transcriptional folding. In the case of the *pbuE* aptamer, the first

FEC obtained after transcription did not exhibit an unfolding behavior substantially different from subsequent FECs, implying that the co-transcriptional aspect of folding may not be important for the formation of an isolated aptamer (18). This result is unsurprising, because structural elements of the aptamer fold in the same order as they are transcribed, hence force-induced refolding mimics co-transcriptional folding in this case. However, for the folding of the complete riboswitch, which includes a downstream terminator hairpin that competes with aptamer formation, we anticipate an important co-transcriptional dependence (10, 26).

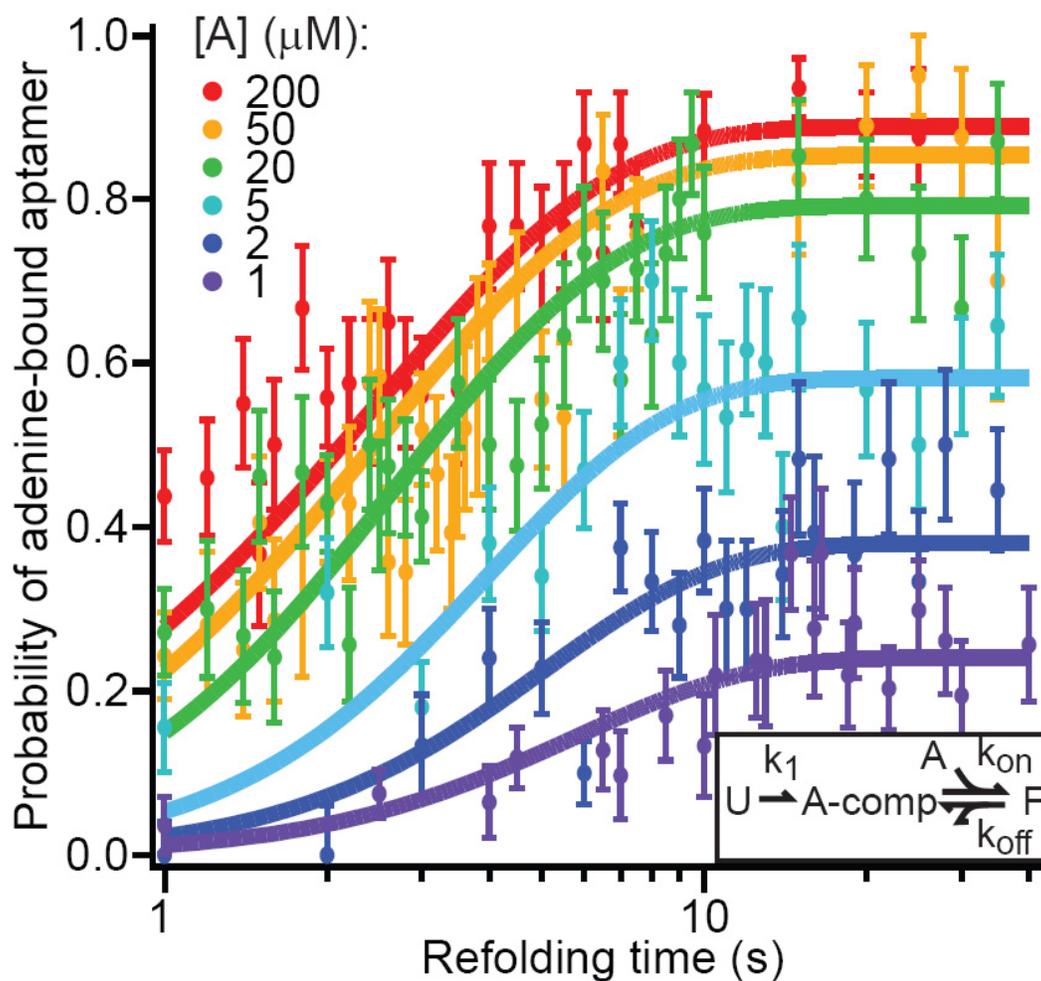
## 5.3 Figures



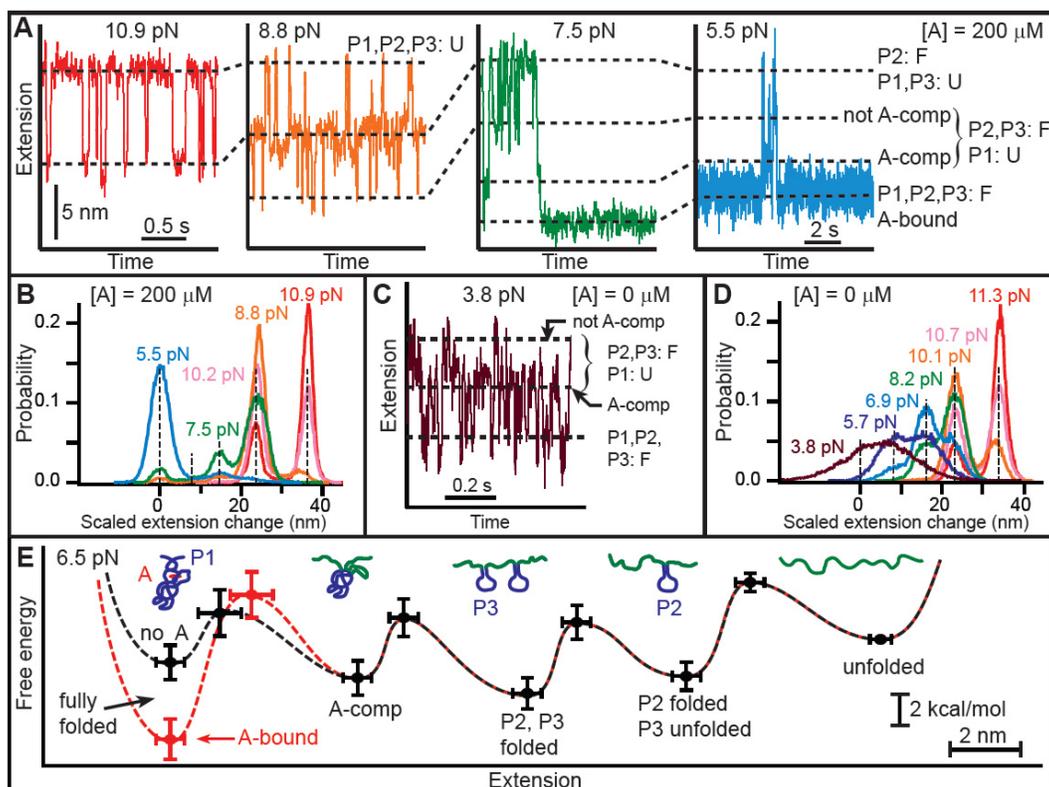
**Figure 5.1** (A) DNA template used for RNA transcription, showing the sequence of the non-transcribed promoter, the 25-bp section hybridizing with the DNA handle, the *pbuE* riboswitch aptamer (base-paired helices highlighted) flanked by short linkers, and the footprint of RNAP when stalled by the terminal roadblock. (B) Schematic of the optical trapping assay showing experimental geometry, with stalled RNAP and initial RNA transcript hybridized to the dsDNA handle (not to scale). (C) Two FECs obtained prior to aptamer transcription show little or no structure in the initial transcript. (D) Template DNA is transcribed *in situ*, producing an aptamer transcript, after which RNAP is stalled by a streptavidin molecule bound to the biotin-based roadblock. (E) FECs obtained after transcription show unfolding transitions in the aptamer. Without adenine, two events are seen (black), corresponding to the unfolding of hairpins P2 and P3 (inset). With adenine bound to the aptamer, larger unfolding events are observed (blue), sometimes involving an intermediate state (red).



**Figure 5.2 (A) Non-equilibrium FECs for folded aptamer display a wide distribution of unfolding forces. WLC fit to the folded state (blue), and double WLC fits to the intermediate (green) and unfolded (black) states, indicate contour length changes of  $39 \pm 1$  nt and  $62 \pm 1$  nt for unfolding to the intermediate and unfolded states, respectively (B) The unfolding force distribution is fit by a model returning the unfolding rate, along with the location and height of the energy barrier to unfolding. [Note added in thesis:  $N=800$  is the number of FECs in the histogram; error bars indicate the standard deviation of counts.]**



**Figure 5.3 Kinetics of aptamer refolding and binding.** The fraction of FECs corresponding to the fully-folded, adenine-bound aptamer (identified by the appropriate unfolding signature) for various adenine concentrations as a function of the variable time delay for refolding between pulls. Solid curves display the global fit to a minimal 3-state kinetic scheme (inset): “U” = unfolded, “A-comp” = competent to bind adenine, “F” = folded; adenine-bound.



**Figure 5.4** Aptamer states and energetics determined by refolding at constant force.

(A) As force is reduced, first P2 refolds (red), then P3 folds (orange). At lower forces, P2 and P3 interact to form a binding pocket and adenine binds, generating two additional states (green). The adenine-bound state is stable over many seconds, even at 5 pN load (blue). (B) Histograms of complete trajectories at different forces, with extension changes scaled by the force-dependent extension per nucleotide. Dashed lines indicate distinct states; the A-comp state is rarely populated. (C, D) Refolding trajectory and histograms in the absence of adenine. P2 and P3 folding occur as with adenine, but the A-comp state is now highly populated at low force, while the folded state is very unstable, even at low force (purple). (E) Quantitative energy landscapes for aptamer folding at 6.5 pN, reconstructed from the experimental data in the presence (red) and absence (black) of adenine. The five potential wells correspond to five observed folding states, illustrated by cartoons. Adenine binding only significantly affects the barrier and energy of the folded state.

(A) As force is reduced, first P2 refolds (red), then P3 folds (orange). At lower forces, P2 and P3 interact to form a binding pocket and adenine binds, generating two additional states (green). The adenine-bound state is stable over many seconds, even at 5 pN load (blue). (B) Histograms of complete trajectories at different forces, with extension changes scaled by the force-dependent extension per nucleotide. Dashed lines indicate distinct states; the A-comp state is rarely populated. (C, D) Refolding trajectory and histograms in the absence of adenine. P2 and P3 folding occur as with adenine, but the A-comp state is now highly populated at low force, while the folded state is very unstable, even at low force (purple). (E) Quantitative energy landscapes for aptamer folding at 6.5 pN, reconstructed from the experimental data in the presence (red) and absence (black) of adenine. The five potential wells correspond to five observed folding states, illustrated by cartoons. Adenine binding only significantly affects the barrier and energy of the folded state.

Adenine binding only significantly affects the barrier and energy of the folded state.

#### 5.4 References and Notes

1. W. C. Winkler, R. R. Breaker, *Annu. Rev. Microbiol.* **59**, 487 (2005).
2. R. L. Coppins, K. B. Hall, E. A. Groisman, *Curr. Opin. Microbiol.* **10**, 176 (2007).
3. R. T. Batey, S. D. Gilbert, R. K. Montange, *Nature* **432**, 411 (2004).
4. A. Serganov *et al.*, *Chem. Biol.* **11**, 1729 (2004).
5. M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, *Cell* **113**, 577 (2003).
6. J. Noeske *et al.*, *Proc. Natl. Acad. Sci. USA* **102**, 1372 (2005).
7. S. D. Gilbert, C. D. Stoddard, S. J. Wise, R. T. Batey, *J. Mol. Biol.* **359**, 754 (2006).
8. J. Noeske *et al.*, *Nucleic Acids Res.* **35**, 572 (2007).
9. O. M. Ottink *et al.*, *RNA* **13**, 2202 (2007).
10. J. F. Lemay, J. C. Penedo, R. Tremblay, D. M. Lilley, D. A. Lafontaine, *Chem. Biol.* **13**, 857 (2006).
11. M. Mandal, R. R. Breaker, *Nat. Struct. Mol. Biol.* **11**, 29 (2004).
12. J. K. Wickiser, M. T. Cheah, R. R. Breaker, D. M. Crothers, *Biochemistry* **44**, 13404 (2005).
13. E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, S. M. Block, *Nature* **438**, 460 (2005).
14. I. Tinoco, Jr., C. Bustamante, *Biophys. Chem.* **101-102**, 513 (2002).
15. M. T. Woodside *et al.*, *Science* **314**, 1001 (2006).
16. M. T. Woodside *et al.*, *Proc. Natl. Acad. Sci. USA* **103**, 6190 (2006).
17. R. V. Dalal *et al.*, *Mol. Cell.* **23**, 231 (2006).
18. Materials and methods are available as supporting material on *Science* online.
19. B. Onoa *et al.*, *Science* **299**, 1892 (2003).
20. O. K. Dudko, G. Hummer, A. Szabo, *Phys. Rev. Lett.* **96**, 108101 (2006).
21. S. B. Smith, Y. Cui, C. Bustamante, *Science* **271**, 795 (1996).

22. Y. Seol, G. M. Skinner, K. Visscher, *Phys. Rev. Lett.* **93**, 118102 (2004).
23. W. Saenger, *Principles of Nucleic Acid Structure* (Springer, New York, 1984)
24. J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., C. Bustamante, *Science* **296**, 1832 (2002).
25. C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
26. J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, *Mol. Cell.* **18**, 49 (2005).
27. W. J. Greenleaf, M. T. Woodside, E. A. Abbondanzieri, S. M. Block, *Phys. Rev. Lett.* **95**, 208102 (2005).
28. Some molecular heterogeneity was present, but we did not observe order-of-magnitude variations in the kinetics from one RNA to the next, as reported in one single-molecule fluorescence study (10).
29. P. T. X. Li, C. Bustamante, I. Tinoco, Jr., *Proc. Natl. Acad. Sci. USA* **103**, 15847 (2006).
30. We thank Robert Landick for providing purified RNAP, Kristina Herbert for providing the plasmid used to make the transcription template, Matt Larson and Olga Dudko for helpful advice, and members of the Block lab and Program Project Grant GM066275 for useful discussions. Our research was supported by grants from the NIGMS (GM057035) and the National Institute for Nanotechnology.

### Supporting Online Material

[www.sciencemag.org](http://www.sciencemag.org)

Materials and Methods

Tables S1, S2

Figs. S1, S2, S3

## 5.5 Supporting Online Material –

### 5.5.1 *Materials and Methods*

#### 5.5.1.1 Constructs.

A 78 bp DNA fragment coding for the *pbuE* riboswitch aptamer was inserted into a pALB3 plasmid 31 bp downstream from a T7A1 promoter (Fig. 5.1A). An extra 9 bp were included in the insert at the 5' end of the aptamer sequence, so that ~40 nt of ssRNA would be transcribed prior to the aptamer-containing region. The first 25 nt of the transcript were designed to hybridize with the 5' cohesive end of a dsDNA handle, leaving a ~15-nt RNA linker region between the DNA handle and the aptamer structure after transcription. The 770-bp dsDNA transcription template was amplified by PCR from the cloned plasmid such that the template ended with a terminal biotin label 41 bp downstream of the aptamer sequence. The PCR product was incubated in a ~20-fold excess of streptavidin (ProZyme) at room temperature for 5 min to bind the streptavidin to the downstream terminus of the DNA template, followed by a ~1000-fold excess of biotin to saturate any unused streptavidin binding sites, and then purified. The final product was a DNA template coding for the *pbuE* riboswitch aptamer carrying a terminal streptavidin roadblock downstream. Because the RNAP molecule has a ~30 bp footprint on the DNA template (*SI*), the aptamer part of the transcript is anticipated to be separated from the polymerase by ~11 nt of RNA once transcription is stopped by the roadblock. This allows the aptamer to be unfolded and refolded without being perturbed by the nearby RNAP enzyme.

Biotinylated RNAP molecules were initiated on the T7A1 promoter of the transcription template. By omitting UTP from the initiation reaction, these became stalled at the first T residue of the template, after first transcribing 29 nt of RNA, as previously described (S2). The stalled transcription elongation complexes (TECs) were purified of free NTPs using a size-exclusion column. A 3057 bp-long dsDNA handle with a 5' overhang of 25 nt ssDNA complementary to the first 25 nt of the RNA transcript was created as previously described (S3) by autosticky PCR of the M13mp18 plasmid, and then incubated with a near-stoichiometric amount of the stalled TEC at room temperature for 1 hr. The resulting mixture was incubated for 1 hr at room temperature with near-stoichiometric amounts of 600-nm diameter, biotin-coated polystyrene beads and 730-nm diameter, digoxigenin-coated polystyrene beads at an overall concentration of each of the components of ~1 nM. This final incubation was diluted ~20-fold into an RNase-free oxygen-scavenging buffer system consisting of 40 U/mL glucose oxidase (Calbiochem), 185 U/mL catalase (Sigma), and 8.3 mg/mL glucose (Sigma) in transcription buffer (50 mM HEPES pH 8.0, 130 mM KCl, 4 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 0.1 mM DTT), and placed in a flow chamber on a microscope slide. Transcription was initiated *in situ* by introducing a buffer containing 1 mM NTPs into the flow chamber.

#### 5.5.1.2 Optical trap.

The optical trapping apparatus used in these measurements has been described previously (S4). Briefly, a laser beam (1064 nm; Spectra-Physics) was split into

two separately controlled traps by polarization, with one trap much stiffer than the other. The strong trap was steered in the specimen plane by acousto-optic deflectors (AODs; IntraAction Inc.), while the position of the bead held in the weak trap was monitored by collecting the light from a second laser (830 nm; Point Source) scattered off the bead onto a position sensitive detector (Pacific Silicon Sensors). Force-extension curves (FECs) were measured by moving the traps apart at a constant velocity using the AODs. Discrete data points in the FECs represented 2.5 ms of signal integration time, acquired using custom Labview software (National Instruments). Refolding at constant force was measured using the passive force clamp technique described previously (S5), in order to avoid spurious instrumental artefacts arising from active feedback loops (S6). Some refolding records were also measured in an “open-loop” configuration, without a force clamp. Data in refolding records were sampled at 4 kHz and filtered at 2 kHz with an 8-pole Bessel filter (Krohn-Hite). All measurements in the optical trap were taken at a temperature of  $\sim 24^{\circ}\text{C}$ . The temperature of the room was clamped to within  $\pm 0.2^{\circ}\text{C}$ .

### ***5.5.2 Force-extension curves.***

The presence of single DNA tethers held in the dumbbell arrangement (Fig. 5.1B) was confirmed by measuring the contour lengths and persistence lengths of the DNA handle from FECs obtained at low forces (dumbbells found to be connected by multiple tethers were excluded from further analysis). FECs without adenine present were measured after transcription without any change of buffer. FECs

with adenine present were measured after washing the flow chamber with at least 3 volumes of transcription buffer containing the desired adenine concentration. For measurement of low concentrations of adenine ( $<20 \mu\text{M}$ ), transcription buffer with  $20 \mu\text{M}$  adenine was flushed into the flow cell and incubated for 10 min. Then, the buffer was replaced with a buffer containing the final adenine concentration desired, followed by  $\sim 5$  min of incubation and at least one more wash with the low adenine concentration buffer. These wash steps ensured that any propensity of adenine to stick to the flow chamber surfaces would not reduce the overall concentration of free adenine in solution. Where not otherwise indicated in the text, FECs were measured with a constant delay time before each pull of 3 s, in the presence of  $200 \mu\text{M}$  adenine.

FECs were partitioned into folded, intermediate, and unfolded states based on the structure of the sawtooth unfolding pattern, as seen in Fig. 5.1E. The results of multiple measurements were aligned to remove the small amount of residual instrumental drift ( $\sim 2$  nm or less) that occurred over the course of the experiment, then averaged. The average FEC for the folded state was fit to a single worm-like chain (WLC) model using a modified Marko-Siggia interpolation formula (S7):

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

where  $L_c$  is the contour length of the handle,  $L_p$  is the persistence length,  $K$  is the elastic modulus, and  $k_B$  is Boltzmann's constant. The average FECs for the

intermediate and unfolded states were then fit to a double WLC model, which included terms for both the extension of the dsDNA handle (as determined by the fit to the folded-state FEC) and for the extension of the now-unfolded ssRNA. Previous work (*S8-S13*) has reported a range of values for the persistence length of ssRNA, which has been found to be sequence-dependent (*S9*); an average value of 1.0 nm was used for the fits performed here. The ssRNA elastic modulus was taken to be 1600 pN/nm (*S8*), and a ssRNA contour length of 0.59 nm/nt, a value expected from the structure of the 3' endo sugar pucker, was assumed (*S14*). The width of the A-form helix was taken to be 2.2 nm (*S14*); this width was subtracted from the extension of the folded state when fitting the FECs to determine the number of nucleotides released during complete unfolding.

We note that the observed contour length change (corresponding to  $62 \pm 1$  nt) indicates that the adenine-bound, folded state likely does not exhibit any significant “fraying” of helix P1 under tension, despite the fact that the two closing basepairs of P1 are comparatively weak AU pairs. This result contrasts with the fraying previously observed in DNA hairpins with different weak closing sequences (*S15, S16*), suggesting that tertiary interactions with the nearby binding pocket may supply additional mechanical stability.

FECs with a structure that indicates that adenine did not bind to the aptamer are displayed in Fig. 5.5 (5.S1). Three states can be observed. As seen previously when adenine does bind (Fig. 5.2A), one portion of the FEC corresponds to a well-defined unfolded state ( $63 \pm 2$  nt of ssRNA), and another corresponds to an

intermediate state whose contour length is  $22 \pm 2$  nt shorter. At low force, however, the adenine-free FEC is fit by a WLC with  $24 \pm 2$  nt of ssRNA still present. Thus, the three states involve  $\sim 39$  nt folded,  $\sim 22$  nt folded, and 0 nt folded, implying that just prior to the first unfolding event in these FECs, only P2 and P3 are folded (40 nt in total), after which P3 unfolds, leaving only P2 folded (21 nt), before P2 finally unfolds.

### 5.5.3 Force distributions.

Unfolding force distributions created by measuring the first unfolding event from the fully-folded state in each FEC were fit to the non-equilibrium model of Dudko *et al* (S17):

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

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

$k_{\text{off}}$  is the unfolding rate at 0 force,  $\Delta x^\ddagger$  is the distance to the transition state from the folded state,  $\Delta G^\ddagger$  is the height of the energy barrier, and  $\nu$  is a parameter characterizing the shape of the energy barrier ( $\nu = 1/2$  for a sharp, cusp-like barrier;  $\nu = 2/3$  for a softer, cubic potential). The shape of the barrier is unknown, but the cusp-like and cubic models represent two reasonable limiting cases. For all fitting parameters, the two models gave results that were equal within uncertainty: Assuming a sharp, cusp-like barrier, we obtained an average

unfolding rate for all molecules measured of  $k_{\text{off}} \sim 0.03 \text{ s}^{-1}$ , an average distance to the barrier of  $\Delta x^\ddagger = 2.1 \pm 0.2 \text{ nm}$ , and an average barrier height of  $\Delta G^\ddagger = 19 \pm 4 \text{ kcal/mol}$ . Assuming instead a softer, linear-cubic potential landscape, the average results were  $k_{\text{off}} \sim 0.04 \text{ s}^{-1}$ ,  $\Delta x^\ddagger = 2.1 \pm 0.2 \text{ nm}$ , and  $\Delta G^\ddagger = 15 \pm 4 \text{ kcal/mol}$ . Thus the principal difference between these alternative models would seem to be in the height of the barrier returned, and even that quantity is found to be the same within our experimental uncertainty. We therefore averaged the results obtained with both these models to obtain the rates and energetics derived from our data reported here. The rate  $k_{\text{off}}$  is highly sensitive to the force, hence the uncertainty in  $k_{\text{off}}$  is expressed in terms of  $\ln(k_{\text{off}})$  in the text.

#### ***5.5.4 Equilibrium free energy.***

The equilibrium folding free energy was calculated from the non-equilibrium FEC measurements using the method of Jarzynski (SI8), which involves an exponentially-weighted mean of the irreversible work done to unfold the molecule. In order to calculate the irreversible work from the measured FEC, the energy required to stretch the molecular handles must also be taken into account, because this energy is not associated with the aptamer structure itself. Therefore, calculation of the relevant irreversible work involves first finding the work performed to stretch out the entire construct (including handles) and the work to unfold the aptamer, by integrating the measured FEC out to a point corresponding to the end of the last unfolding event, including any intermediates (Fig. 5.S2, red shading). Then, the energy required to stretch out the handles and ssRNA to the

identical force must be subtracted from this work. The latter energy is calculated by integrating the FEC expected for the fully-unfolded state out to the same point (Fig. 5.S2, black shading). This procedure is equivalent to integrating the FECs between constant extension endpoints (*e.g.*, 0 and some extension value above all unzipping events for a given molecule), calculating the weighted average, and subtracting the work done to stretch the handles, under the assumption that the handles are at equilibrium throughout the experiment. To reduce fluctuations from noise in the experimental FECs, these integrals were calculated from the unfolding forces measured in each FEC using the fits for the averaged FECs for folded, intermediate, and unfolded states (as shown in Fig. 2A).

The Jarzynski estimator is known to have a systematic bias when only finite numbers of measurements are sampled, due to nonlinear weighting of the data. We estimate the bias, under the assumption of a Gaussian free energy distribution (*S19*), as  $3 \pm 1$  kcal/mol. This bias is included in the equilibrium free energy reported in the main text.

#### ***5.5.5 Kinetics from FECs.***

To measure refolding and binding kinetics, the aptamer was first completely denatured by exerting a large load ( $\sim 20$  pN). The load was then immediately reduced to  $\sim 0$  pN by reducing the bead separation to  $\sim 350$  nm (transition time  $< 5$  ms), a distance much less than the contour length of the 3-kb DNA handle. After a variable delay time to allow refolding to occur, a FEC was measured by immediately jumping the tension to  $\sim 2$  pN and then moving the two traps apart at

a rate of  $\sim 300$  nm/s, until the force reached  $\sim 20$  pN and the aptamer was once again denatured. FEC measurements were repeated 10-20 times for each desired value of the delay time, for at least two different molecules, and the fraction of FECs that had refolded to the fully-folded, adenine-bound structure was determined from the unfolding signature, as in Fig. 1.

The rates plotted in Fig. 5.3, with standard binomial error estimates, were fit to the time-dependent probability for the folded state as calculated from the rate equations corresponding to the 3-state reaction pathway (Fig. 5.3, inset). Because our constant-force measurements imply that the equilibrium free energy difference between fully-folded and adenine-competent states is  $\sim 6 k_B T$ , (Table S1), the folding of the adenine-competent state was taken to be effectively irreversible. We note that the experimental value for the refolding/binding saturated at  $\sim 90\%$  at high adenine concentrations (rather than 100%), possibly due to some misidentification of a minor fraction of FECs, or to the presence of a long-lived unfolded or misfolded state. This saturation level was included when calculating rates for the three-state model.

From the measured binding and dissociation rates of adenine, the dissociation constant is  $K_d = 2.5 \pm 0.7 \mu\text{M}$ , implying a standard Gibbs free energy of  $7.6 \pm 0.2$  kcal/mol at  $24^\circ\text{C}$ . This value is comparable to a previous result reported for the standard Gibbs free energy for 2AP binding (S20),  $\sim 8.3$  kcal/mol (uncertainty not reported).

### 5.5.6 Hairpin unfolding predictions.

A number of predictive models have been developed to describe the folding of hairpins under tension (*S16, S21-S23*). Here, we use an energy landscape model developed previously by our groups that has been extensively tested for DNA hairpin folding/refolding in an optical trap (*S15, S16*). We assume an ssRNA contour length of 0.59 nm/nt, and nearest-neighbor stacking energies that are 25% lower than the canonical values measured in 1 M salt, to account for the reduced salt concentration used in our assay buffer (*S24*). The results of the model are listed in Table 5.2(S2) for hairpins P2 and P3. For comparison, the unfolding force, unfolding distance, and distance to transition state observed for the folding events correlated to P2 and P3 folding are listed in Table 5.1(S1). To convert measured rates into barrier heights, we assumed that the rates vary as  $k = k_0 \cdot \exp(-\Delta G^\ddagger / k_B T)$ , where the prefactor  $k_0$  was taken to be  $10^5 \text{ s}^{-1}$ , in accordance with previous measurements of the kinetics of unfolding DNA hairpins in an optical trap (*S16*). We estimated the uncertainties in the model predictions from the standard deviation of the results calculated using a range of ssRNA persistence lengths compatible with previous measurements (0.8-1.3 nm). Uncertainties in the stacking interactions and salt correction (assumed to introduce ~7% error) were added in quadrature.

We note that there is good overall agreement between the experimental observations and model predictions, with the exception of the opening force (hence also the folding free energy) for hairpin P2: the measured value is ~1-

2 kcal/mol larger than predicted. Given the prior success of the model in predicting the behavior of hairpins for a wide range of stem sequences, we postulate that this discrepancy may be due to some extra stability in the loop of hairpin P2 imparted by intraloop base stacking and/or hydrogen bonding, or from the formation of additional, non-canonical base pairs in the stem of this hairpin, neither of which is considered by the model.

#### ***5.5.7 Refolding trajectories.***

Refolding trajectories were measured both in the presence and absence of 200  $\mu\text{M}$  adenine. Data were filtered offline using a median filter with a 10 ms window, then partitioned into two or more states using a threshold algorithm similar to that previously described (*S16*). The extension change for each transition,  $\Delta x$ , was measured directly from Gaussian fits to position histograms formed from the extension records (*S15*, *S16*). The number of RNA nucleotides involved in each folding step was calculated from the extension change at a given force by adding (or subtracting, as appropriate) the width of an A-form helix, 2.2 nm, and dividing by the extension expected per nucleotide at that force, assuming a WLC model for ssRNA, as described above. For folding the P2 and P3 helices, one helix width was added to the observed extension change. For folding to the adenine-competent state, one helix was subtracted, because published structures of the purine riboswitches suggest that the loop-loop interaction that causes binding-pocket pre-organization arranges the 5' end of P2 and the 3' end of P3 approximately one helix-width away from each other (*S25*, *S26*). For folding P1,

no helix correction was applied. The extension change upon fully unfolding the aptamer at the start of the refolding measurements, in the presence of adenine, was found to be  $63 \pm 1$  nt, fully consistent with the FEC measurements.

Each folding transition was analyzed as a separate two-state system, as described previously (S16). Briefly,  $F_{1/2}$ , defined as the force at which a structural element spent equal time in folded and unfolded states, was determined by plotting the fraction of time spent unfolded as a function of force, and fitting this to the Boltzmann relation for a two-state system:

$$P_u(F) = \{1 + \exp[(F_{1/2} - F) \cdot \Delta x / k_B T]\}^{-1} \quad (S24).$$

The free energy for each folding step was then computed from the quantity  $F_{1/2} \cdot \Delta x$ , minus the energy required to stretch out the unfolded ssRNA (calculated from an integral of a worm-like chain of appropriate length). The average rate for transitions from state A to state B was calculated by dividing the number of transitions from A to B by the total amount of time spent in state A. The distance along the reaction coordinate to the transition state,  $\Delta x^\ddagger$ , was determined for each folding step from the force dependence of the rates. Assuming a sharp energy barrier,  $\Delta x^\ddagger$  is given by the slope of the logarithm of the rates as a function of force, since  $k(F) = k_0 \exp(F \Delta x^\ddagger / k_B T)$  (S27). The results for each transition are summarized in Table 5.1(S1). Extracting the rates for the final transition to the fully-folded, adenine bound state presented a special challenge, because adenine-binding shifts  $F_{1/2}$  for this state to a value higher than  $F_{1/2}$  for folding P3. Since P3 folding is required to form the adenine binding pocket, the  $F_{1/2}$  value for the adenine-bound

folded state had to be extrapolated from data collected well below  $F_{1/2}$ , resulting in a comparatively large uncertainty.

The proper identification of the two folding events observed at the highest forces, corresponding to P2 and P3 folding, was confirmed by blocking the folding of each of these hairpins separately, using anti-sense oligomers complementary to the loop and 3' stem of hairpin P2 (denoted "anti-P2") or the loop and 5' stem of hairpin P3 (denoted "anti-P3"). Comparing folding trajectories measured at ~11 pN with and without the anti-P2 oligomer, the bistable behavior attributed to folding and unfolding of hairpin P2 (Fig. 5.S3A, red) is suppressed by the presence of the anti-P2 oligomer (Fig. 5.S3A, blue), confirming the identification. Similarly, the transition at ~8 pN attributed to hairpin P3 is suppressed by the anti-P3 oligomer (Fig. 5.S3B). Notably, however, the P2 folding transition is not affected by the anti-P3 oligomer, as seen in Fig. 5.S3B (blue) from the upward spike, corresponding to P2 folding.

The conformation of the adenine-competent state was determined by comparing the length changes observed between this state and both P2/P3-folded and the fully folded states. The observed length change of  $7 \pm 1$  nt between the P2/P3-folded and adenine-competent states implies that the stems of P2 and P3 are brought into contact in the adenine-competent state, so that J2-3 (the 8-nt junction segment between P2 and P3) is no longer stretched out. This contact is likely maintained by interactions between the loops of the hairpins and by base-pairing between the terminal G and C residues of J2-3 (S25). On the other hand, the

length change of  $14 \pm 1$  nt between the adenine-competent and fully-folded states indicates that the 15 residues in P1, J1-2 (the junction segment between P1 and P2), and J3-1 (the junction segment between P3 and P1) are still unfolded in the adenine-competent state. Thus, in the adenine-competent state, the residues in P1, J1-2, and J3-1 are stretched away from the junction, and hence not available for creating tertiary contacts with J2-3. This forms the basis for our claim that the adenine-competent state is stabilized largely by contacts between the loops, rather than in the junction. This interpretation finds additional support in NMR structures of the guanine riboswitch aptamer (S28), in which structural changes upon ligand binding were seen to occur predominantly in the J2-3 and P1 regions, suggesting that most tertiary interactions in the junction are not present in the absence of ligand.

We do not observe evidence for an additional intermediate state between the P2/P3 folded and adenine-competent states, as postulated by Lemay *et al.* to explain their single-molecule FRET data (S29). This may be attributable to differences in the experimental conditions, including buffer conditions or the effects of tension on the RNA molecule.

Extension histograms obtained at different forces were compared (Figs. 5.4B, D) by first scaling the measured extension changes by the fractional extension of the ssRNA (*i.e.*, the extension per nucleotide divided by the contour length per nucleotide) at a particular force, as given by the WLC model. This is not quite the

same as the true contour length change, since corrections for the creation or dissolution of helices are not included.

### ***5.5.8 Free energy landscapes.***

Free energy landscapes (Fig. 5.4E) are depicted at a constant force based on the relative free energies of the 5 different states determined in the refolding trajectories, and on the locations and heights of the energy barriers between the various states determined from the force-dependent kinetics. Rates under tension were converted into energy barriers assuming the prefactor  $k_0 = 10^5 \text{ s}^{-1}$ , as determined previously in an optical trapping study of DNA hairpins (SI6). Due to experimental uncertainties, the sum of the distances to the transition state from folded and unfolded states did not always precisely equal the total unfolding distance. In these cases, the distances to the transition state were scaled so that their sum was equal to the total distance. Note that these diagrams represent the effective potential landscape under tension, and hence include the stretching energy of the unfolded ssRNA.

Note that the transition state for leaving the adenine-bound, folded state, as measured from the refolding trajectories at constant force, involves unfolding the folded state by a distance corresponding to  $\sim 6$  nt. This distance matches the result from the non-equilibrium unfolding measurements, reinforcing the conclusion that the barrier to P1 unfolding occurs near the G:C pair in helix P1.

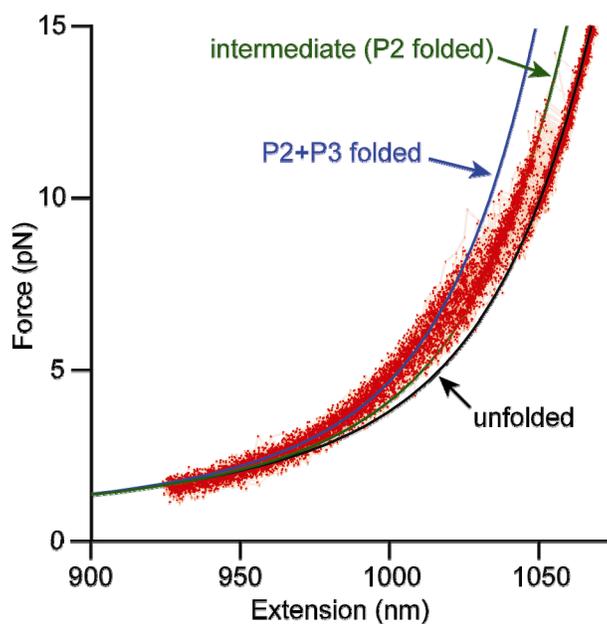
### ***5.5.9 Co-transcriptional folding.***

The question of whether aptamer folding is different when occurring concurrently with transcription (co-transcriptional folding) was addressed by comparing the first FEC measured immediately after transcription of the aptamer to all the FECs measured subsequently, for each molecule. Such comparisons were made for 23 molecules. No differences in the unfolding distance, unfolding force, or fraction of events displaying unfolding intermediates were evident in our data, suggesting that folding of the aptamer is the same whether formed co-transcriptionally or by refolding after mechanical disruption.

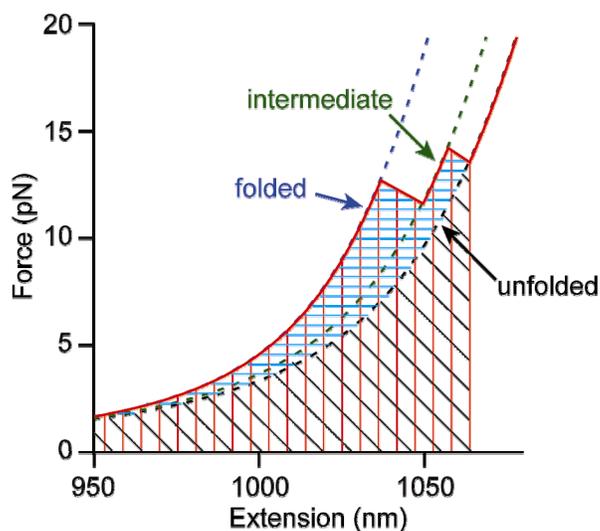
### ***5.5.10 Primary Sources of Experimental Error.***

The experimental uncertainties cited above and in the main text and figures include the effects of both statistical and systematic errors. The principal source of random error in these experiments is the variation in the sizes of beads, which leads directly to an uncertainty in position measurements (estimated at ~5–10%) and also, indirectly, to calibration errors in force associated with changes in the distribution of scattered light that lead to an uncertainty in the effective trap stiffness (estimated ~10%). Such errors are reduced by making measurements on as many beads as possible. An additional source of error in constant force measurements comes from uncertainty in the location of the zero-stiffness position in the optical trap (S5). We estimate that residual stiffness, associated with being displaced from the center of the zero-stiffness region of the trap, could introduce an uncertainty of up to 5-10% in distance measurements.

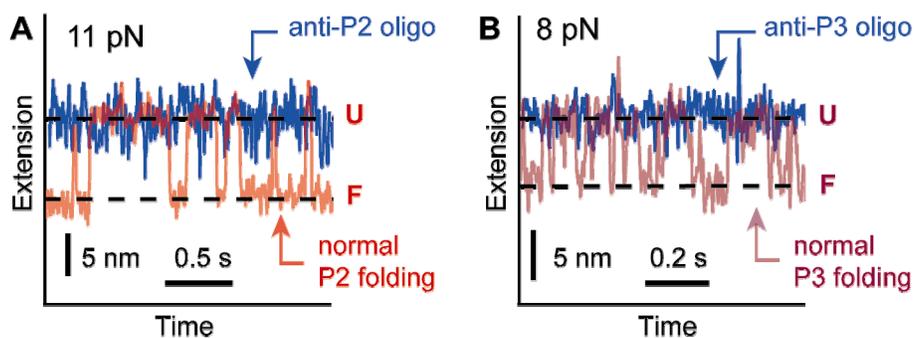
The principal source of systematic error arises from the calibration of trap stiffness. We estimate our systematic error in force to be ~10-15%. A second source of systematic error, relevant to the FEC analysis, comes from the fact that unfolding events at very low force ( $<5$  pN) are typically difficult to observe, effectively truncating the force distribution. This may serve to increase the effective bias in the Jarzynski estimate of the equilibrium free energy (*S19*).



**Figure 5.5 (Fig. 5.S1.)** FECs observed when adenine is not bound to the aptamer (same molecule as Fig. 5.2A). Two unfolding events are seen. The first event involves unfolding of P3; the second event involves the unfolding of P2, the same as the intermediate seen in Fig. 5.2A.



**Figure 5.6 (Fig. 5.S2)** Graphical representation of the work performed during stretching. The non-equilibrium work done to unfold the aptamer (blue) is given by the area under the measured FEC out to the last unfolding event (red), minus the area under the unfolded state FEC out to the same point (black).



**Figure 5.7 (Fig. 5.S3) Refolding traces in presence and absence of anti-sense oligos. (A) In the presence of an anti-P2 oligo, the normal folding of P2 at ~9-11 pN (red) is suppressed (blue). (B) In the presence of an anti-P3 oligo, the normal folding of P3 at ~7-8 pN (purple) is suppressed (blue). Note that the normal unfolding of P2, evident as a brief increase in extension (blue), is not suppressed by the presence of an anti-P3 oligo. Extensions of records obtained with and without oligos have been aligned on the common unfolded state to remove any drift occurring during buffer exchange.**

**Table 5.1 (S1). Measured values describing individual folding steps of *pbuE* aptamer. The four steps are: P2 folding, denoted “P2”; P3 folding, denoted “P3”; adenine-competent state folding, denoted “AC” (adenine-competent); and full folding with adenine bound, denoted “AF” (adenine; folded), or full folding in the absence of adenine, denoted “NAF” (no adenine; folded).**

**Distances are reported in nm for forces as close as possible to  $F_{1/2}$ , except in the case of AF (for which  $F_{1/2}$  could not be reached). Values were determined from measurements on 3 different molecules in the absence of adenine and 7 in the presence of adenine.**

Folding step	$\Delta x$ (nm)	$\Delta x$ (nt)	$F_{1/2}$ (pN)	$\ln k_{1/2}$ ( $s^{-1}$ )	$\Delta G_{1/2}^\ddagger$ (kcal/mol)	$\Delta x_r^\ddagger$ (nm)	$\Delta x_u^\ddagger$ (nm)
P2	$6.4 \pm 0.5$	$22 \pm 1$	$10.0 \pm 0.8$	$2.9 \pm 0.3$	$5.2 \pm 0.6$	$2.5 \pm 0.4$	$5.1 \pm 0.6$
P3	$4.6 \pm 0.5$	$20 \pm 1$	$7.0 \pm 0.6$	$4.1 \pm 0.3$	$4.4 \pm 0.6$	$1.8 \pm 0.3$	$4.1 \pm 0.6$
AC	$4.1 \pm 0.5$	$7 \pm 1$	$5.1 \pm 0.5$	$4.5 \pm 0.6$	$4.2 \pm 0.7$	$1.4 \pm 0.4$	$3.7 \pm 0.6$
AF	$4.2 \pm 0.5$	$14 \pm 1$	$9 \pm 1$	$0 \pm 1$	$7 \pm 1$	$2.3 \pm 0.7$	$3 \pm 0.7$
NAF	$4.4 \pm 0.5$	$18 \pm 3$	$3.1 \pm 0.6$	$5 \pm 1$	$4 \pm 1$	$1.8 \pm 0.6$	$5.1 \pm 1.5$

**Table 5.2 (S2). Model calculation results for folding of the P2 and P3 hairpins. Predicted errors include the effects of uncertainties in the nearest-neighbor stacking parameters and in the value of the ssRNA persistence length.**

Hairpin	$\Delta x$ (nm)	$F_{1/2}$ (pN)	$\Delta G^\ddagger$ (kcal/mol)	$\Delta x_r^\ddagger$ (nm)	$\Delta x_u^\ddagger$ (nm)
P2	$6.5 \pm 0.4$	$6.9 \pm 0.6$	$6.9 \pm 0.5$	$2.2 \pm 0.2$	$4.3 \pm 0.4$
P3	$4.8 \pm 0.4$	$6.7 \pm 0.6$	$5.0 \pm 0.4$	$1.4 \pm 0.2$	$3.3 \pm 0.4$

### 5.5.11 References

- S1. S. A. Darst, *Curr. Opin. Struct. Biol.* **11**, 155 (2001).
- S2. K. C. Neuman, E. A. Abbondanzieri, R. Landick, J. Gelles, S. M. Block, *Cell* **115**, 347 (2003).
- S3. R. V. Dalal *et al.*, *Mol. Cell* **23**, 231 (2006).
- S4. E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, S. M. Block, *Nature* **438**, 460 (2005).
- S5. W. J. Greenleaf, M. T. Woodside, E. A. Abbondanzieri, S. M. Block, *Phys. Rev. Lett.* **95**, 208102 (2005).
- S6. M. Mañosas *et al.*, *Biophys. J.* **92**, 3010 (2007).
- S7. M. D. Wang, H. Yin, R. Landick, J. Gelles, S. M. Block, *Biophys. J.* **72**, 1335 (1997).
- S8. Y. Seol, G. M. Skinner, K. Visscher, *Phys. Rev. Lett.* **93**, 118102 (2004).
- S9. Y. Seol, G. M. Skinner, K. Visscher, *Phys. Rev. Lett.* **98**, 158103 (2007).
- S10. F. Vanzi, Y. Takagi, H. Shuman, B. S. Cooperman, Y. E. Goldman, *Biophys. J.* **89**, 1909 (2005).
- S11. P. Mangeol, D. Cote, T. Bizebard, O. Legrand, U. Bockelmann, *Eur. Phys. J. E* **19**, 311 (2006).
- S12. G. Caliskan *et al.*, *Phys. Rev. Lett.* **95**, 268303 (2005).
- S13. F. Kienberger *et al.*, *Biomaterials* **28**, 2403 (2007).
- S14. W. Saenger, *Principles of Nucleic Acid Structure* (Springer, New York, 1984).
- S15. M. T. Woodside *et al.*, *Science* **314**, 1001 (2006).
- S16. M. T. Woodside *et al.*, *Proc. Natl. Acad. Sci. USA* **103**, 6190 (2006).
- S17. O. Dudko, G. Hummer, A. Szabo, *Phys. Rev. Lett.* **96**, 108101 (2006).
- S18. C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- S19. J. Gore, F. Ritort, C. Bustamante, *Proc. Natl. Acad. Sci. USA* **100**, 12564 (2003).
- S20. J. K. Wickiser, M. T. Cheah, R. R. Breaker, D. M. Crothers, *Biochemistry* **44**, 13404 (2005).
- S21. S. Cocco, J. F. Marko, R. Monasson, *Eur. Phys. J. E Soft Matter* **10**, 153 (2003).
- S22. C. Hyeon, D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **102**, 6789-6794 (2005).
- S23. J. R. Viereg, I. Tinoco Jr., *Mol. Phys.* **104**, 1343 (2006).
- S24. J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr., C. Bustamante, *Science* **292**, 5517 (2001).
- S25. R. T. Batey, S. D. Gilbert, R. K. Montange, *Nature* **432**, 411 (2004).
- S26. A. Serganov *et al.*, *Chem. Biol.* **11**, 1729 (2004).
- S27. C. Bustamante, Y. R. Chemla, N. R. Forde, D. Izhaky, *Annu. Rev. Biochem.* **73**, 705 (2004).
- S28. O. M. Ottink *et al.*, *RNA* **13**, 2202 (2007).
- S29. J.-F. Lemay, J. C. Penedo, R. Tremblay, D. M. J. Lilley, D. A. Lafontaine, *Chem. Biol.* **13**, 857 (2006).

## Chapter 6: Future Work

### 6.1 Single-molecule force spectroscopy of riboswitches

The work done for this thesis on the *pbuE* riboswitch aptamer has shown that individual steps in the folding can be resolved, and the kinetic and energetic properties of the intermediate states can be related to biological activity in a cellular context. There are many other important questions and issues that remain to be addressed about the folding of this riboswitch, purine riboswitches, and other types riboswitches in general, however, which will be the focus of future work.

First, this study has considered only the folding of the aptamer in isolation—the full riboswitch (with expression platform) has not yet been investigated as thoroughly with single-molecule methods. Similarly, the aptamer has previously been the subject of several mutational studies, but not addressing the full riboswitch (Lemay 2006, Gilbert 2007). We expect that there may be interesting opportunities to observe the competition between the kinetics of transcription, aptamer folding, and terminator hairpin formation. This competition may be explored by changing the rate of transcription (e.g. using low nucleotide-triphosphate (NTP) concentrations) or by mutating the sequence of the terminator hairpin, whether to change the folding rate (via changing the loop length, for example) or to change the thermodynamic stability of the hairpin. While attempts to observe differences between initial co-transcriptional folding and subsequent refolding were not successful when studying the aptamer alone, the full

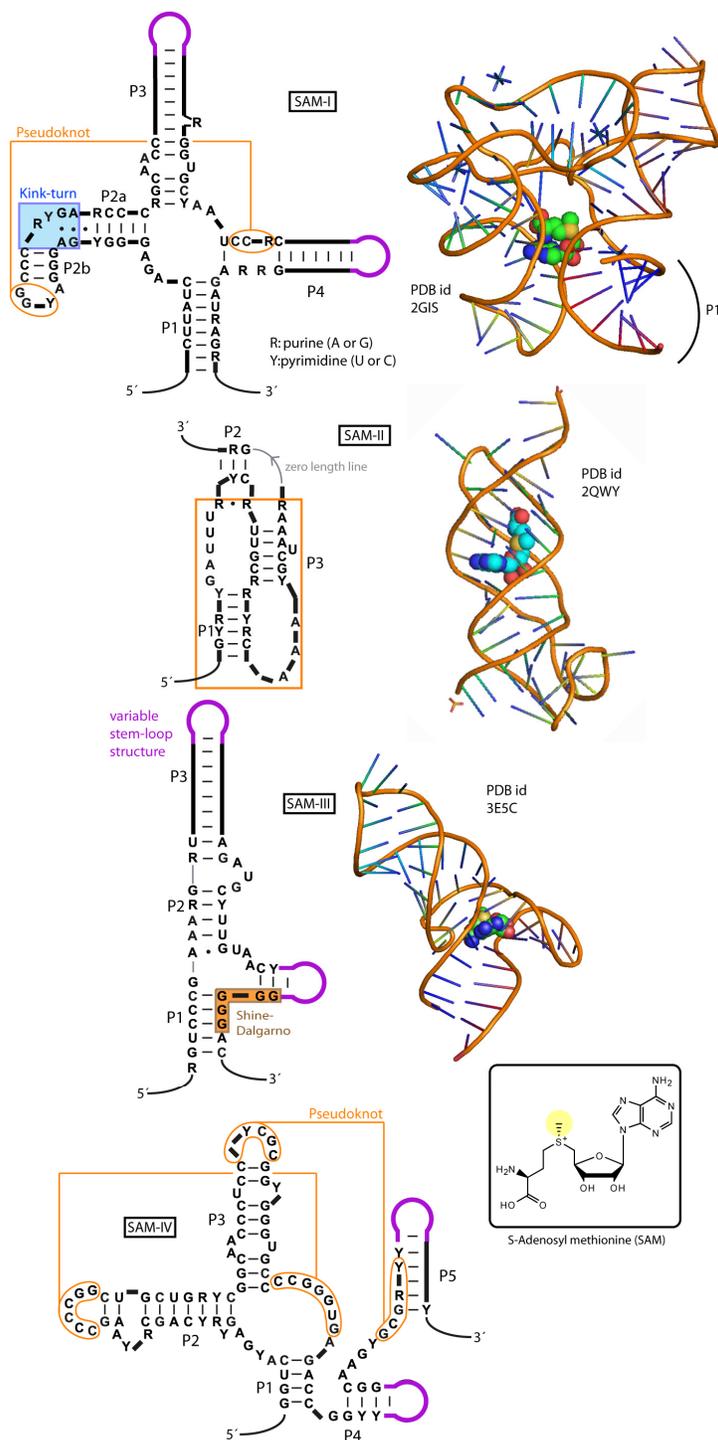
riboswitch would presumably be much more likely to show such a difference due to the importance of the polymerase kinetics for determining the folding outcome. Studies like these are being carried out by our collaborators at Stanford University. Additional studies could also be carried out on one of the interesting mutants with modified function that have been discovered by colleagues of D.A. LaFontaine's laboratory at Sherbrooke University (unpublished).

A second issue to investigate involves comparing the folding of the *pbuE* aptamer to the folding of other adenine riboswitch aptamers. As mentioned before, all known purine riboswitch aptamers have very similar tuning-fork structures, but there are significant functional differences: some regulate through transcription termination, others through anti-termination, and some at the level of translation. There are also some minor sequence and structural differences. A detailed comparison of the folding behaviour should provide insight into how the minor differences between the aptamers relate (if at all) to the functional differences.

One target of interest is the *add* adenine-riboswitch of *V. vulnificus* (Fig6.1). Unlike the *pbuE* riboswitch, it acts at the level of translation: the expression platform involves the formation of a hairpin that sequesters the Shine-Dalgarno sequence. This hairpin is much less stable than the terminator hairpin of the *pbuE* riboswitch, and the P1 helix with which it competes structurally (analogously to the *pbuE* riboswitch) is much more stable than P1 in the *pbuE* aptamer. It would be interesting to investigate whether the translational repression in the *add* riboswitch involves kinetic competition as with the transcription



structure and function in these molecules. Two riboswitches of particular interest include the thiamine pyrophosphate (TPP) riboswitch of *Arabidopsis thaliana* (Thore 2006) and the SAM-I riboswitch of *Thermoanaerobacter tengcongensis* (Montange 2006). The aptamer of this TPP riboswitch is based on a three-helix junction, like for the purine riboswitches, but the structure contains several additional components and involves an induced fit in the binding pocket (Thore 2008). The SAM-I riboswitch, on the other hand, is built around a four-helix junction, and a tertiary pseudoknot (Gilbert 2006, Wang 2008). These structures will offer a unique laboratory for understanding in detail the folding of relatively complex RNA molecules through single molecule force spectroscopy, pushing the current state of the art. Being more complex, they may also offer more opportunity to observe differences between RNAs folded co-transcriptionally and those refolded after transcription. Finally, there is a lot of structural diversity among different classes of the S-Adenosyl methionine (SAM) riboswitch (Gilbert 2008, Wang 2008, Lu 2008), all capable of binding the same ligand with high efficiency, presenting an interesting opportunity to study further how different structures relate to function (Fig 6.2).



**Figure 6.2** Secondary and crystal structures of SAM riboswitch aptamers. The consensus structures of the four known classes are shown. The first involves a four-helix junction; three involve pseudoknots in some fashion (Lu 2008, Wang 2008). Inset (bottom right): structure of SAM ligand

A final issue that could be addressed through experiments like these relates to the whether RNA folds in a strictly hierarchical fashion (Brion 1997), or whether the folding is better described as kinetic partitioning between multiple folding pathways (Thirumalai 2001). The first step will be to re-examine the data obtained from the *pbuE* aptamer, using careful statistical analysis to determine whether some small portion of the folding events involved multiple structures coming apart in a single event—i.e., not simply two-state transitions. The initial analysis showed that such events, if they occur for the *pbuE* aptamer, are quite rare. A careful analysis of the distribution of lifetimes will be needed to account correctly for “missing” events, in which an apparently non-two-state transition occurs because the lifetime of the intervening state is too short to register experimentally. Riboswitches with more complex structures will provide an even better source of data for such studies, given that many more intermediate states should be possible and observable.

The field of riboswitch folding thus holds a lot of promise, with many different research opportunities (Serganov 2009). Continued study should lead not only to a better understanding of how specific riboswitches function, but also to a deeper appreciation for how structure, folding, and function are related in RNA, as well as to insights into the fundamental biophysics governing RNA folding.

## 6.2 References

- Brion P, Westhof E (1997), *Hierarchy and dynamics of RNA folding*, Annu Rev Biophys Biomol Struct. **26**:113-37.
- Gilbert SD, Montange RK, Stoddard CD, Batey RT (2006), *Structural studies of the purine and SAM binding riboswitches*, Cold Spring Harb Symp Quant Biol. **71**:259-68.
- Gilbert SD, Love CE, Edwards AL, Batey RT (2007), *Mutational analysis of the purine riboswitch aptamer domain*, Biochemistry, **46**:13297-309.
- Gilbert SD, Rambo RP, Van Tyne D, Batey RT (2008), *Structure of the SAM-II riboswitch bound to S-adenosylmethionine*, Nat Struct Mol Biol. **15**:177-82.
- Kim JN, Breaker RR (2008), *Purine sensing by riboswitches*, Biol Cell. **100**:1-11.
- Lemay JF, Penedo JC, Tremblay R, Lilley DM, Lafontaine DA (2006), *Folding of the adenine riboswitch*, Chem Biol. **13**:857-68,
- Lu C, Smith AM, Fuchs RT, Ding F, Rajashankar K, Henkin TM, Ke A (2008), *Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism*, Nat Struct Mol Biol. **15**:1076-83.
- Montange RK, Batey RT (2006), *Structure of the S-adenosylmethionine riboswitch regulatory mRNA element*, Nature **441**:1172-5
- Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ (2004), *Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs*, Chem Biol. **11**:1729-41.
- Serganov A, (2009) *The long and the short of riboswitches*, Curr Opin Struct Biol. **19**:251-9.
- Thirumalai D, Lee N, Woodson SA, Klimov D (2001), *Early events in RNA folding*, Annu Rev Phys Chem. **52**:751-62.
- Thore S, Leibundgut M, Ban N (2006), *Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand*, Science. **312**:1208-11.
- Thore S, Frick C, Ban N (2008), *Structural basis of thiamine pyrophosphate analogues binding to the eukaryotic riboswitch*, J Am Chem Soc. **130**:8116-7.
- Wang JX, Breaker RR (2008), *Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine*, Biochem Cell Biol. **86**:157-68.