

# University of Alberta

Single-molecule studies of prion protein folding and misfolding

by

Hao Yu

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

Doctor of Philosophy

Department of Physics

©Hao Yu

Fall 2013

Edmonton, Alberta

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

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

## **Dedication**

In memory of my maternal grandfather,

Zhipan Qi

1926-2013

## Abstract

Protein folding involves a stochastic search through the configurational energy landscape towards the native structure. Although most proteins have evolved to fold efficiently into a unique native structure, misfolding (the formation of non-native structures) occurs frequently *in vivo* causing a wide range of diseases. The prion protein PrP has the unique ability to propagate an infectious disease without transmitting any genetic material, based instead on a misfolded conformation which can reproduce itself. The mechanism of prion misfolding and propagation remains unsettled, from details about the earliest stages of misfolding to the structure of the infectious state. Part of the difficulty in understanding the structural conversion arises from the complexity of the underlying energy landscape. Single-molecule methods provide a powerful tool for probing complex folding pathways as in protein misfolding, because they allow rare and transient events to be observed directly.

We used custom-built high resolution optical tweezers to study PrP one molecule at a time. By measuring folding trajectories of single PrP molecules held under tension, we found that the native folding pathway involves only two states, without evidence for partially folded intermediates that have been proposed to mediate misfolding. The full energy profile was reconstructed for the native folding of PrP, revealing a double-well potential with an extended partially-unfolded transition state. Interestingly, three different misfolding pathways were detected, all starting from the unfolded state. A mutant PrP with higher aggregation propensity showed increased occupancy of some of the misfolded states, suggesting these states may act as intermediates during aggregation. To investigate the mechanism of PrP misfolding further, we characterized the folding pathways of PrP when two molecules interact to form a dimer. Remarkably, the dimer invariably formed a stable misfolded structure, via multiple partially-folded intermediates.

We mapped the energy landscape for PrP dimer misfolding and identified a key intermediate that leads to misfolding by kinetically blocking the formation of the native structure. These results provide mechanistic insight into the formation of non-native structures of PrP and demonstrate a general platform for studying protein misfolding and aggregation at the single-molecule level, with wide applicability for understanding disease and biological function.

## Acknowledgement

First of all, I would like to express my deepest gratitude to my supervisor, Dr. Michael T. Woodside. He introduced me to the fantastic world of biophysics. He taught me to build the first optical tweezers in our lab, from the overall design of a high-resolution instrument to each detailed steps of alignment and measurement. My productive PhD career was inseparably inherited from his tremendous support over the years. As a foreign student, I am really appreciated his help in writing papers as well as the thesis. His great talent in biophysics inspired my research a lot and I look forward to make use of everything he taught into my future work.

I would also like to thank my committee members: Dr. Jack Tuszynski and Dr. Frank Hegmann. The feedbacks from them in each committee meeting have been invaluable over the past six years of my PhD program. I would like to extend my thanks to other examiners: Dr. Mark Freeman, Dr. David Wishart and especially Dr. Nancy Forde. I am really appreciated for their careful reading of my thesis and participation in my defense with lots of valuable questions and suggestions.

The research in this thesis would not have been possible without the contributions from all members of the Woodside's group. The optical tweezer system was built with the help of Daniel Foster, Zhiyuan Tang, Krishna Neupane and Miro Belov. The special talent of Daniel on electronics contributes an important half of the effort in building the instrument. As the first project after the birth of our optical tweezers, the riboswitch project was done together with Krishna Neupane, Daniel Foster and Feng Wang. Feng, an experienced molecular biologist, taught me a lot of experimental skills as I first started in the biochemistry labs. The days and nights spent in taking large amount of data are the most wonderful time inside the NINT building. The prion project was done together with

Xia Liu, Derek Dee, Amar Nath Gupta, Angela Brigley, Allison Solanki and Iveta Sosova.

Xia has contributed a lot in developing new experimental protocols for the handle attachment. All other group members, Cuauh Garcia-Garcia, Elena Grandi, Abhilash Vincent, Dustin Ritchie, Armin Hoffmann, Marion Becker, Bethany Griffith, Jingchuan Soong and Megan Engel, have been helpful in all aspects of my research. It is my great pleasure to work with such a big group of talented people. Special thanks to Armin Hoffman and Daniel Foster for their critical reading of my thesis.

Finally I will thank my parents and grandparents back in China. Their selfless love and encouragement over the years is the major spiritual support on my way to the PhD degree. And to Huilin, thank you for always being there with me. It is you who make my life in Edmonton more colorful and meaningful. It is you who make Edmonton such a significant and memorable place for me. Thank all of my friends for the wonderful life I live with you.

I dedicate this thesis to my grandfather. I can still remember the first English he taught me when I was young. His integrity and dignity during his entire life has influenced me a lot. May he rest in peace.

## Table of Contents

Dedication .....	
Abstract .....	
Acknowledgement .....	
List of Tables .....	
List of Figures .....	
List of Symbols, Nomenclature, or Abbreviations.....	
1. Molecular folding, misfolding and diseases.....	1
2. Biophysical description of the folding problem.....	10
2.1 Theoretical description: the energy landscape .....	10
2.2 Diffusion constant, transition path time, and rates.....	14
3. Single-molecule approaches to the folding problem.....	18
3.1 Advantage of single-molecule approaches.....	18
3.2 Summary of single-molecule methods.....	20
3.3 The effect of force in single-molecule force spectroscopy measurements.....	26
4. Optical tweezers.....	31
4.1 Principles of optical trapping .....	31
4.2 Construction of ultra-stable high-resolution optical tweezers system.....	35
4.3 Dumbbell assay .....	40
4.4 SMFS measurements and information to be learned.....	41
4.4.1 Non-equilibrium force-extension curve measurements.....	42
4.4.2 Equilibrium constant force measurements.....	51
Appendix: Optical trapping measurements .....	63
5. SMFS of the <i>add</i> adenine riboswitch relates folding to regulatory mechanism .....	67
5.1 Introduction of the <i>add</i> adenine riboswitch.....	67
5.2 RNA constructs .....	70
5.3 Folding intermediates of the aptamer .....	73
5.4 Energy landscape of the aptamer.....	81
5.5 Misfolded states in the aptamer.....	84
5.6 Full length riboswitch.....	88
5.7 Discussion .....	93
6. Prion misfolding and diseases.....	99
6.1 The prion hypothesis .....	99
6.2 PrP <sup>C</sup> native folding .....	100
6.3 PrP misfolding and aggregation .....	103

6.4 Mutations and prion diseases .....	106
6.5 Single-molecule studies of PrP .....	108
7. Direct observation of multiple misfolding pathways of PrP monomers .....	111
7.1 Sample preparation.....	111
7.2 PrP native folding is two-state.....	114
7.3 Identification of off-pathway misfolded states.....	123
7.4 Mutation studies .....	126
7.5 PrP folding/misfolding pathways and implications on aggregation.....	129
Appendix: experimental details.....	134
8. Energy landscape analysis of PrP native folding pathway: the diffusion constant, transition time and rates .....	136
8.1 Validation of the Hummer-Szabo method for energy landscape reconstruction .....	137
8.1.1 Two-state DNA hairpins.....	137
8.1.2 <i>add</i> adenine riboswitch aptamer with intermediate states .....	144
8.2 Energy landscape reconstruction of PrP native pathway .....	146
8.3 Energy landscape analysis yields the diffusion constant, transition path time and rates .....	151
8.4 Discussion .....	159
9. Structural conversion of prion protein dimers via a multi-step pathway observed at the single-molecule level .....	164
9.1 SMFS of PrP dimers and the observation of stable misfolded structures .....	164
9.2 Metastable sequential intermediates on the misfolding pathway .....	169
9.3 PrP dimer energy landscape .....	176
9.4 Topology of the PrP dimer .....	180
9.5 PrP dimer invariably misfolds.....	180
9.6 Structure of the misfolded dimer and relevance of the misfolded states.....	185
10. Conclusions and future directions.....	188
10.1 Summary and implications of the work .....	188
10.2 Future directions.....	193
10.2.1 SMFS of oligomers and aggregates.....	193
10.2.2 Disease-related mutations.....	199
10.2.3 Structural prediction in combination with computer simulations.....	201
10.2.4 Biological relevance of single-molecule results .....	203
Bibliography .....	206

## List of Tables

Table 3.1 Single-molecule techniques. ....	21
Table 5.1 Sequences of RNA molecules measured: <i>add</i> riboswitch aptamer, expression platform, full-length riboswitch, and oligomer sequences complementary to portions of the aptamer. ....	72
Table 5.2 Aptamer unfolding contour length changes. ....	76
Table 5.3 Constant force results. ....	80
Table 9.1 Structure, kinetic and stability parameters of the misfolded dimer and intermediate states. ....	167

## List of Figures

Figure 1.1 Complex folding/misfolding pathways accessible to a protein. ....	7
Figure 2.1 Energy landscape of protein folding/misfolding. ....	11
Figure 2.2 Schematic free-energy profile for a two-state folding reaction. ....	14
Figure 3.1 The effect of force on free energy landscapes. ....	27
Figure 4.1 Principles of optical trapping. ....	32
Figure 4.2 The trapping profile of optical tweezers. ....	34
Figure 4.3 Design of a dual-trap optical tweezers system. ....	36
Figure 4.4 Spatial resolution. ....	38
Figure 4.5 Experimental assay for the dual trap apparatus. ....	40
Figure 4.6 Summary of FEC analysis. ....	42
Figure 4.7 Force-extension curve measurements. ....	47
Figure 4.8 Force remains constant during constant-force measurements. ....	52
Figure 4.9 Summary of constant force analysis. ....	53
Figure 4.10 Signature of an intermediate state in constant-force measurements. ....	54
Figure 4.11 Lifetime determined from constant force measurements. ....	56
Figure 4.12 Point-spread function. ....	57
Figure 4.13 Instrument response function. ....	60
Figure 4.14 The distance and force calculations in SMFS measurements. ....	63
Figure 4.15 Force calibration using DNA hairpins for constant force measurements. ....	65
Figure 5.1 The structure of <i>add</i> adenine riboswitch. ....	69

Figure 5.2 Force spectroscopy of <i>add</i> aptamer alone. ....	74
Figure 5.3 Confirmation of hairpin folding identification using anti-sense oligos. ....	77
Figure 5.4 Unfolding force distributions and force-dependent rates determined from FECs. ....	82
Figure 5.5 Folding energy landscapes of the <i>add</i> aptamer. ....	84
Figure 5.6 Misfolded states. ....	86
Figure 5.7 Folding pathways of the aptamer. ....	87
Figure 5.8 FECs of the full-length riboswitch. ....	89
Figure 5.9 Switching and ligand binding. ....	92
Figure 5.10 Folding kinetics and riboswitch mechanism. ....	96
Figure 5.11 Schematic of the riboswitch mechanism. ....	97
Figure 6.1 The structure and sequence of PrP. ....	103
Figure 7.1 Gel electrophoresis of the handle attachment reaction. ....	112
Figure 7.2 Force spectroscopy measurements of PrP show apparent two-state folding. ....	115
Figure 7.3 On-pathway intermediates are not observed. ....	117
Figure 7.4 Simulation of intermediate states. ....	119
Figure 7.5 Constant-force records reveal misfolded states. ....	121
Figure 7.6 Misfolded states form via the unfolded state only. ....	123
Figure 7.7 Residuals for extension histogram fits. ....	124
Figure 7.8 Free energy of the misfolded states. ....	126
Figure 7.9 Aggregation of wild-type and C179A/C214A PrP. ....	127
Figure 7.10 Force-extension curves of C179A/C214A mutant. ....	128

Figure 7.11 Mutant PrP shows increased misfolding. ....	131
Figure 7.12 Schematic diagram of folding/misfolding pathways. ....	132
Figure 8.1 Optical trapping measurements of DNA hairpins. ....	138
Figure 8.2 FECs of DNA hairpins. ....	139
Figure 8.3 Free energy landscape reconstruction at zero force. ....	140
Figure 8.4 Comparison of equilibrium and non-equilibrium reconstructions. ....	142
Figure 8.5 FECs at different pulling rates. ....	143
Figure 8.6 Riboswitch aptamer landscape reconstruction. ....	145
Figure 8.7 Free-energy profile reconstruction for PrP folding. ....	147
Figure 8.8 Comparison of reconstructed landscape to WLC energies. ....	147
Figure 8.9 Point-spread function for landscape reconstruction. ....	149
Figure 8.10 Energy landscape deconvolution. ....	150
Figure 8.11 Force-extension curves of PrP unfolding. ....	152
Figure 8.12 Free energy estimation from Crooks theorem. ....	154
Figure 8.13 Constant-force trajectories of PrP folding. ....	155
Figure 8.14 Direct estimate of an upper bound for the transition time. ....	157
Figure 8.15 Comparison between experimental and predicted rates. ....	158
Figure 9.1 SMFS measurements of PrP dimers show the presence of stable misfolded states. ....	167
Figure 9.2 Intermediates on the misfolding pathway of PrP dimers. ....	171
Figure 9.3 Quasi-equilibrium transitions between on-pathway intermediates in the shoulder region of FECs. ....	174

Figure 9.4 Pathway and energy landscape for PrP dimer misfolding. ....	178
Figure 9.5 Topology of the PrP dimers. ....	179
Figure 9.6 Intermediate state $I_D1$ . ....	182
Figure 9.7 CD spectrum of PrP dimers. ....	184
Figure 10.1 SMFS of PrP trimers. ....	195
Figure 10.2 FECs of a PrP aggregate showing the complex aggregation pathway. ....	197
Figure 10.3 SMFS of monomeric H187R mutant molecule. ....	198
Figure 10.4 SMFS of F198S dimers showing a similar behavior as the wild-type dimers. .....	200
Figure 10.5 FECs of H187R dimers showing a different behavior from the wild-type dimers. ....	202

## List of Symbols, Nomenclature, or Abbreviations

AOD	acousto-optic deflector
AFM	atomic force microscopy
CD	circular dichroism
CJD	Creutzfeldt-Jakob disease
EOD	electro-optic deflector
FCS	fluorescence correlation spectroscopy
FEC	force-extension curve
FFI	fatal familial insomnia
FRET	Förster resonance energy transfer
GPI	glycophosphatidylinositol
GSS	Gerstmann-Sträussler-Scheinker syndrome
NMR	nuclear magnetic resonance
PMCA	protein misfolding cyclic amplification
PrP	prion protein
PrP <sup>C</sup>	cellular form of the prion protein
PrP <sup>Sc</sup>	scrapie form of the prion protein
PSD	position sensitive detector
SHaPrP	Syrian hamster PrP
SMFS	single-molecule force spectroscopy
TSE	transmissible spongiform encephalopathies
WLC	worm-like chain
U	unfolded state
N	native state
I	intermediate state
M	misfolded state
‡	transition state
$L_c$	contour length
$L_p$	persistence length
$K$	stretch modulus
$D$	diffusion constant
$\Delta G$	free energy difference

$\Delta G^0$	free energy difference at zero force
$\Delta G^\ddagger$	height of the energy barriers
$\Delta G^{\ddagger,0}$	height of the energy barriers at zero force
$\Delta x^\ddagger$	distance to the transition state
$\Delta x^{\ddagger,0}$	distance to the transition state at zero force
$\tau_{tp}$	transition path time
$\kappa$	curvatures of the energy profile
$k$	Kramers rate
$k_0$	Kramers pre-factor
$k_B$	Boltzmann constant
$k_{\text{fold/unfold}}$	folding/unfolding rate at zero force
$F_{1/2}$	equilibrium force in a two-state system
$F_{unf}$	unfolding force

# 1. Molecular folding, misfolding and diseases

Folding is the molecular self assembly process by which the linear one-dimensional protein chain takes on a specific three-dimensional structure. According to the central dogma of molecular biology, genetic information flows from DNA into RNA through transcription, and finally into protein through translation. However, the expression of genetic information does not end here. Proteins have to fold into specific three-dimensional structures in order to carry out their functions precisely. The folding process is thus required for the information stored in DNA to take effect, constituting an important step in the flow of genetic information (Koonin et al., 2002). Moreover, RNA, rather than being just an intermediate factor assisting protein synthesis, has been found to play important roles in regulating different biological processes (Castel and Martienssen, 2013; Serganov and Nudler, 2013). Finally, the structural dynamics of molecules like proteins and RNA also plays essential roles in biology (Agarwal, 2006; Al-Hashimi and Walter, 2008; Hall, 2008; Henzler-Wildman and Kern, 2007; Smock and Gierasch, 2009).

At the most basic level, sequence determines conformation and conformation determines function. Only correctly folded molecules have long-term stability and are able to carry out their biological function properly, except for intrinsically disordered proteins (Rezaei-Ghaleh et al., 2012), which remain functional despite the lack of unique three-dimensional structures. In contrast, improper folding resulting in non-native states (i.e. misfolding) can cause the dysfunction of the molecules. Unfolded, partially folded or misfolded molecules may be subjected to some housekeeping procedures (Araki and Nagata, 2012; Kubota, 2009; Pechmann et al., 2013; Tyedmers et al., 2010) inside organisms and degraded, except some special events in the cell, such as translocation

which requires complete or partial unfolding of a molecule (Matouschek, 2003; Prakash and Matouschek, 2004). However, misfolded molecules that escape the proper regulatory mechanisms may aggregate and induce diseases (Chiti and Dobson, 2006).

Understanding how macromolecules fold and misfold is thus a crucial part of deciphering the flow of genetic information, with important implications for understanding how cells function and origins of many diseases. These questions are discussed below in the context of two general aspects: first, the folding code and mechanisms; second, misfolding mechanisms and diseases.

### 1. Folding code and mechanisms

The first aspect of the folding problem involves how the information defining the three-dimensional native conformation is encoded in the one-dimensional sequence of a biopolymer. This question was addressed by Anfinsen in 1957, who showed that the native structure of a protein can be fully regenerated spontaneously, with its activity fully recovered, after denaturing and then renaturing the protein (Sela et al., 1957). In this pioneering work, the structure of a natively-folded enzyme, ribonuclease A, was fully denatured with the addition of 2-mercaptoethanol and urea. The catalytic activity was destroyed as the enzyme lost its native conformation. Upon the removal of the denaturant, the activity was essentially fully restored, indicating the spontaneous recovery of the native conformation. Based on this observation, it was hypothesized that all the information required to determine the three-dimensional structure of a protein is incorporated in the amino acid sequence. Anfinsen hypothesised furthermore that the native structure was the lowest-energy state of the system (Anfinsen, 1973). The question arises of whether there exists a universal code to store the information regarding the

three-dimensional conformation of proteins in their one-dimensional sequence, and how the information is encoded.

Many structures of proteins and nucleic acids have been solved by high-resolution structural determination methods such as nuclear magnetic resonance (NMR) and X-ray crystallography. However, these procedures are often time consuming and technically demanding, so that structures are available for only a relatively small proportion of the vast universe of possible proteins (Slabinski et al., 2007). Computational prediction of structures is possible, but not completely reliable (Moult, 2005), and physics-based structure prediction is still not very effective (Cao and Chen, 2011; Faver et al., 2011; Zhang, 2009). Molecular dynamics simulations can successfully predict the folding structures of some proteins, especially small ones (Piana et al., 2013; Shaw et al., 2010), but further improvements in computational capabilities will be needed to improve reliability for large proteins or macromolecular assemblies. Given the limited information we have connecting sequences to structures, compared to the combinatorial complexity of the possible sequences, the folding code has yet to be deciphered. Complicating the matter is the fact that there is likely no universal, one-to-one relationship between the primary sequence and three-dimensional structure of proteins; rather the structure formed by a given sequence is context-dependent (Minor and Kim, 1996; Schellenberg et al., 2010).

Another important aspect of the folding problem is how the structure forms, i.e. the folding mechanism. As Levinthal pointed out, the time required for a protein to fold into a specific conformation based on a random search would be enormously long because of the astronomical number of possible configurations assessable to a protein (Levinthal, 1968; Levinthal, 1969). Assuming simplistically that each amino acid has 3

possible conformations and that each conformation can be sampled in 1 ps (the time scale of elastic vibration and hydrogen bonding (Benkovic and Hammes-Schiffer, 2003)), a 100-amino-acid-long protein with a random sequence would take  $10^{28}$  years ( $3^{100} \times 10^{-12}$  s =  $5 \times 10^{35}$  s) to search the entirety of conformation space sequentially, very much longer than the age of the universe ( $\sim 10^{10}$  years). In practice proteins usually fold on the timescale of microsecond to seconds (Kubelka et al., 2004), or as slow as minutes to hours (Kim and Baldwin, 1990). This contradiction is the well-known “Levinthal’s paradox”, which indicates that a random search would not allow a protein to find its native structure on a physiological time scale.

Instead of a random search, proteins are generally believed to find their conformations quickly by folding along specific pathways through the set of a conformational landscape, biased by the interactions within the molecule and with the solvent (Frauenfelder et al., 2007; Thirumalai et al., 2010; Wolynes, 2008). However, the microscopic details of the transition from a disordered polypeptide chain into a highly ordered structure are still not fully understood. Classical biochemistry methods, such as circular dichroism (CD) spectroscopy, infrared spectroscopy, NMR spectroscopy and mass spectrometry, are able to provide important information about the structure and kinetics during the folding process (Buchner and Kiefhaber, 2005), although these methods generally cannot provide enough temporal resolution for observing the detailed interconversion along folding pathways. Stopped flow (Fabian and Naumann, 2004), temperature jump (Hofrichter, 2001) and hydrogen exchange techniques (Konermann et al., 2011) allow the direct observation of fast kinetics during the folding process (Schechter, 1970). However, these are ensemble measurements that observe the average behavior of many molecules. It can sometimes be difficult to discern intermediate,

partially-folded states (which can play important roles along the pathway to the final structure), especially if they are rare or short-lived. It can be even more difficult to distinguish between different populations, if different types of behavior occur at the same time.

## 2. Misfolding mechanism and diseases

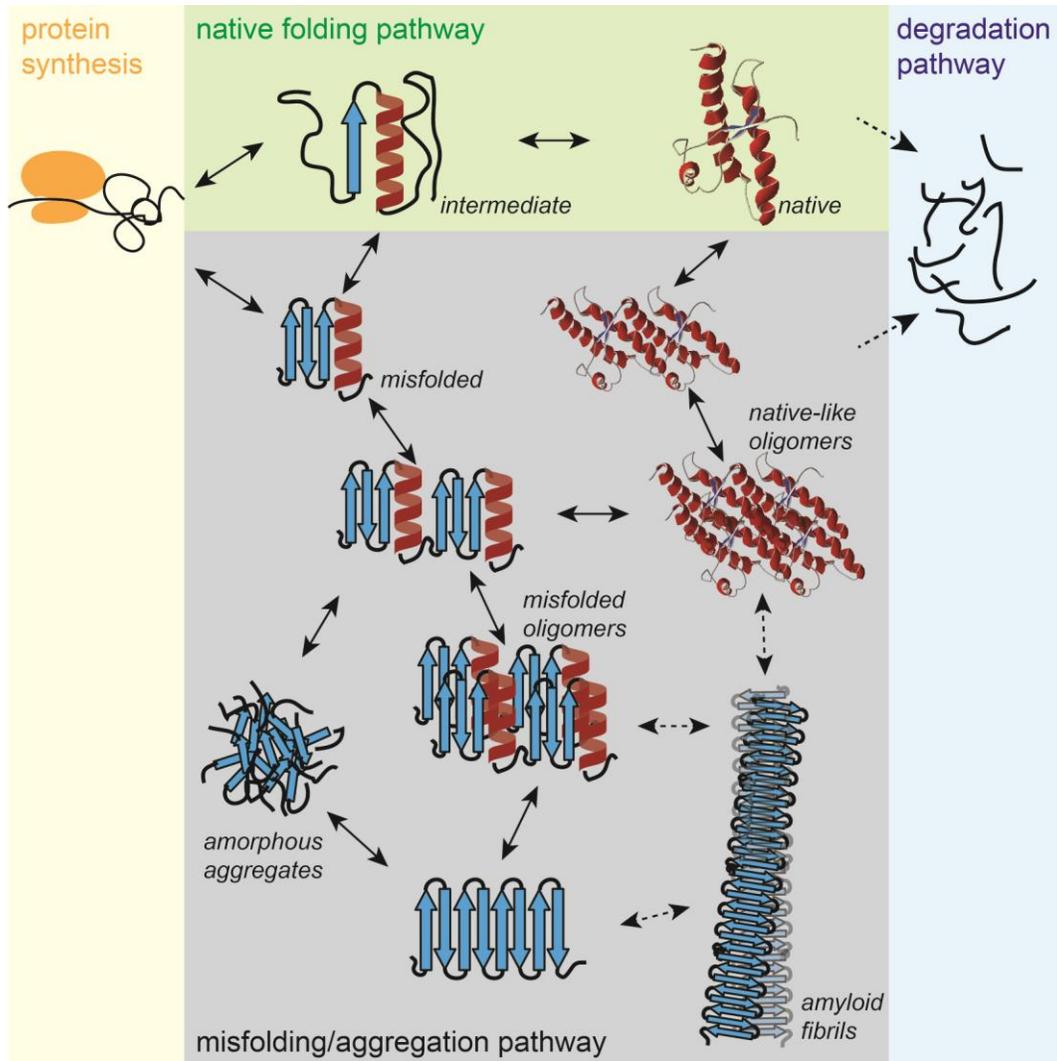
Although most proteins have evolved to maintain a specific conformation in order to carry out their functions properly, misfolding (the formation of non-native structures) still occurs frequently *in vivo* (Hartl and Hayer-Hartl, 2009). Cellular processes act to mitigate the effects of misfolding, *e.g.* by guiding folding with molecular chaperones or by removing misfolded proteins once they have formed through the action of the proteasome (Hartl and Hayer-Hartl, 2009; Tyedmers et al., 2010). Misfolded proteins, which escape such quality-control pathways, however, can lead to a wide range of diseases. Such “protein misfolding diseases” include, for example, Alzheimer’s disease, Parkinson’s disease, fronto-temporal dementia and the prion disorders, which are associated with the misfolding respectively of amyloid  $\beta$ ,  $\alpha$ -synuclein, tau and prion proteins (Chiti and Dobson, 2006).

The accumulation of misfolded protein molecules into large-scale aggregates is a common feature of numerous protein misfolding diseases, however in most cases the identity of the toxic species remains unresolved. Disease-related proteins can misfold and aggregate into oligomers, ultimately forming amyloid fibrils, insoluble aggregates of misfolded proteins which is usually composed of highly-ordered  $\beta$ -sheet-rich structures. The deposition of amyloid fibrils in different tissue types has been found in a large number of protein misfolding diseases (Chiti and Dobson, 2006), but the significance of

the amyloid fibrils in disease propagation is not clear. Exactly which species are toxic is still controversial, although it is believed that soluble oligomers early in the aggregation pathway are, in many cases, more toxic than the mature fibrils. In fact, the fibrils may actually serve a protective role by reducing the level of the toxic particles (Caughey and Lansbury, 2003; Chiti and Dobson, 2006).

Despite the importance of protein misfolding and aggregation to disease, the molecular mechanisms of the misfolding and aggregation processes remain imperfectly understood. Amyloid formation is often found to be consistent with a nucleated-growth model, in which a misfolding (or aggregation) nucleus forms slowly during an initial lag phase, after which amyloid polymerization proceeds rapidly (Chiti and Dobson, 2006). During this process, many different conformational states and folding pathways may, in principle, be accessible to a polypeptide chain. This is illustrated schematically in Fig. 1.1. Only the native state is able to function properly. If the quality control systems that regulate *in vivo* folding fail, the misfolding pathways can overwhelm the native folding pathways and cause diseases. Understanding the fundamental molecular events and the detailed mechanism underlying protein misfolding will help elucidate the origin of protein misfolding diseases and develop potential therapies. Part of the challenge in deciphering misfolding mechanisms stems from the fact that ensemble averaging makes it challenging to resolve the key features, since the initiating events are likely rare and entangled with further interconversions (Straub and Thirumalai, 2011). Few techniques have the ability to access the early nucleation stage and follow the complex aggregation process thereafter, leaving protein misfolding and aggregation a “black box” needing further exploration. New methods that are able to monitor transient protein states and

follow their fast interconversions are required to clarify the complex misfolding and aggregation pathways.



**Figure 1.1 Complex folding/misfolding pathways accessible to a protein.** After being synthesized by ribosomes, nascent polypeptides are subjected to different folding/assembly/disassembly pathways. Under normal conditions, proteins follow the native folding pathway into a fully functional structure, possibly through some intermediate states. *In vivo*, these processes are highly regulated and improperly folded proteins are subjected to the degradation pathways and other quality control systems. However, misfolding can occur. The protein may be able to form a variety of non-native conformations, including misfolded monomers, native-like oligomers, misfolded oligomers, amorphous aggregates and amyloid fibrils. Under certain conditions, misfolded or aggregated proteins may be toxic, causing diseases.

The prion diseases, which result from the misfolding of PrP, are notable among protein-misfolding diseases not only because they are infectious, but also because of their mode of infection (Aguzzi et al., 2007). A specific pathogenic conformation of PrP (the scrapie form, PrP<sup>Sc</sup>) is able to recruit and template the conformational change of the cellular form of PrP (PrP<sup>C</sup>). Therefore, the information required for the propagation of a prion disorder is not genetic but rather, a protein conformation. Recent studies have suggested that such conformational infectiousness may not be unique to PrP. For instance, prion-like propagation (Brundin et al., 2010; Frost and Diamond, 2010; Goedert et al., 2010) seems to occur in several neurodegenerative diseases such as Alzheimer's disease (Stohr et al., 2012), Parkinson's disease, fronto-temporal dementia, amyotrophic lateral sclerosis (Grad et al., 2011; Munch et al., 2011). As with the prion disease, all of these diseases are related with one or more types of amyloid-forming proteins: amyloid  $\beta$ ,  $\alpha$ -synuclein, tau and superoxide dismutase-1. Although not infectious, these diseases share similar features with the prionopathies, such as the accumulation and spread of misfolded protein aggregates, as well as a diversity of disease phenotypes. Therefore, the propagation of protein misfolding and neuropathogenicity in these diseases may occur in a similar prion-like manner. These similarities suggest that a detailed understanding of the protein misfolding/aggregation pathway of one protein could provide insight into the misfolding of other proteins as well.

In this thesis, I focus on PrP as a model system to study protein folding and misfolding mechanisms in the context of a disease-causing protein. A basic conceptual biophysical picture of folding and misfolding will be introduced in chapter 2, followed by two chapters talking about the methodology – single-molecule approaches (chapter 3), and specifically single-molecule force spectroscopy (SMFS) using optical tweezers

(chapter 4). Chapter 5 demonstrates SMFS as an approach to measure the hierarchical folding/misfolding of a molecule using a system with alternative structural outcomes, namely RNA riboswitches. The rest of the thesis focuses on SMFS studies of folding and misfolding of PrP. Chapter 6 describes the current background on prion research. Chapter 7 presents the SMFS results of PrP, showing the direct observation of folding and misfolding of single PrP monomers. In Chapter 8, we first discuss the validation of the Hummer-Szabo formalism for energy landscape reconstruction, and then use this method to recover the free energy profile of the PrP native folding pathway, yielding the diffusion constant, transition path time and rates. In Chapter 9, we take the first step to studying PrP aggregation using individual dimers, which we found to form exclusively misfolded structures. I conclude the thesis with an outlook on single molecule approaches to prion research and discuss some considerations for future research.

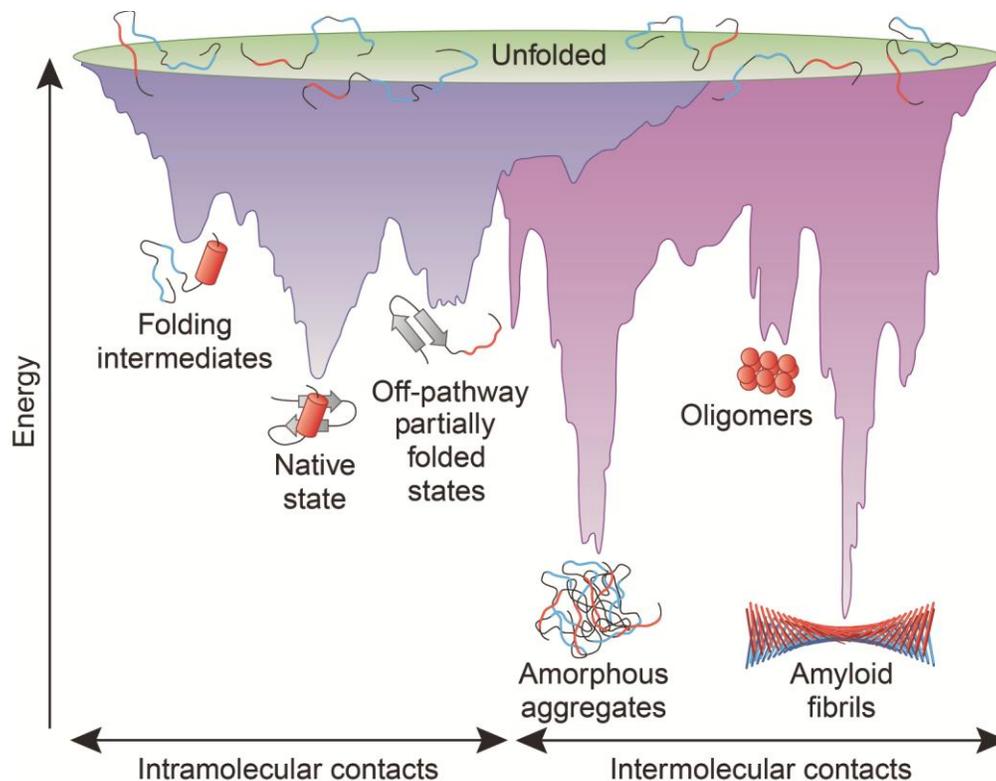
## **2. Biophysical description of the folding problem**

The folding problem can be understood in terms of the energy landscape theory. The energy landscape encoded by the amino acid sequence contains, in principle, all the information needed to describe the conformational dynamics of a protein, from the folding kinetics to the locations of energy barriers and the existence of intermediates or non-native pathways. The detailed characterization of energy landscapes provides a quantitative view of the protein folding problem. Fast native folding is typically viewed as the consequence of a funneled shape of the energy landscape (Dill and Chan, 1997; Wolynes, 2008), whereas misfolding and aggregation can result from partially folded states which are kinetically trapped by high barriers in a rugged landscape (Chiti and Dobson, 2009; Hartl and Hayer-Hartl, 2009).

### **2.1 Theoretical description: the energy landscape**

The energy landscape (Bryngelson and Wolynes, 1987; Dill and Chan, 1997; Onuchic and Wolynes, 2004; Wolynes, 2008) describes the free energy of each conformational state a molecule can adopt as a function of conformational coordinates, such as the dihedral angles of each peptide bond. The folding process then involves diffusion across this conformational landscape. Since the number of microscopic degrees of freedom for a given polypeptide is large, energy landscapes are typically high-dimensional hypersurfaces. A random polypeptide sequence will generally result in a very rugged landscape, where the protein is highly “frustrated” as the global minimum is difficult to reach kinetically (Dill and Chan, 1997; Wolynes, 2008). This is the origin of the Levinthal paradox. However, evolution has generally selected for proteins that fold rapidly along well-defined pathways, avoiding the need for a random search of the entire

conformational space. This can be achieved based on a funnel-shaped energy landscape with minimum frustration.



**Figure 2.1 Energy landscape of protein folding/misfolding.** The purple surface shows a multitude of conformations funnelling to the native state via intramolecular contacts, whereas the pink area shows conformations moving toward amorphous aggregates or amyloid fibrils via intermolecular contacts. The complexity of the folding pathways for aggregation is indicated by the complex energy profile with numerous free energy minima and maxima, representing different stable or metastable states with transition barriers in between. The figure is adapted by permission from (Hartl and Hayer-Hartl, 2009).

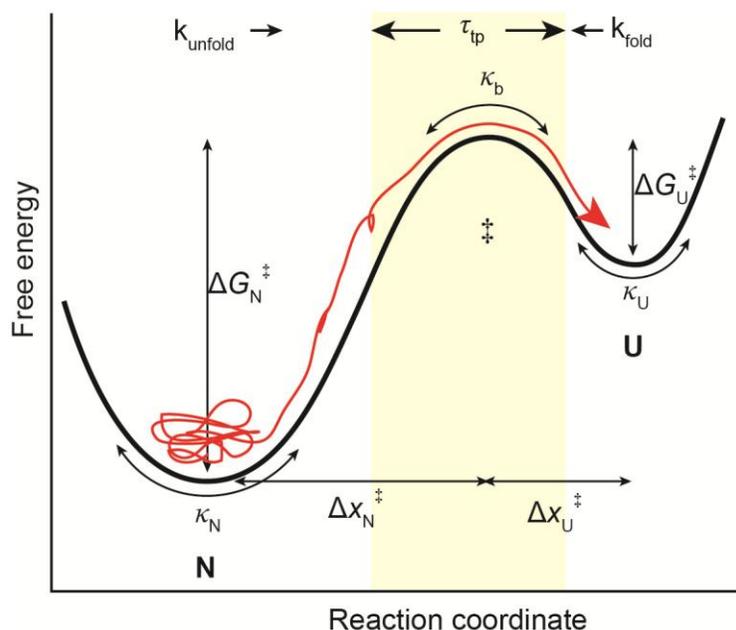
Consider folding in a funnelled energy landscape (Fig. 2.1). An unfolded protein molecule is randomly coiled, with a high conformational entropy. The funnel will guide the molecule towards changes in conformation that decrease the internal energy by making favorable intramolecular interactions. The bias to the conformational diffusion provided by these interactions speeds up folding exponentially, resolving Levinthal's

paradox (Zwanzig et al., 1992). As interactions are formed, however, there is a decrease in entropy as the conformational degrees of freedom are restricted. This competition between entropy and energy usually funnels the free energy through a barrier, explaining why in most cases protein folding is an activated process (Lorimer, 1996).

As the energy landscape becomes more rugged, incompletely folded metastable states start to appear, either on or off the native folding pathway (Fig. 2.1). On-pathway intermediate states usually represent mileposts leading to the correct folding of the native structure. Off-pathway misfolded states, which are often kinetically trapped by high barriers in the energy landscapes or involve contacts between different molecules, could result in further misfolding and aggregation. The folding/misfolding network of a polypeptide in Fig. 1.1 is quantified by energy landscape theory, resulting in a complex energy profile represented in Fig. 2.1. The conformation of a polypeptide chain under a given condition depends on the relative thermodynamic stabilities of different states and the kinetics of the interconversions between the states. Therefore, characterizing a complete picture of the energy landscape, including all partially-structured states, helps us to understand the folding pathways and misfolding mechanisms of a protein. However, because these partially-structured conformations are often energetically unfavourable, they are usually short-lived or rare, which raises technical difficulties for detection and characterization.

If protein folding could be measured by tracking the position of each atom in the molecule in real time, then the full multi-dimensional energy landscape could be measured. This is not possible with current experimental methods, hence the full landscape can only be studied virtually, through computational simulations (Piana and Laio, 2007; Piana et al., 2013; Shaw et al., 2010). One of the approaches used to

overcome this difficulty and treat experimental data more completely is to project the high-dimensional energy landscape onto the specific experimental reaction coordinate measured. This projection reduces the landscape to a one-dimensional profile (Bryngelson et al., 1995; Piana et al., 2013; Sabelko et al., 1999; Shaw et al., 2010). Examples of reaction coordinates include the extension of the molecule (Best et al., 2008), an experimental observable, or the fraction of native contacts, which is especially useful for simulations. Under normal experimental conditions of constant temperature and pressure, the Gibbs free energy is more convenient to describe the properties of the system than the Helmholtz free energy (Pathria, 1996). In this framework, the reduction in entropy as the molecule folds along the funnelled landscape towards a minimum of free energy competes with the enthalpy of intra-chain interactions. In some cases, the entropy loss during folding is compensated sufficiently by a corresponding internal energy gain so that the molecule reaches the minimum-energy state without encountering any barriers. Molecules that fold without barriers are known as “downhill folders” (Bryngelson et al., 1995; Onuchic et al., 1997). More commonly, however, a free energy barrier appears because the entropy loss and the energy decrease are not synchronized (Fig. 2.2). This barrier represents the “transition state,” the highest-energy, unstable intermediate through which the molecule must pass during folding. In the case of a single barrier separating two states, the transition state represents the decision point for committing to the forward or reverse reaction (in the context of classic chemical physics (Truhlar et al., 1996), and it also defines the rate-limiting step during the folding process. The height of the transition state barrier is related to the stability of the molecule (both energetic and kinetic).



**Figure 2.2 Schematic free-energy profile for a two-state folding reaction.** A typical two-state protein folding between the native state (N) and the unfolded state (U) across the transition barrier (‡) is characterized by the diffusion constant over the energy barrier ( $D$ ) and the shape of the energy landscape in terms of the height of the energy barriers ( $\Delta G_{U^\ddagger}$  from the unfolded state and  $\Delta G_{N^\ddagger}$  from the folded state, respectively), the position of the energy barriers ( $\Delta x_{U^\ddagger}$  from the unfolded state and  $\Delta x_{N^\ddagger}$  from the folded state, respectively) and the stiffnesses of the potential wells ( $\kappa_U$  for the unfolded state and  $\kappa_N$  for the folded state, respectively) and barrier ( $\kappa_b$ ). The folding rate given by Kramers theory ( $k$ ) is set primarily by the barrier height, which determines the length of time spent diffusing within the potential well waiting for a sufficiently large thermal fluctuation to cross the barrier. The transition path time required to cross the barrier ( $\tau_{tp}$ ) is much shorter, and is determined primarily by the diffusion constant across the barrier.

## 2.2 Diffusion constant, transition path time, and rates

Energy landscape theory conceptualizes folding as configurational diffusion over the surface of the landscape. A molecule diffuses within the folded/unfolded wells and across the barrier as it unfolds/folds (Fig. 2.2). In this picture, the molecule spends most of its time diffusing within the potential wells, with only rare and fast transitions between them. Unfortunately, key information about how the molecule folds is contained in the

transition path, the actual path taken while crossing the barrier (Fig. 2.2, yellow highlighted part). It is within this path that the effective folding processes during which the molecule explores different intermediate conformations occur, forming and breaking native/non-native contacts and finally reaching the folded state. As a single-molecule property, the transition paths themselves are extremely challenging to observe experimentally since they are very brief (Chung et al., 2009; Chung et al., 2012; Lee et al., 2007). Until recently, only all-atom simulations could provide insight into the mechanistic details of the transition paths (Piana et al., 2012, 2013; Shaw et al., 2010), and these were limited to very simple proteins (Best, 2012). However, the average time spent during the transition path, the transition path time  $\tau_{ip}$ , has now started to become accessible experimentally. To date,  $\tau_{ip}$  has been estimated for protein folding from single-molecule fluorescence trajectories (Chung et al., 2009; Chung et al., 2012), but it has not yet been calculated from folding landscapes themselves.

The results from chemical kinetics, in particular Kramers' theory for diffusive motion in a harmonic landscape (Hanggi et al., 1990; Kramers, 1940), can be used to relate the kinetics of protein folding to the underlying energy landscape. Kramers theory has now been widely applied to interpret the timescales for protein conformational dynamics (Best and Hummer, 2010; Dudko et al., 2006; Kubelka et al., 2004; Nettels et al., 2007; Plotkin and Onuchic, 2002). The folding/unfolding rate, which primarily depends on the timescale of the molecule spent inside a well (Fig. 2.2), can be expressed as:

$$k = k_0 \exp(-\Delta G^\ddagger/k_B T), \text{ where the prefactor is } k_0 = \frac{\sqrt{\kappa_w \kappa_b}}{2\pi k_B T} D, \quad (2.1)$$

in which  $\Delta G^\ddagger$  is the height of the potential barrier,  $D$  is the diffusion constant over the barrier,  $\kappa_w$  is the stiffness (curvature) of the potential well,  $\kappa_b$  is the stiffness of the barrier, and  $k_B$  is the Boltzmann constant. In this Kramers rate equation, the pre-factor  $k_0$  gives a measure of the fastest rate for folding, that is when there is no barrier ( $\Delta G^\ddagger = 0$ ).

Therefore,  $D$  effectively describes the roughness of the landscape and sets the fundamental “speed limit” for protein folding (Kubelka et al., 2004). The rate of contact formation within short peptides of various lengths has been measured with energy transfer (Bieri et al., 1999) and quenching (Krieger et al., 2003; Lapidus et al., 2000), as has the intrachain diffusion constant for a variety of naturally-folding proteins, but only in their denatured or nearly-denatured states (Borgia et al., 2012; Chen et al., 2009; Hagen et al., 1996; Nettels et al., 2007; Soranno et al., 2012). Importantly, these approaches do not probe chain dynamics as they are approaching the transition state, when hydrophobic collapse has likely occurred or structure is partly formed. Since the diffusion constant is generally coordinate-dependent and for proteins with high energy barriers the folding rate depends mostly on the diffusion constant at the top of the barrier (Best and Hummer, 2010), proper characterization of the diffusion over a barrier could be significant in terms of understanding the protein folding problem, which has never been done experimentally.

The transition path time,  $\tau_{\text{tp}}$ , may also be related to the shape of the energy landscape, by calculating the mean first passage time (Chaudhury and Makarov, 2010; Chung et al., 2009; Hummer, 2004). For an harmonic barrier with  $\Delta G^\ddagger > 2 k_B T$  (Chung et al., 2009),

$$\tau_{\text{tp}} \approx \frac{\ln(2e^\gamma \Delta G^\ddagger / k_B T)}{D \kappa_b / k_B T} = \frac{\ln(2e^\gamma \Delta G^\ddagger / k_B T)}{2\pi k_0 \sqrt{\kappa_b / \kappa_w}}, \quad (2.2)$$

where  $\gamma$  is Euler's constant and the expression becomes exact in the limit of a large barrier height. Generally  $\tau_{\text{tp}} \ll k^{-1}$ , since a significant amount of time is spent diffusing within the potential well before the barrier is actually crossed. It is interesting to note the differences in the dependence of the folding rate and the transition path time on the barrier. The folding rate is highly sensitive to the barrier height (in fact exponentially-dependent), since the molecule must wait for a thermal fluctuation that is sufficiently large to get over the barrier. As the barrier gets higher, the molecule will spend more time trapped in the potential well, resulting in a lower folding rate. However, since the transition path time only takes into account successful transition events, it is much less sensitive to the barrier height, depending only logarithmically on the height. To date, these landscape theories have been used to predict protein folding rates, but only using landscapes derived from simulations (Kubelka et al., 2004), because free-energy profiles are difficult to measure. Free energy landscapes measured experimentally (Gebhardt et al., 2010; Woodside et al., 2006a) must be integrated with the prediction of folding rates in order to validate the corresponding theories for further investigations.

### **3. Single-molecule approaches to the folding problem**

#### **3.1 Advantage of single-molecule approaches**

Protein folding has been studied for many decades with biochemical and biophysical methods such as fluorescence spectroscopy, CD spectroscopy, neutron scattering, and NMR. These methods look at the behavior of an ensemble of molecules in solution. Such ensemble biochemical and biophysical measurements have provided the bulk of our knowledge of folding (Buchner and Kiefhaber, 2005), but they have some significant limitations. In particular, the ensemble averaging can make it challenging to characterize the properties of states that occur rarely or are very short-lived. It is also challenging to distinguish the properties of any subpopulations present in a mixture of states. In recent years, approaches that study the folding behavior of individual molecules have been used to overcome these challenges, complementing the more traditional methods and offering ever more detailed pictures of the microscopic behavior (Borgia et al., 2008; Woodside et al., 2008). By monitoring one molecule at a time, such single-molecule techniques provide a promising new approach in understanding protein folding, misfolding and aggregation.

Single-molecule techniques provide a unique window into folding reactions because of the ability to measure properties of the free energy landscape directly. Single reaction paths for each individual folding event can be monitored independently, allowing the observation of specific pathways along the energy landscape and the full distribution of kinetic behaviors. Changes in the molecular configurations can be measured with sub-nanometer resolution as the molecule unfolds and refolds, allowing the structural elements involved in each folding event to be determined (Greenleaf et al., 2007). The

folding energies and rates for each step along the folding pathway can be characterized with high precision. The diffusion constant for the denatured (Borgia et al., 2012; Chen et al., 2009; Hagen et al., 1996; Nettels et al., 2007; Soranno et al., 2012), transition, and folded states (Hinczewski et al., 2010), can in principle be resolved separately, providing detailed insights into the structural dynamics of the molecule at different stages of the folding process. The transition path time, which is a single-molecule property, has only started to be measured experimentally using high-resolution single-molecule techniques (Chung et al., 2009; Chung et al., 2012). All information about a structural transition can be explored and analyzed from the energy landscape, the full profile of which can be reconstructed directly from single-molecule techniques. In this way, single-molecule techniques provide a more quantitative and physical description of the biological folding problem in terms of the energy landscapes.

Single-molecule techniques provide insights and detailed information about the complex picture of protein misfolding and aggregation as well (Hoffmann et al., 2013; Yu et al., 2013). The many different conformations involved, the numerous alternative pathways, and the likely importance of rare or transient states (Fig. 1.1) all pose key technical challenges for characterizing misfolding mechanisms. Determining how they interconvert and their position along the folding/aggregation pathways is a key challenge in aggregation studies. Single-molecule approaches are well-suited to overcome these challenges: not only are they well-established for studying protein folding mechanisms (Borgia et al., 2008), but they can directly characterize distinct subpopulations, map out folding pathways, and observe rare or transient states (Hoffmann et al., 2013). A particular concern in the context of aggregation is that the concentration required to achieve a sufficient signal strength from a given species might be incompatible with

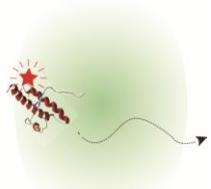
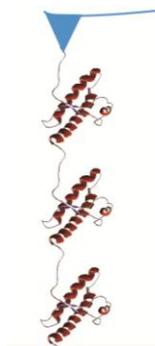
measurement because it triggers rapid aggregation. In the single-molecule regime, single protein monomers, oligomers, or even aggregates can be isolated to achieve a local concentration of the protein monomer in an oligomer construct as high as 10 M (e.g. as the effective concentration of proteins linked together covalently), leading to the formation of intermediate species via intermolecular interactions (Hoffmann et al., 2013). At the same time, the overall protein concentration in solution can be monitored from as low as 1 pM to high molar concentration to mimic different stages of the aggregation process (Fung et al., 2003). A detailed picture of aggregation mechanisms can thus be built up step-wise.

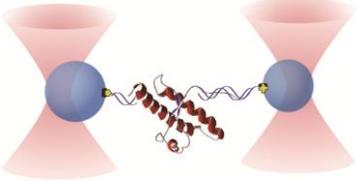
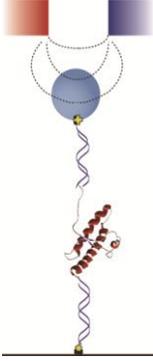
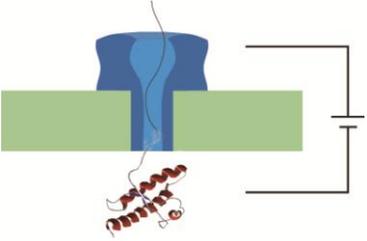
Nowadays, single-molecule spectroscopy is just beginning to be applied to characterize protein misfolding and aggregation. For example, misfolding has been observed in a few protein constructs using both force (Bechtluft et al., 2007; Oberhauser et al., 1999; Stigler et al., 2011; Xi et al., 2012; Yu et al., 2008) and fluorescence spectroscopies (Borgia et al., 2011). Likewise the formation and growth of aggregates has been monitored with fluorescence spectroscopy (Hillger et al., 2007; Nath et al., 2010; Orte et al., 2008). The network of pathways available for misfolding has only begun to be mapped out in any detail (Stigler et al., 2011; Xi et al., 2012), but not yet for any disease-related protein.

### **3.2 Summary of single-molecule methods**

A variety of techniques is available to monitor the structure, stability, and dynamics of molecules at the single-molecule level, based most commonly on fluorescence or force measurements (Table 3.1).

**Table 3.1 Single-molecule techniques.** The basic principles, method of measurement, and requirements for protein samples are outlined for several commonly used single-molecule techniques (Förster resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS), atomic force microscopy (AFM), optical tweezers, magnetic tweezers and nanopores).

	<b>Principle</b>	<b>Measurement</b>	<b>Requirements</b>
<b>FRET</b>	<p>Molecules are labeled with a pair of fluorophores which can exchange energy resonantly. The strength of the energy exchange depends on the inter-fluorophore distance.</p> 	<p>The FRET efficiency is measured as the conformation of the molecule changes. FRET efficiency is high when the molecule is folded so that the dyes are close together, whereas it is low for structures in which the dyes are far apart. FRET efficiency can be used as a molecular ruler to determine the distance between the dyes.</p>	<p>Specific labeling of protein with donor and acceptor dyes at different locations.</p>
<b>FCS</b>	<p>Dye-labeled molecules diffusing freely through a confocal excitation volume generate a fluctuating fluorescence signal.</p> 	<p>Correlations in the fluctuating signal produced by the diffusing molecules are analyzed to determine the diffusion constant for the molecule, which is related to its size and shape.</p>	<p>Specific labeling of protein with fluorescent dyes.</p>
<b>AFM</b>	<p>A molecule is tethered between a flat surface and a sharp tip on the end of a flexible cantilever. The tip is moved to apply force to the molecule and stretch it out.</p> 	<p>The force on the molecule is measured from the cantilever deflection, while the molecular extension is measured from the tip-surface distance. When the structure changes under tension, there is a sudden change in the extension.</p>	<p>Protein attached specifically or non-specifically to surface and AFM tip.</p>

<p style="text-align: center;"><b>Optical tweezers</b></p>	<p>A molecule is tethered between small beads trapped by tightly-focused laser beams. The laser beams are moved apart to apply force to the molecule and stretch it out.</p> 	<p>The extension of the molecule is measured from the distance between the beads, whereas the force is measured from the displacement of the beads in the traps. Conformational transitions change the molecular extension.</p>	<p>Protein attached covalently to DNA handles, in turn attached specifically to beads.</p>
<p style="text-align: center;"><b>Magnetic tweezers</b></p>	<p>A molecule is tethered between a flat surface and a superparamagnetic bead, held in magnetic fields. The trapping force is generated from the large field gradient.</p> 	<p>The position of the bead is tracked by video microscopy. The extension of the molecule was measured from the height of the bead, which is extract from the diffraction pattern. The force is calibrated from the thermal motion of the bead.</p>	<p>Protein attached covalently to DNA handles, in turn attached specifically to the bead and surface.</p>
<p style="text-align: center;"><b>Nanopore</b></p>	<p>A molecule is driven through the nanopore by electronic current. Confined by the geometry of the pore, the molecule may be forced to unfold before translocation.</p> 	<p>Current fluctuations are monitored as a molecule goes through the pore. The current depends on the folding as well as the sequence of the molecule inside the pore.</p>	<p>A nanopore with membrane and charged molecules</p>

## 1. Fluorescence-based techniques

FRET between co-localized fluorophores is a very sensitive probe of conformation and dynamics (Joo et al., 2008). Molecules are labeled by a donor fluorophore and an acceptor fluorophore, between which energy transfer can occur via resonance due to the overlap of the absorption spectrum of the acceptor and emission spectrum of the donor. The distance between the two fluorophores can be monitored based on the FRET efficiency, which can be determined from the donor and acceptor fluorescent intensities. Chemical denaturation such as urea or guanidine hydrochloride is typically used to manipulate the conformation of the molecule, and the distance changes estimated from the FRET efficiency represent the conformation dynamics. For example, proteins labeled with two different fluorescent dyes usually have smaller FRET values in their unfolded and larger FRET values in their native state. In combination with techniques such as total internal reflection microscopy, FRET can be imaged at single-molecule level. As a highly sensitive method, single-molecule FRET is typically used to probe distance changes between 1 to 10 nm, over timescales usually ranging from  $10^{-3}$  to  $10^2$  s (Greenleaf et al., 2007). Single-molecule FRET has been applied widely to study both protein (Borgia et al., 2008) and RNA folding (Tinoco et al., 2010).

Another common fluorescence-based technique is FCS. In FCS, the diffusion of fluorophore-labeled molecules across a small confocal volume results in fluorescence fluctuations that are temporally correlated. The autocorrelation of the fluorescence intensity contains information about the size and the shape of the molecule (Fitzpatrick and Lillemeier, 2011). Therefore FCS can be used to observe the formation and growth of structure formation during early aggregation process, for example, the multi-phase aggregation of PrP initiated by an early dimerization process has been observed using

FCS in conjunction with other methods (Post et al., 1998). However, unless immobilized, each molecule passes the effective volume only for a short time, so FCS reports the collective behavior of an ensemble of molecules and the ability for tracking the kinetics of individual molecules is limited.

## 2. Force-based techniques

Force can be applied directly to single molecules using different techniques, such as AFM, optical tweezers, or magnetic tweezers (Greenleaf et al., 2007; Neuman and Nagy, 2008). Typically the molecule of interest is attached to large objects to allow the application of forces and manipulation of structures. AFM utilizes a sharp cantilever tip, the position of which is monitored via deflected laser beams, to apply force to a molecule. Because of the high stiffness of the cantilever probe, AFM works best at large forces, typically in the range 10 to 10,000 pN, which makes it most suitable to study high-force phenomena such as breaking of covalent interactions, unfolding of stable proteins, and unfolding of aggregates. Magnetic tweezers use a superparamagnetic particle attached to a molecule to apply force with a magnetic field (Neuman and Nagy, 2008). In contrast to AFM measurements, with magnetic tweezers the applied force is constant regardless of the motion of the force probe, since the length scale over which the magnetic field gradient changes (~mm) is much larger than the scale of the motion (~nm), making magnetic tweezers good for monitoring structural transitions under equilibrium constant-force conditions with typical force range from 0.01 to 10,000 pN (Neuman and Nagy, 2008).

Optical tweezers employ dielectric beads trapped in electromagnetic fields to manipulate the molecule of interest (Woodside and Valentine, 2009). The stiffness of

optical tweezers is generally lower than for an AFM, and therefore optical tweezers are more suitable to study processes in the low-force regime (0.1 to 100 pN). Optical tweezers gives the best temporal resolution ( $10^{-4}$  to  $10^3$  s) among all the force measurement techniques, with very high spatial precision (as low as 0.1 nm). Optical tweezers have been successfully used to study a wide range of biological processes, such as the motion of molecular motors and the mechanisms of enzymes, including the transcription and translation process at the single-molecule level (Abbondanzieri et al., 2005; Wen et al., 2008). The folding of DNA, RNA and proteins has been studied extensively by this technique as well (Borgia et al., 2008; Li et al., 2008; Moffitt et al., 2008). However, optical tweezers are limited by the highest force that can be applied, which is determined by the laser intensity, the optical geometry, the index of refraction of the samples, and various aspects of the experimental protocols.

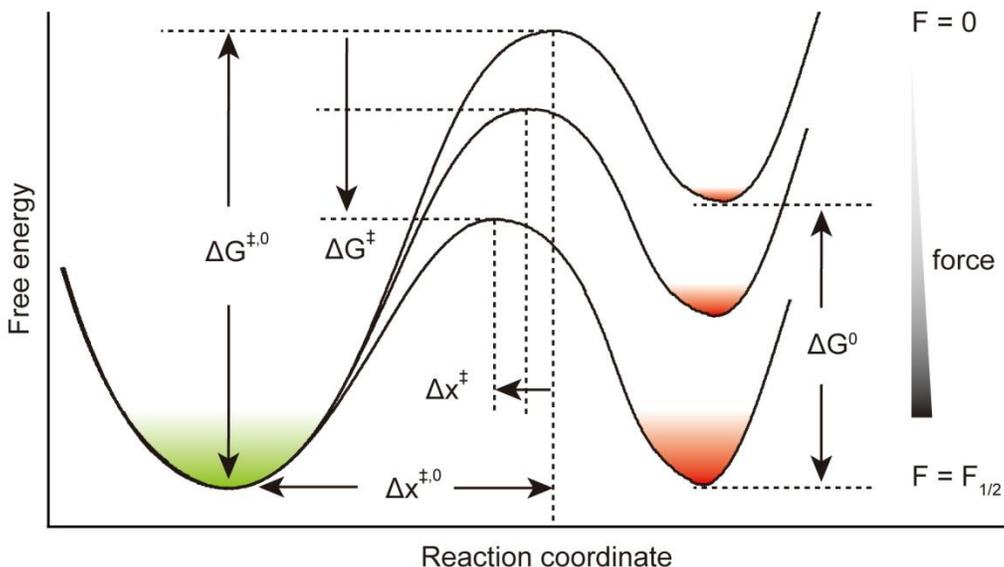
Force and fluorescence techniques have been combined to generate force probing systems with FRET-detection capability, such as FRET-based optical tweezers (Hohng et al., 2007) or magnetic tweezers (Hugel et al., 2007). These new hybrid systems allow simultaneous observation of different variables describing the conformation of the molecule, such as the force and distance from optical tweezers and the FRET measurement of intramolecular distance. By allowing multiple reaction coordinates to be followed at once, such measurements can in principle allow the probing of multiple dimensions in the energy landscape, as well as providing more complete information about the conformation of the molecule.

### **3.3 The effect of force in single-molecule force spectroscopy measurements**

The focus of my work has been to use single-molecule force spectroscopy (SMFS) as a tool for studying protein folding. SMFS is by now a standard method for measuring structural transitions. A force probe (e.g. AFM, optical tweezers, or magnetic tweezers) is used to apply tension to a molecule. The force acts as a denaturant to induce structural changes, such as switching proteins between folded and unfolded states. In more traditional ensemble measurements, other variables such as chemicals, temperature changes, pressure changes, or pH changes are used as denaturants to perturb the protein structure. Although most denaturants are “artificial”, in the sense that the protein does not normally encounter the conditions being imposed, forces are in fact encountered naturally in the cell in many contexts, from gene regulation (Oluwole et al., 1997) to the translocation of proteins across membranes (Matouschek and Bustamante, 2003), making mechanical unfolding a biologically-relevant condition.

SMFS provides distinct advantages regarding misfolding and aggregation studies. Extremely stable structures such as amyloids can be disrupted with relative ease, and the use of force as a denaturant allows the solution conditions to be maintained constant while denaturing and renaturing the protein. This last point is especially important because misfolding and aggregation are often very sensitive to conditions such as pH, temperature, or chemical denaturants (Baskakov, 2002; Gerber et al., 2008; Hornemann and Glockshuber, 1998; O'Sullivan et al., 2007). Since these conditions are typically used to trigger unfolding/refolding, it can be challenging to isolate the effects that relate specifically to misfolding and aggregation. Misfolding can therefore be studied in the single-molecule regime under solution conditions which would not normally be

considered conducive to misfolding. Conversely, since individual molecules are measured in isolation, the folding dynamics of individual monomers may be studied under solution conditions in which the protein would normally aggregate rapidly.



**Figure 3.1 The effect of force on free energy landscapes.** The force tilts the energy landscape, changing the barrier height, the barrier position, the free energy difference as well as the occupancy of states (represented by the depth of the shading inside the wells).

SMFS also has advantages for applying energy landscape theory to study protein folding, because it allows free energy profiles to be measured directly, as will be discussed in greater detail in chapter 8. Here, we describe briefly the effect of force on the folding through changes to the landscape. The application of force introduces a natural, well-defined reaction coordinate, the end-to-end distance, onto which the energy landscape can be projected. Force tilts the energy profile along the reaction coordinate in a controlled and well-understood way (Fig. 3.1). From classical thermodynamics, at constant temperature and pressure, the free energy change is equal to the reversible work done to the system:

$$dG = F \cdot dx \quad (3.1)$$

Therefore, the force tilts the energy landscape by the amount depending linearly on the distance along the reaction coordinate  $\Delta x$ :

$$\Delta G(F) = \Delta G^0 - F \cdot \Delta x, \quad (3.2)$$

where  $\Delta G^0$  is the free energy difference at zero force. Meanwhile, the probability density of states along the reaction coordinate  $P(x)$  is related to the free energy  $\Delta G(x)$  via the Boltzmann transform:

$$P(x) = \exp(-\Delta G(x)/k_B T). \quad (3.3)$$

In a two-state system as illustrated in Fig. 3.1, an equilibrium force,  $F_{1/2}$ , exists at which the molecule spends equal time in the folded and unfolded states. According to Eq. 3.3, at  $F_{1/2}$  the free energy difference between the two adjacent states within the landscape is zero. Therefore,

$$\Delta G^0 = F_{1/2} \cdot \Delta x, \quad (3.4)$$

thus the free energy difference at an arbitrary force is given by Eq. 3.2 as:

$$\Delta G(F) = (F_{1/2} - F) \cdot \Delta x. \quad (3.5)$$

As a result, the occupancy of states can be shifted by an applied force. For example, the population of the high-energy unfolded state varies with force as:

$$P_u(F) = \{1 + \exp[(F_{1/2} - F) \Delta x / k_B T]\}^{-1}, \quad (3.6)$$

where  $\Delta x(F)$  is the force-dependent distance between adjacent states. Analogously, the population of high-energy rare or transient intermediate states can be tuned by force in a similar way, giving an excellent chance for observing them using SMFS.

The force-dependent rates can be obtained from the Kramers rate theory based on different kinds of assumptions. By assuming the distances to the transition states are not force-dependent, Eq. 2.1 becomes:

$$k(F) = k_0 \exp(-(\Delta G^{\ddagger,0} - F\Delta x^{\ddagger})/k_B T) = k_{\text{fold/unfold}} \exp(F\Delta x^{\ddagger}/k_B T) \quad (3.7)$$

which is the widely-used Bell's rate formula (Bell, 1978). In this formula,  $k_{\text{fold/unfold}}$  represents the folding/unfolding rate at zero force:  $k_{\text{fold/unfold}} = k_0 \exp(-\Delta G^{\ddagger,0}/k_B T)$ , where  $\Delta G^{\ddagger,0}$  is the height of the barrier at zero force and  $\Delta x^{\ddagger}$  is the distance to the transition state, which is considered to be positive in case of unfolding and negative in case of refolding.

However, force-independent transition state positions are usually not an accurate description. Alternatively, by specifying a certain shape of the energy landscape and using the Kramers theory at the high barrier limitation, Dudko *et al.* worked out an expression for force-dependent rate analytically (Dudko *et al.*, 2006):

$$k(F) = k_{\text{fold/unfold}} \left( 1 - \frac{\Delta x^{\ddagger,0} F}{\Delta G^{\ddagger,0}} \nu \right)^{1/\nu - 1} \exp \left\{ \frac{\Delta G^{\ddagger,0}}{k_B T} \left[ 1 - \left( 1 - \frac{\Delta x^{\ddagger,0} F}{\Delta G^{\ddagger,0}} \nu \right)^{1/\nu} \right] \right\} \quad (3.8)$$

in which  $\nu$  is a parameter characterising the shape of the energy barrier ( $\nu = 1/2$  for a cusp-like barrier,  $U_0(x) = \Delta G^{\ddagger} (x/x^{\ddagger})^2$  for  $(x < x^{\ddagger})$  and  $-\infty$  for  $(x \geq x^{\ddagger})$ ;  $\nu = 2/3$  for a softer cubic potential,  $U_0(x) = 3/2 \Delta G^{\ddagger} x/x^{\ddagger} - 2 \Delta G^{\ddagger} (x/x^{\ddagger})^3$ ). For  $\nu = 1$ , the expression reduces to the simple Bell model. The Dudko model tracks the dependence of the position of the transition state with force, therefore is expected to provide a more accurate way for

characterizing rates. In cases where the reconstruction of the full energy profile, which is based on high-resolution single-molecule measurements, is not essential, these theories (the Dudko theory in combination of the Kramers theory), provide an accurate prediction of the most important parameters of the energy landscape ( $\Delta G^\ddagger$ ,  $\Delta x^\ddagger$ ) and of protein folding itself ( $D$ ,  $\tau_{fp}$ ). (The comparison between the experimental and predicted rates will be discussed in chapter 8.)

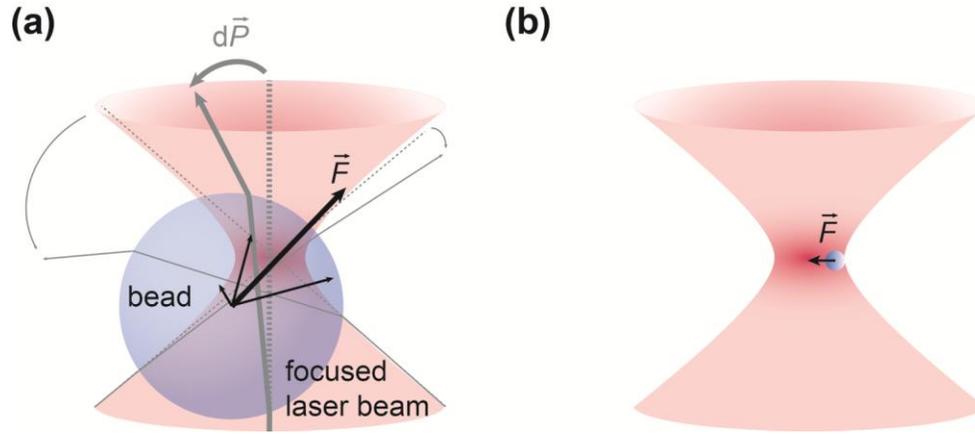
Because of the effect of force applied over the entire energy landscape, the free energy of states and the kinetics of the transitions between different states can be measured directly from SMFS. In principle, all accessible states along the folding pathway should be sampled in equilibrium force measurements. Rare or transient states may be more populated at different forces, which provides a more complete picture of the protein folding and aggregation network.

## 4. Optical tweezers

Since their invention by Arthur Ashkin in 1986 (Ashkin et al., 1986), optical tweezers have become a powerful tool for studying biological systems at the single-molecule level. In this thesis, we use optical tweezers as the major technique to investigate protein folding. The basic principles of optical tweezers are introduced in this chapter, followed by a brief description of the construction of our high-resolution optical tweezer system. After that I describe two basic types of SMFS measurements, which will be used extensively in this thesis, and what information is expected out of the measurements.

### 4.1 Principles of optical trapping

Optical tweezers use high optical intensity gradients to manipulate microscopic dielectric objects with electromagnetic forces (Neuman and Block, 2004). The mechanism of optical trapping can be understood simply based on the momentum conservation in the Mie regime (Ashkin, 1992), by assuming the dimension of the object is much larger than the wavelength of the laser (Fig. 4.1a). As a tightly focused laser beam passes through a dielectric object, such as a plastic bead, light is deflected with changes in momentum. From the momentum conservation law, the momentum change of the object must be equal but opposite to the momentum change of the light, pushing the object back to the centre of the laser trap. The situation is complicated when considering the scattering force coming from light reflected from the object, as a result of which the object is pushed axially away from the focal point. For micron-sized objects, the axial gradient force typically balances the scattering force, allowing stable trapping.



**Figure 4.1 Principles of optical trapping.** (a) The principle of optical trapping in Mie regime. The momentum of the laser changes as the laser beams (grey) are deflected by the object, which results in the restoring force (black) of the object. (b) The principle of optical trapping in Rayleigh regime. A Lorentz force is generated as the dipole of the object interacts with light.

An alternative view of optical trapping considers the opposite limit, when the trapped object is much smaller than the wavelength of the laser (known as the Rayleigh scattering regime). In this case, the object can be considered as a dipole which interacts with the electric field via a Lorentz force (Fig. 4.1b). The absorption and re-radiation of the light by the object gives the scattering force, which is proportional to the intensity of the laser:

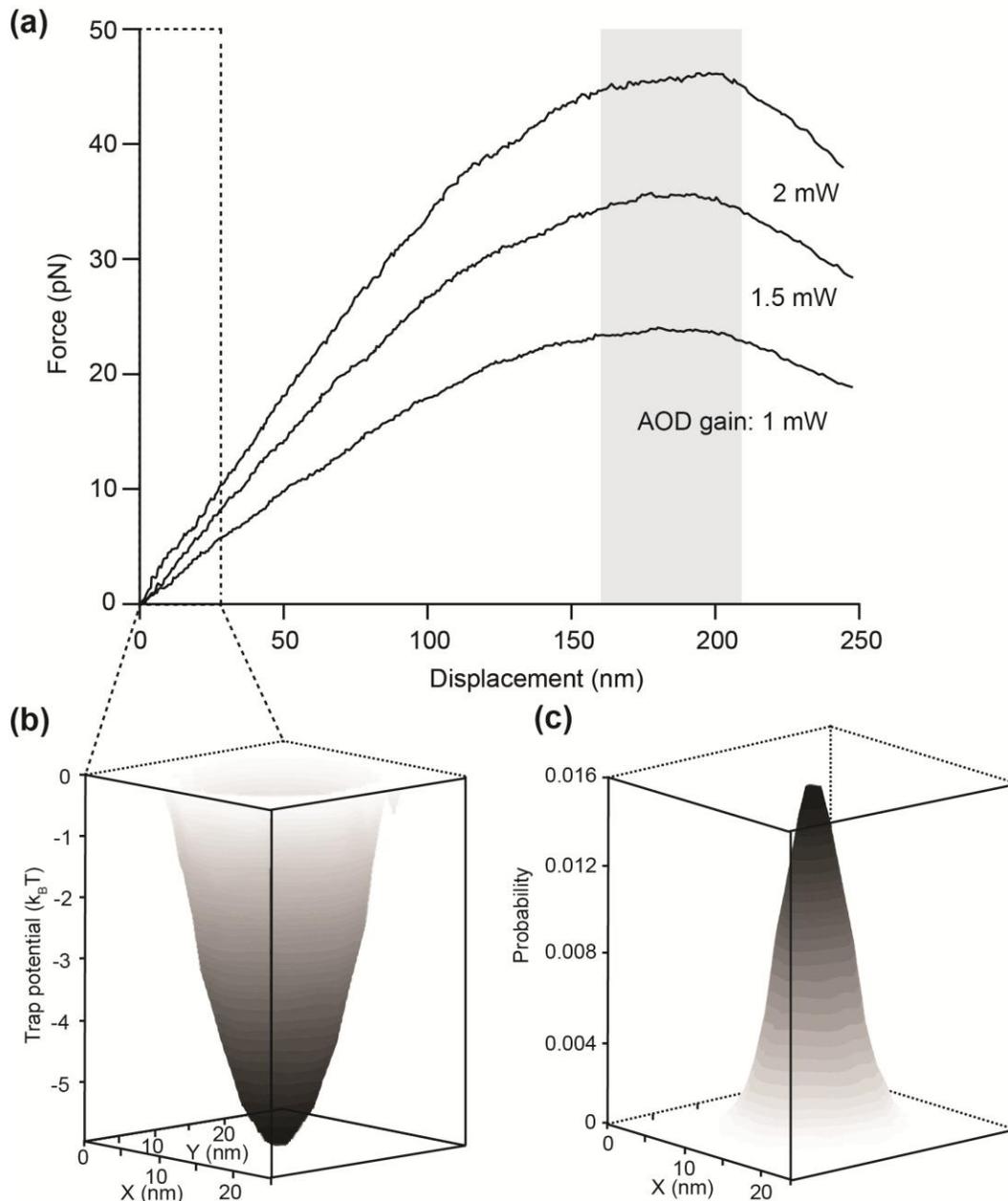
$$F_{scat} = \frac{n_m C}{c} I \quad (4.1)$$

where  $I$  is the laser beam intensity,  $c$  is the speed of light in vacuum,  $n_m$  is the refraction index of the medium, and  $C = (8/3)\pi k^4 r^6 [(m^2-1)/(m^2+2)]$  is the scattering cross-section of the sphere. The time-averaged Lorentz force gives the gradient restoring force exerted on the object (Harada and Asakura, 1996; Woodside and Valentine, 2009):

$$\langle \bar{F}_{grad} \rangle = \langle (\bar{p} \cdot \bar{\nabla}) \bar{E} \rangle = \frac{2\pi m_m}{c} \left( \frac{m^2 - 1}{m^2 + 2} \right) r^3 \bar{\nabla} I \quad (4.2)$$

where  $r$  is the radius of a sphere,  $m = n_p/n_m$  is the ratio of the indices of refraction of the sphere ( $n_p$ ) and medium ( $n_m$ ). The restoring force is thus proportional to the gradient of the light intensity.

For a laser beam with a Gaussian-shape intensity profile, the force rises linearly with displacement near the centre of the trap. The restoring force can be then considered as  $F = -k_T x$ , where  $k_T$  is the stiffness of the trap. The optical trap acts like a three-dimensional spring made out of light with a constant stiffness in the centre of the trap, giving an harmonic potential  $V = (1/2)k_T x^2$ , as shown by the trap potential (Fig. 4.2b) measured from the Brownian motion of a trapped bead (Fig. 4.2c). However, as the object is displaced further from the trapping centre, the gradient of the laser intensity becomes non-linear with the displacement. The force, proportional to the gradient, then rolls over and generates a non-harmonic trapping potential outside the centre region (Fig. 4.2a). This non-harmonic window of  $\sim 30$ - $50$  nm width where the force is constant to within a few percent effectively has zero stiffness, and hence can be used to clamp the force passively, without the need for feedback loops which can introduce experimental artifacts (Greenleaf et al., 2005). This is the principle of the passive constant force measurement used in our studies.



**Figure 4.2 The trapping profile of optical tweezers.** The Force-displacement curves (a) for beads held in an optical trap at different light intensity (AOD gain) showed a linear harmonic potential (b) near the trap centre and a non-harmonic potential as the bead moved farther out. As a result the force rolled over to generate a zero stiffness region of  $\sim 50$  nm wide (grey). The trap potential near the centre of the trap is measured from the Gaussian position histogram (c) of a trapped bead based on the Boltzmann relation between the probability for the displacement of a trapped bead with the potential. These data were taken from our instrument using 600-nm-diameter polystyrene beads.

In a typical optical trapping experiment, beads with diameters  $\sim 1 \mu\text{m}$  are used, and the wavelength of the trapping laser is often in the near infrared region ( $\sim 1 \mu\text{m}$ ) for the least damage to biological samples. Therefore, the above two limiting cases break down. More complete electromagnetic theories have been developed to calculate the force in the intermediate regime that characterizes most measurements (Rohrbach and Stelzer, 2001, 2002). However, these theories are not considered here as the physical principles of trapping remain the same, and the basic dependence of the force on the gradient of the light intensity is unchanged.

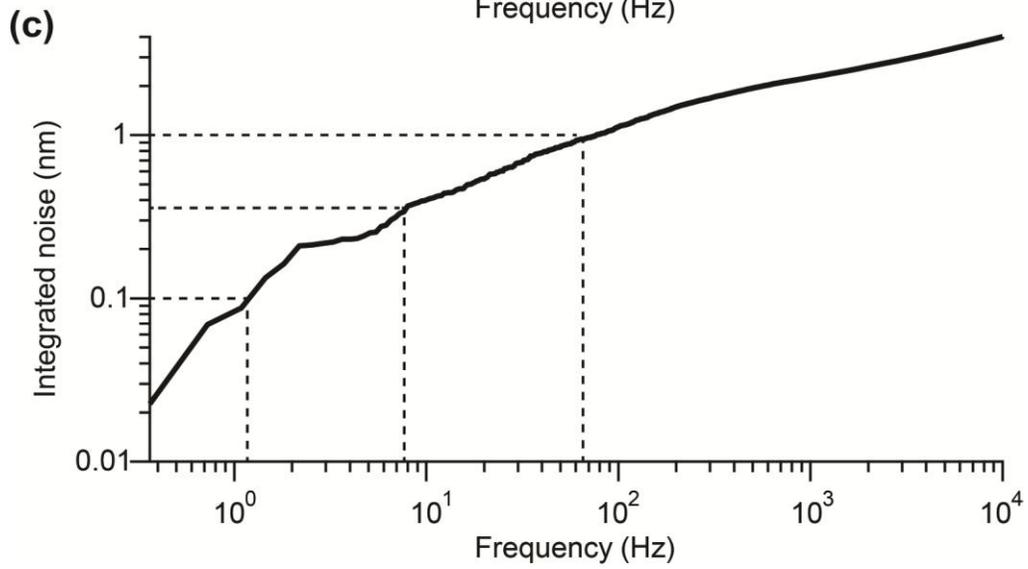
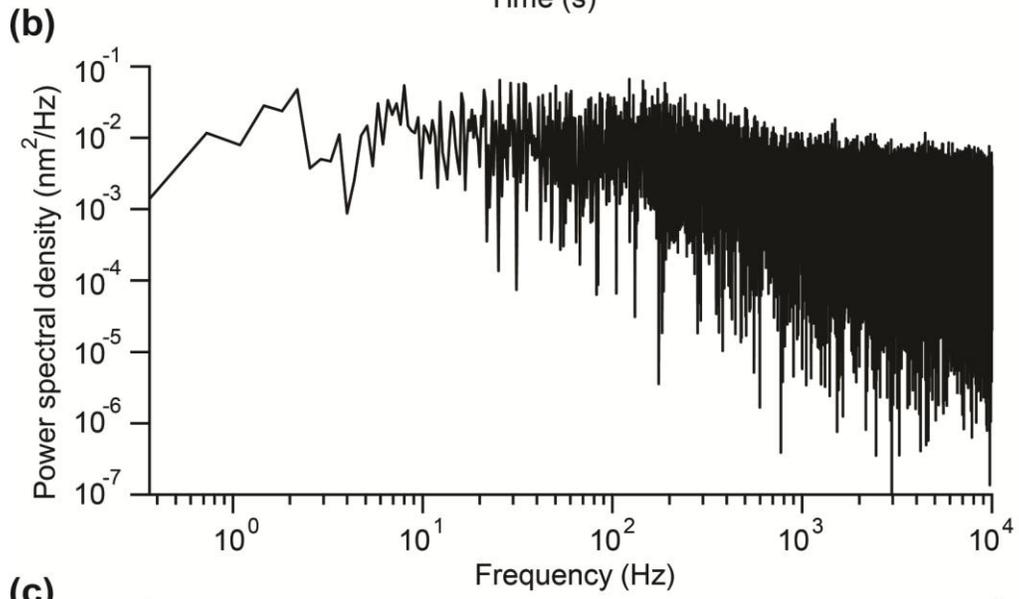
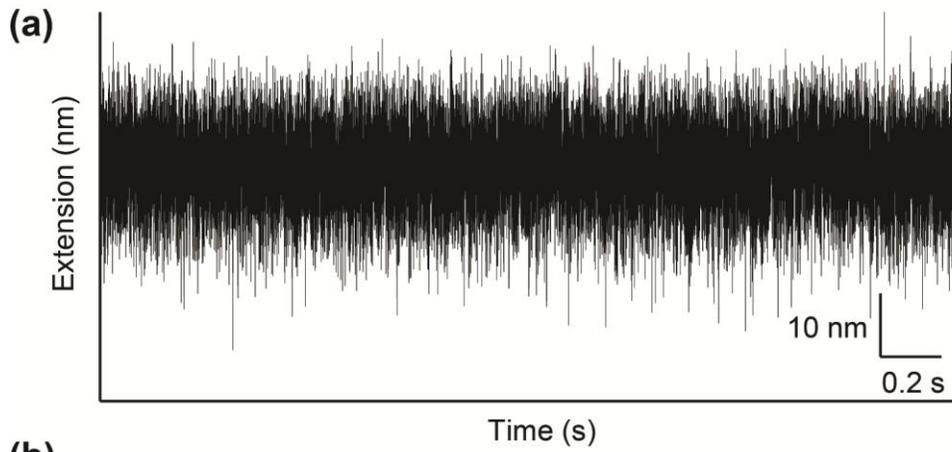
## **4.2 Construction of ultra-stable high-resolution optical tweezers system**

A dual-trap optical tweezer system with a separate detection laser (Fig. 4.3) was used for all our measurements. A 5W, 1064-nm Nd:YVO<sub>4</sub> solid state laser from Spectra-Physics (BL106C) was used as the trapping laser. The laser beam was separated by polarization to generate two traps. The position of each trap was controlled independently by one acousto-optic deflector (AOD, AA Optoelectronique) and one electro-optic deflector (EOD, Conoptics) in each axis. The traps were moved apart during all measurements using the EODs only, because of their highly linear response (Valentine et al., 2008). The AODs were used to steer the beam in the perpendicular axis during calibration processes and to control the stiffness of each trap by modulating the laser beam intensity. Several Keplerian telescopies were used to expand the beam to fill the objective at the back focal plane, to maximize trap stiffness (Neuman and Block, 2004).



**Figure 4.3 Design of a dual-trap optical tweezers system.** The system consists of three major parts: a trapping laser with beam steering (orange path), a detection laser with beam steering (red path), and a microscope with detection. Two orthogonally-polarized laser beams from the same 1064-nm laser were used to generate two traps. The position of each trap was controlled independently by one AOD and one EOD in each axis. The traps were moved apart during all measurements using the EODs only. The stiffness of each trap was controlled by the AODs. Three Keplerian telescopes were used in each laser path to expand the beam to overfill the objective lens (full width at  $1/e \sim 5$  mm) for maximum stiffness. Bead positions within the traps were measured by collecting the light from two orthogonally-polarised beams at 633 nm scattered by the beads onto independent PSDs. The flow chamber is made by plasma-cleaned microscope slides and cover slips with double sided tapes (yellow), with the sample channel in between. The channel is sealed with glue (cyan) to prevent drying during the measurement. Labels in the figure:  $\lambda/2$ , half waveplate; PBS, polarizing beam splitter; DM, dichroic mirror; CCD, charge-coupled device; P, polarizing film; IF, interference filter. The dashed arrows label conjugate planes.

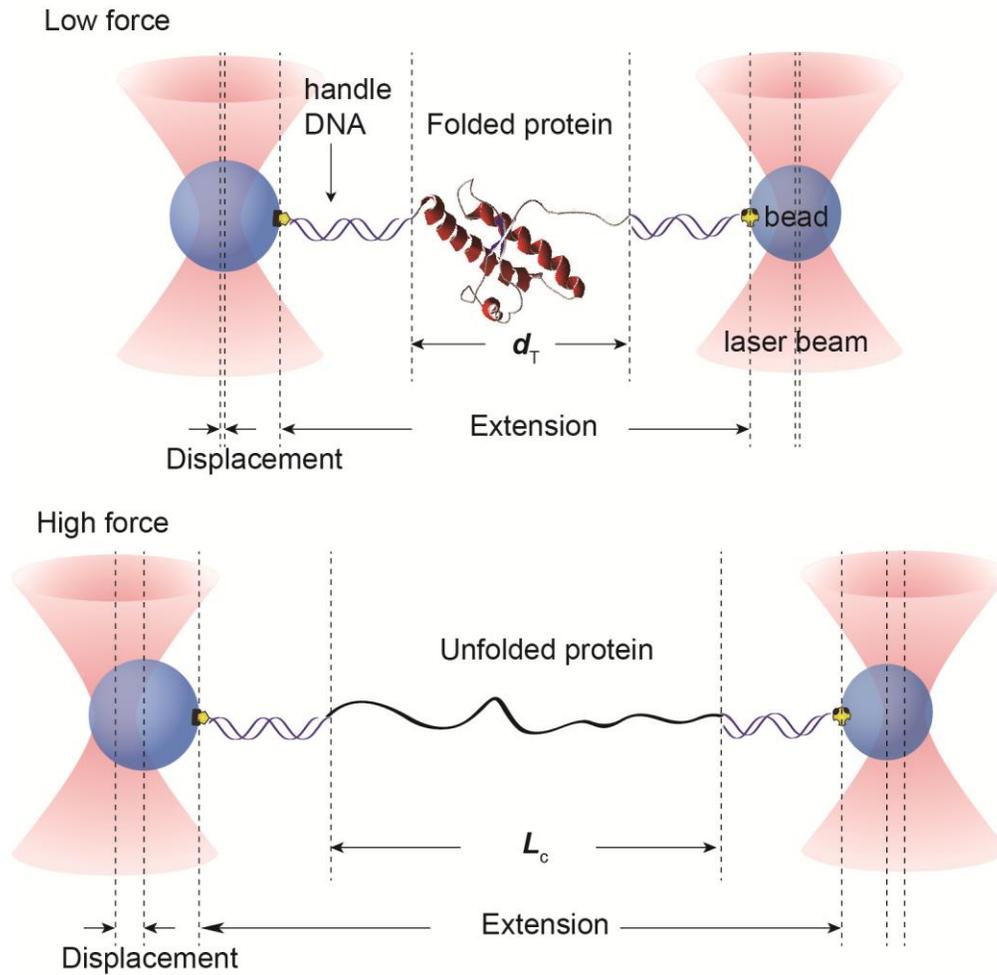
The position of the trapped bead is monitored by the scattered light from a separate detection laser. In our system, a 633-nm HeNe gas laser is used as the detection laser. This laser was coupled into a single mode optical fiber for enhanced pointing stability, and then divided into two orthogonally-polarized beams, one for each trap. The beams were then passed through steering lenses to control their position in the specimen plane, collimated, expanded, and combined with the trapping beams before sending into the microscope. Detection light scattered by the beads was separated by polarization and collected by independent position-sensitive diodes (PSD, Pacific Silicon Sensors) for position calibration. Further details of the construction and calibration of the optical tweezers have been reviewed in the master's thesis of Daniel A. Foster (Foster, 2010).



**Figure 4.4 Spatial resolution.** (a) A short segment of constant force data can be used to estimate the spatial resolution of the optical tweezers. Here the data was taken from the extension trajectory of PrP when the protein was fully folded at 9.2 pN. (b) The power spectral density from the Fourier transform of the extension trajectory. (c) Integrated noise from the power spectral density. Dashed lines represent the bandwidth needed to achieve a resolution of 1 Å, 3.6 Å (the length of one amino acid), and 1 nm.

A system with ultra high spatial and temporal resolution is required to calibrate the displacements and forces accurately and observe fast dynamics effectively. The spatial resolution is limited by instabilities from two main sources: (1) environmental factors, such as temperature stability, acoustic noise, mechanical stability and air currents; and (2) system instabilities, such as fluctuations of the laser beams. Our system has been designed to minimize these instabilities in several ways. First of all, the instrument is housed in a quiet room with the temperature controlled to within 0.1 °C. Second, measurements were made with the molecule uncoupled mechanically from the microscope slide: the molecule was held between two beads levitated about 600 nm above the slide surface by the traps. The sample was therefore isolated from mechanical motions of the microscope. Furthermore, this dual-trap assay also reduced noise from sources such as laser pointing fluctuations, by allowing differential measurement between the two traps. Third, the effects of air currents were reduced by covering the optical paths with plastic tubes, especially at focal points. All the arrangements implemented to our system have been contributed to bring the spatial resolution down to sub-nanometer range (Fig. 4.4). On the other hand, the temporal resolution of the system is limited to ~ 0.1 ms by the bead size and trap/construct compliance. Further improvements could be made using smaller beads and stiffer constructs.

### 4.3 Dumbbell assay



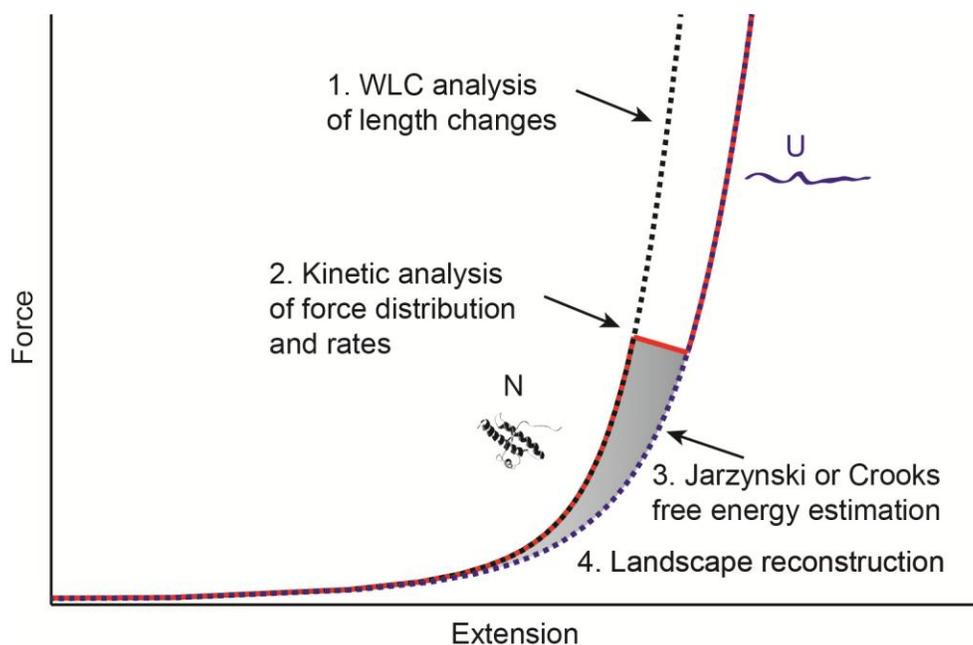
**Figure 4.5 Experimental assay for the dual trap apparatus.** A protein labeled at both termini with Cys residues is attached to sulfhydryl-labeled DNA strands bound to beads held in optical traps. The extension of the molecule held under tension by the traps is measured as the protein unfolded or refolded. The force is proportional to the displacement of the bead from the centre of the trap. Note: figures are not drawn to scale.

A macromolecule itself is generally too small to be trapped, since the trapping force depends on the size of the molecule (Eq. 4.2). Although techniques have been developed to trap single protein molecules (Pang and Gordon, 2012), the direct manipulation of the molecule by force is still not available. In a standard optical tweezers SMFS assay, force is instead applied through micrometer-sized polystyrene beads, which

are linked to the molecule through kilobase-long double-stranded DNA handles (~ 300 nm each) (Fig. 4.5). The handles also serve as spacers between the two traps, preventing interactions of the molecule with the surface of the beads. A number of attachment schemes have been developed to provide strong, specific binding. Antibody-antigen interactions are widely used to attach the handles to the beads and a common approach for attaching DNA handles to proteins is through a disulfide bridge (Cecconi et al., 2008). Beads of different sizes are typically used for easy identification under the microscope.

#### **4.4 SMFS measurements and information to be learned**

In SMFS measurements, the dynamics of a molecule is monitored from the motions of the beads in the traps. The extension of the construct as well as the force exerted on the molecule can be derived from the relative positions of the beads with respect to their equilibrium positions. SMFS is often done in two different regimes: non-equilibrium force-extension curve measurements, and equilibrium constant-force measurements. A wealth of information, such as the detailed folding pathways, the presence of intermediate or misfolded states, the kinetics of the folding reactions, the transition path time, the diffusion constant as well as the whole profile of the energy landscapes can be determined from SMFS.



**Figure 4.6 Summary of FEC analysis.** Analyses of FEC measurements involve WLC fittings of different states along the folding pathway, kinetic analysis of force distributions and rates, free energy estimation from the Jarzynski and/or Crooks theorem, as well as full energy landscape profile reconstruction using the Hummer-Szabo formalism.

#### 4.4.1 Non-equilibrium force-extension curve measurements

The molecule can be unfolded/refolded while moving the traps apart/together at a constant velocity to ramp the force up/down. The extension of the constructs (Fig. 4.5) can be measured as we monitor the force at the same time (see appendix for a detailed protocol of the distance and force calculation), creating force-extension curves (FECs). In our experiments on PrP, ramping velocities between 10–300 nm/s were typically used to generate loading rates (rate of change of the applied force) ranging from ~1–50 pN/s. The analysis of FECs yields: (1) folding pathways and structures of states involved in terms of contour length change; (2) kinetic rates and energy barriers; (3) estimation of free energy changes; as well as (4) energy landscape profiles, as summarized in Fig. 4.6. Note that

rapid changes to the force mean that these measurements are typically under non-equilibrium conditions.

Typical FECs of a simple DNA tether, made by connecting two DNA handles together without any other molecules in between, show a monotonic but non-linear increase in extension with force as the handles are stretched (Fig. 4.7a). For mechanical loads below  $\sim 35$  pN (Gross et al., 2011), DNA elasticity can be well described by extensible worm-like chain (WLC) model (Marko and Siggia, 1995; Wang et al., 1997), which treats the double-stranded DNA as an entropic chain with flexible bonds. The extensible WLC model yields the relation between the extension of the molecule ( $x$ ) and the force exerted on the molecule ( $F$ ) as follows (Wang et al., 1997):

$$F(x) = \frac{k_B T}{L_p} \left[ \frac{1}{4} \left( 1 - \frac{x}{L_c} + \frac{F}{K} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} - \frac{F}{K} \right]. \quad (4.3)$$

Here  $L_p$  is the persistence length of the polymer,  $L_c$  is the contour length,  $K$  is the stretch modulus which accounts for the enthalpic compliance of polymers due to the stretching of covalent bonds at higher forces. The handle FEC fits well to the extensible WLC model at force up to  $\sim 40$  pN, but deviates at higher forces (Fig. 4.7a) due to the twisting effect of the double stranded DNA handles (Gross et al., 2011). At forces  $\sim 60$  pN, the handles become “overstretched,” generating a plateau (Smith et al., 1996) (Fig. 4.7a). At this stage, the handles are being denatured by force and the sawtooth pattern is generated by unpeeling of segments of the DNA handles, until the handles are fully melted and begin to act as single-stranded DNA (Gross et al., 2011; Paik and Perkins, 2011). As a result, the DNA-handle-based assay is only valid for studying folding events lower than 60 pN.

When a molecule is attached, the FEC measurements provide information regarding the folding pathways, kinetics and energy landscapes for the molecule of interest.

### 1. Mapping out folding pathways and determination of contour length changes

In the presence of a molecule between two handles, the molecular structure may change with force, for example unfolding at a certain force, causing a deviation of the actual FECs from the WLC behavior of the handle alone. Typically, unfolding of a structure causes an abrupt increase in extension and concomitant drop in force, which indicates an apparently two-state cooperative unfolding process (Fig. 4.7b). The contour length change  $\Delta L_c$  of the unfolded molecule can be determined by fitting the second part of FECs with two WLC in series, one for the DNA handles and one for the unfolded molecule.  $L_c$ ,  $L_p$ , and  $K$  of the handle were treated as free parameters for fitting the folded branch of the FECs, but thereafter as fixed parameters for fitting the unfolded branch.  $L_p$  and  $K$  for the unfolded biopolymers were also treated as fixed parameters when fitting the unfolded branch of the FECs. Hence the latter fit involved only a single free parameter, the contour length change,  $\Delta L_c$ , which can be then related to the structures involved during the unfolding event. Any intermediate state (I) should have been seen as an extra step in between the folded native state (N) and the unfolded state (U) (Fig. 4.7b). However, non-equilibrium FECs are usually less sensitive for mapping out the folding pathways compared with the equilibrium constant force method, for example, as we can see in case of RNA aptamer in the next chapter.

## 2. Folding kinetics

The unfolding forces ( $F_{unf}$ ) encode a wealth of information regarding the kinetics of the folding reaction. Because of the stochastic nature of unfolding, a given molecule will unfold at a slightly different force each time it is pulled apart, producing a distribution of unfolding forces. This distribution can be related to the force-dependent rate for unfolding as well as the shape of the energy landscape (the folded well and the barrier from the folded state). The distribution,  $p(F)$ , can be described very well by a quantitative model based on Kramers theory developed by Dudko et al. (Dudko et al., 2006):

$$p(F) \propto \frac{k(F)}{r} \exp \left\{ \frac{k_{\text{unfold}} k_B T}{\Delta x_N^{\ddagger,0} r} - \frac{k(F) k_B T}{\Delta x_N^{\ddagger,0} r} \left( 1 - \frac{\Delta x_N^{\ddagger,0} F}{\Delta G_N^{\ddagger,0}} \nu \right)^{1-1/\nu} \right\}, \quad (4.4)$$

$$\text{where } k(F) = k_{\text{unfold}} \left( 1 - \frac{\Delta x_N^{\ddagger,0} F}{\Delta G_N^{\ddagger,0}} \nu \right)^{1/\nu-1} \exp \left\{ \frac{\Delta G_N^{\ddagger,0}}{k_B T} \left[ 1 - \left( 1 - \frac{\Delta x_N^{\ddagger,0} F}{\Delta G_N^{\ddagger,0}} \nu \right)^{1/\nu} \right] \right\}$$

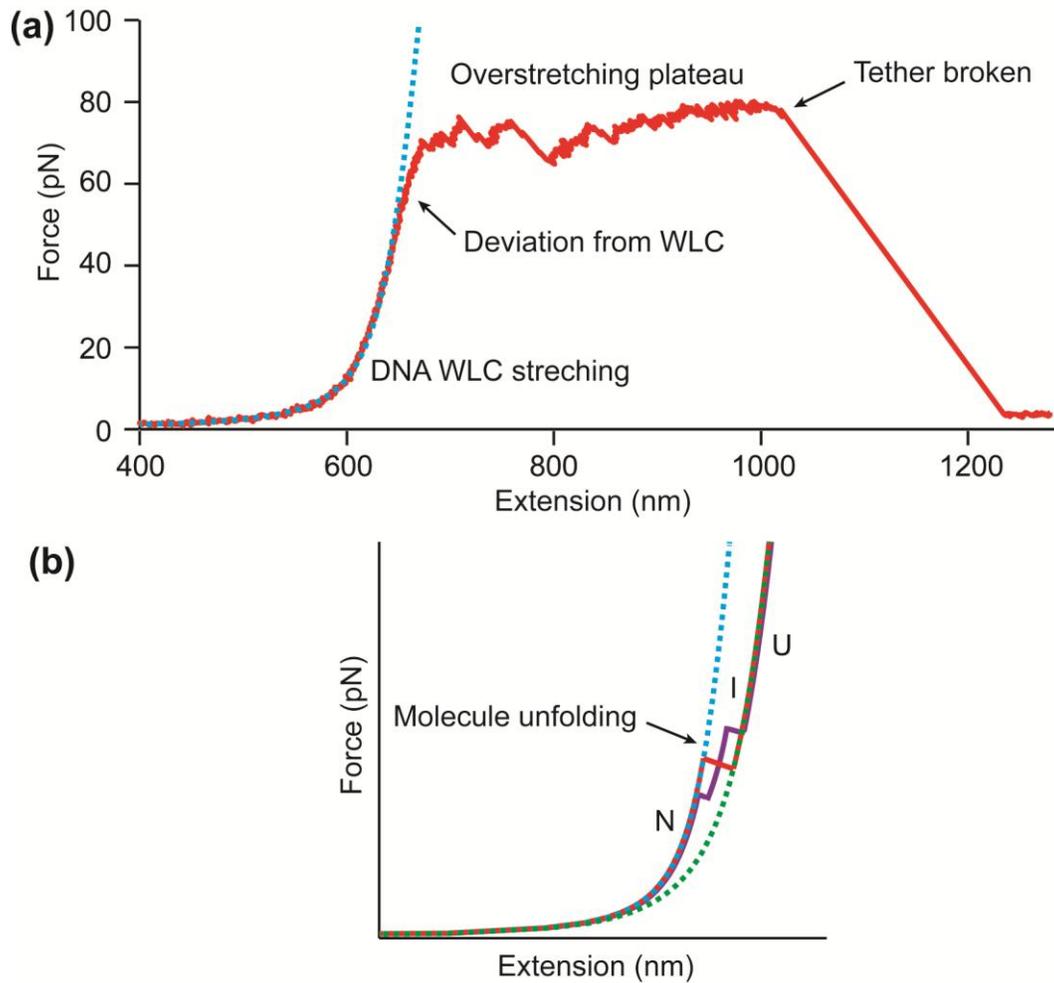
is the force dependent rate introduced in the last chapter (Eq. 3.8). In these expressions,  $k_{\text{unfold}}$  is the unfolding rate at zero force,  $\Delta x_N^{\ddagger,0}$  is the distance to the transition state from the native (folded) state at zero force,  $\Delta G_N^{\ddagger,0}$  is the height of the energy barrier from the native state at zero force,  $r$  is the loading rate,  $k_B$  is Boltzmann's constant, and  $\nu$  is a parameter characterising the shape of the energy barrier ( $\nu = 1/2$  for a cusp-like barrier,  $\nu = 2/3$  for a softer cubic potential). Based on the same theory, the refolding force distribution depends on the folding rates as well as the energy barrier from the unfolded state in a similar way:

$$p(F) \propto -\frac{k(F)}{r} \exp \left\{ -\frac{k(F)k_B T}{\Delta x_U^{\ddagger,0} r} \left( 1 - \frac{\Delta x_U^{\ddagger,0} F}{\Delta G_U^{\ddagger,0}} \nu \right)^{1-1/\nu} \right\} \quad (4.5)$$

where  $\Delta x_U^{\ddagger,0}$  and  $\Delta G_U^{\ddagger,0}$  is the position and height of the energy barrier from the unfolded state at zero force. Here,  $\Delta x^{\ddagger}$  and  $r$  are taken as negative numbers, since the extension is decreasing and the force is ramping down during the refolding experiments. A complementary analysis of the kinetics based on the cumulative probability of unfolding yields the lifetime of the folded state as a function of force and hence the unfolding rate  $k(F)$  (Dudko et al., 2008). Values of  $k(F)$  from datasets at different loading rates for the same type of molecules collapse to a single curve (Dudko et al., 2008), that is well fit by the same type of landscape model using Eq. 3.8.

### 3. Equilibrium free energy estimation

Equilibrium thermodynamic properties can be extracted from non-equilibrium measurements. Although work is irreversibly dissipated in such measurements, the equilibrium free energy can be reconstructed from the distribution of the non-equilibrium work using fluctuation theorems such as the Crooks theorem (Crooks, 1999) or the Jarzynski equality (Jarzynski, 1997). These methods have been tested through nanomechanical measurements of unfolding transitions in single RNA hairpins (Collin et al., 2005; Liphardt et al., 2002) and applied widely (Ritort, 2008).



**Figure 4.7 Force-extension curve measurements.** (a) FECs of DNA handles alone. The FEC deviates from the standard extensive WLC starting from  $\sim 40$  pN. At  $\sim 70$  pN there is a plateau from overstretching. Although  $L_p$  for long double-stranded DNA is measured to be 50 nm,  $L_p$  for short DNA may significantly differ from this number (20 - 30 nm in our measurements). The slope after the breakage of the tether is related to the stiffness of the trap as the bead moving back to the centre of the trap. (b) Schematic FEC (red) of cooperative two-state unfolding between the native state (cyan) and the unfolded state (green). In presence of intermediate, steps appear in between (purple).

The Crooks fluctuation theorem predicts a time reversed symmetry in the non-equilibrium work done on a system. It states that the distribution of non-equilibrium work  $P_U(W)$  associated with the forward change of a system (in our case the unfolding of a

molecule) relates to the reverse work distribution  $P_R(W)$ , corresponding to the refolding of a molecule, via following equation:

$$\frac{P_U(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G}{k_B T}\right), \quad (4.6)$$

assuming the forward and reverse changes of the system follow the same (but time-reversed) protocol. In FEC measurements, this protocol requires the loading rate to be the same for the unfolding and the refolding process. Here  $\Delta G$  is the equilibrium free energy difference between the initial and the final state of the system. As shown in Eq. 4.6, under equilibrium conditions,  $\Delta G = W$  since no work is dissipated. Therefore, free energy can be estimated by  $\Delta G = \langle W \rangle$ , given the measurement is done near the equilibrium regime. However, since a typical FEC measurement is done in the non-equilibrium regime, from Eq. 4.6,  $\Delta G$  can be determined as the work value at which the unfolding and refolding work distributions cross, i.e. where  $P_U(W) = P_R(-W)$ .

The Jarzynski equality, on the other hand, provides a simple identity between the equilibrium free energy and non-equilibrium work for a one-sided reaction (e.g. folding or unfolding, separately, without the need to measure both). It states that the free energy change of a system,  $\Delta G$ , relates to the exponential average of the non-equilibrium work done on the system,  $W$ , as follows:

$$\exp\left(-\frac{\Delta G}{k_B T}\right) = \left\langle \exp\left(-\frac{W}{k_B T}\right) \right\rangle. \quad (4.7)$$

In FEC measurements,  $W$  can be estimated from the area under the curve (integrating force over distance). The value estimated for  $\Delta G$  converges to the correct value as the

number of FECs goes to infinity. However practically, the exponentially-weighted average heavily emphasizes the low-energy tail of the work distribution, which has contributions from only a limited number of rare events. Therefore, the Jarzynski estimator has a systematic bias towards overestimating the value of  $\Delta G$ . This bias becomes larger as the measurement goes further out of equilibrium and as the number of samples becomes smaller. Different methods have been developed to estimate the Jarzynski bias (Gore et al., 2003; Palassini and Ritort, 2011). In the near equilibrium regime, the bias  $B_N$  depends on the averaged work and the number of samples by the following relation (Gore et al., 2003):

$$B_N = \frac{W_{dis}}{N^\alpha}, \quad (4.8)$$

in which  $W_{dis}$  is the dissipated work defined as the difference between the non-equilibrium work and equilibrium free energy difference:

$$W_{dis} = \langle W \rangle - \Delta G, \quad (4.9)$$

and  $\alpha$  is a decreasing function of  $W_{dis}$ :

$$\alpha = \frac{\ln[2CW_{dis}/k_B T]}{\ln[C(e^{2W_{dis}/k_B T} - 1)]}. \quad (4.10)$$

Here  $C$  is a constant induced by Gore *et al.* that defines how small the bias must be before reaching the large  $N$  limit and was assumed to be 15 for bias calculation (Gore et al., 2003).

#### 4. Energy landscape reconstruction from non-equilibrium measurements

Going beyond the determination of key parameters describing the energy landscape, Hummer and Szabo proposed an approach to reconstructing the whole free energy profile from non-equilibrium force-ramp measurements (Hummer and Szabo, 2001). Briefly, assume a time-dependent Hamiltonian of the single molecule pulling system of the form  $H(x,t) = H_0(x) + V(x,t)$ , where  $H_0(x)$  is the Hamiltonian of the unperturbed system and  $V(x,t)$  is the time-dependent perturbation from the trap. The perturbation is described in terms of the molecular extension  $q(x)$ , effective trap stiffness  $k_s$ , and center position  $z(t) = z_0 + vt$  of the trap moving at velocity  $v$ , as:

$V(x,t) = V[q(x)-z(t)] = 1/2 k_s[q(x)-z(t)]^2$ . The external work along each non-equilibrium trajectory at time  $t$  is then found from:

$$W_t = \int_0^t \frac{\partial H(x(t'),t')}{\partial t} dt' = k_s v \left[ vt^2 / 2 + z_0 t - \int_0^t z_{t'} dt' \right] \quad (4.11)$$

For each FEC, indexed by  $k$ , we determine the trap position  $z_{ik}$  at discrete time  $t_i$  and then calculate the external work  $W_{ik}$  by numerical integration of the data using:

$$W_{ik} = k_s v \left[ 0.5 vt_i^2 + z_0 t_i - \sum_{j=1}^i (t_j - t_{j-1}) (q_{jk} + q_{j-1,k}) / 2 \right] \quad (4.12)$$

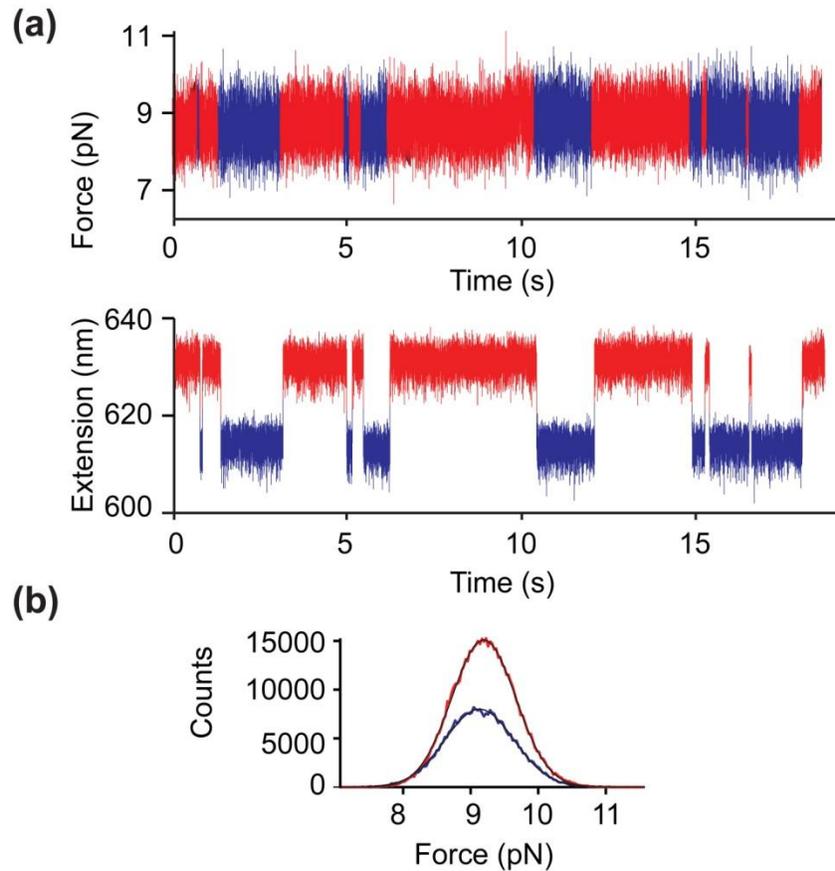
To obtain an optimal estimate of the free energy surface, averages are made over many time slices and repeated trajectories. For each time slice  $t$ , ensembles of positions  $z_t$  and corresponding  $W_t$  values are obtained. The extensions are binned with respect to the time slices and the corresponding histogram values are weighted by  $\exp(-W_t/k_B T)$ , leading to the “weighted-histogram” calculation of the unperturbed molecular free energy:

$$G_0(q) = -\beta^{-1} \ln \frac{\sum_t \frac{\langle \delta(q - q_t) \exp(-\beta W_t) \rangle}{\langle \exp(-\beta W_t) \rangle}}{\sum_t \frac{\exp[-\beta V(q, t)]}{\langle \exp(-\beta W_t) \rangle}} \quad (4.13)$$

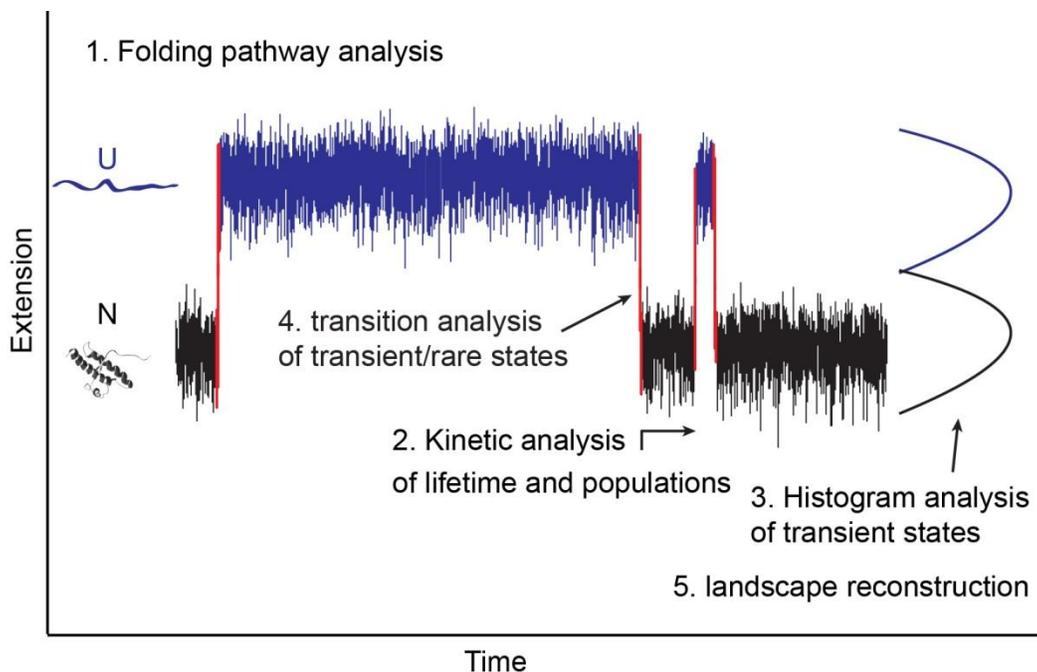
#### 4.4.2 Equilibrium constant force measurements

FECs probe the folding dynamics out of equilibrium, due to the changing force. To investigate the folding under equilibrium conditions, the extension of the molecule can be measured as a function of time while the force is held constant using a force clamp (see appendix for a detailed protocol of the distance and force calculation). Constant force can be achieved by an active force clamp or a passive force clamp. The active force clamp can cause instability of the force from the resonant oscillations of the system, which is not suitable for studying fast dynamics (a good example of this in real data can be found in a recent paper using AFM to study the dynamics of an unfolded protein (Berkovich et al., 2012)). The passive force clamp makes use of the non-harmonic part of the trapping potential in optical tweezers, as demonstrated previously (Greenleaf et al., 2005). A passive force clamp maintains a constant force during the transitions themselves, avoiding artifactual transients from feedback loop closure. As Fig. 4.8 shows, during the transition from a typical two state system, the force changes very little (here, less than 0.07 pN; the apparent force change is probably due mostly to drift in the force signal, which is single-ended in contrast to the differential extension measurement and hence more susceptible to drift). The same kind of information that can be obtained from FECs can also be recovered from the analysis of constant force measurements (Fig. 4.9), including the folding pathways, energetic stability, kinetic rates and energy barriers, and

energy landscape profiles, allowing the results to be checked for consistency. Meanwhile, under the equilibrium condition provided by the passive force clamp, the folding pathways of a molecule can be characterized in a much more detailed way, because it is easier to detect rare or transient states from histogram and transition analysis of the extension trajectories.



**Figure 4.8 Force remains constant during constant-force measurements.** (a) Force and extension records measured simultaneously using the passive force clamp. The force remains unchanged whether the protein is folded (blue) or unfolded (red); no transients are observed as the folding/unfolding takes place. (b) Histograms of the force in the folded (blue) and unfolded (red) states are well fit by Gaussians whose peaks differ by less than 0.07 pN.

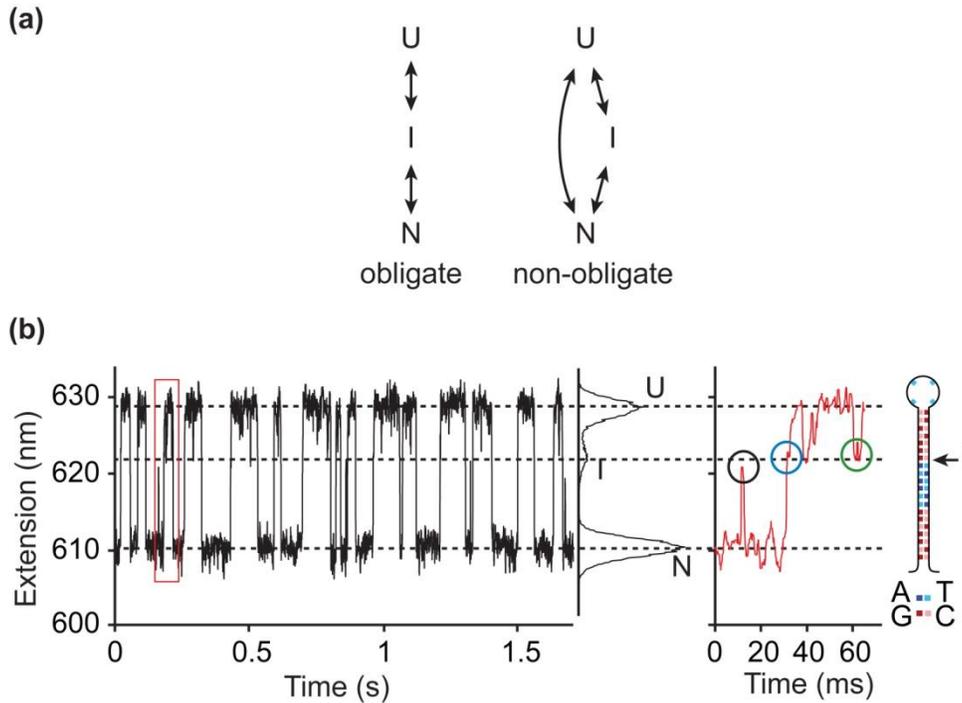


**Figure 4.9 Summary of constant force analysis.** The folding pathway analysis, kinetics analysis of lifetime and populations, histogram analysis and transition analysis for searching for transient states and energy landscape reconstructions can be applied to constant force measurements. Most of the methods for researching rare or transient states were first developed based on the work done in this thesis.

#### 1. Map out folding pathway and detection of intermediate/misfolded states

In constant force measurements, conformational changes of a molecule are usually accompanied by a change in extension, as shown in Fig. 4.8 for a two state protein and in Fig. 4.10 for a three-state DNA hairpin. Since constant force measurements are made under equilibrium conditions, the molecule will generally sample all possible transitions between different conformations. For example, if there is an intermediate state, I, on the folding pathway between the unfolded state U and the native state N, then it should be seen not only as a step between U and N when the molecule unfolds or refolds, but also as a transient excursion from U and, separately, from N (Fig. 4.10). If an intermediate is only seen as a transient excursion from U or from N (but not both) and

does not occur as a step between U and N, then it must represent an off-pathway state. Constant force measurements generally provide a more sensitive way to detect rare or transient states and map out the entire folding network compared with FEC measurements, because they allow for more extended observation times.



**Figure 4.10 Signature of an intermediate state in constant-force measurements.** (a) An intermediate state (“I”) on the native folding pathway could be obligate, if there is only one pathway from unfolded (“U”) to native (“N”) (schematic at left), or non-obligate, if there is more than one pathway from U to N (schematic at right). (b) Constant-force measurement of a DNA hairpin with a stem sequence designed to produce a single obligate intermediate at the location marked by the arrow. An intermediate is clearly seen in the extension histogram. It is also seen directly in the extension record: as transient spikes from N to I (black circle) and from U to I (green circle), as well as in the form of brief steps during the motion between N and U (blue circle). The red box represents the area displayed with an expanded timescale.

## 2. Folding free energy and kinetics

As introduced in chapter 3, force tilts the energy landscape, changing the relative free energy difference between states and the height of the energy barrier. Therefore, the thermodynamics and kinetics of different states can be determined by directly measuring the populations and lifetimes of different states at different forces. The free energy difference between adjacent states within the landscape can be calculated from Eq. 3.5. The equilibrium force,  $F_{1/2}$ , can be determined by fitting the probability distribution for the unfolded state,  $P_u(F)$ , to Eq. 3.6. Considering that the unfolded molecule is stretched in the presence of the force, the free energy difference at zero force can be estimated by:

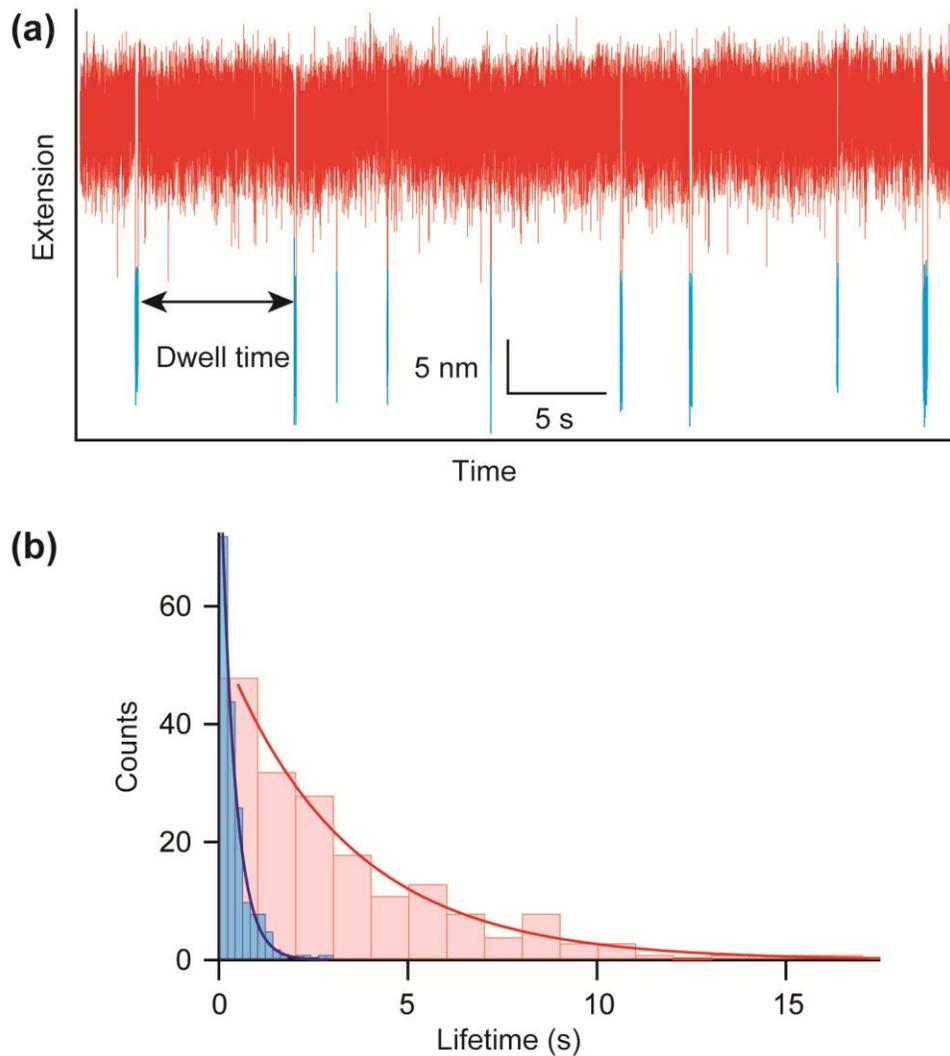
$$\Delta G^0 = F_{1/2} \cdot \Delta x(F_{1/2}) - \Delta G_{stretch}, \quad (4.14)$$

where  $\Delta G_{stretch}$  gives the energy for stretching the amino acids that unfolded in the transition from zero force to  $F_{1/2}$ . This stretching energy can be estimated by integrating the non-extensible version of the WLC equation, to obtain:

$$\Delta G_{stretch}(x) = \frac{k_B T}{L_p} \frac{L_c}{4(1-x/L_c)} \left[ 3 \left( \frac{x}{L_c} \right)^2 - 2 \left( \frac{x}{L_c} \right)^3 \right] \quad (4.15)$$

Whereas the kinetics for folding and unfolding at a given force must be extracted from the distribution of unfolding forces based on theoretical models in FEC measurements, in the case of constant-force measurements they can be read straight from the data, by measuring the dwell times in each state between the folding/unfolding transitions (Fig. 4.11a). Under constant force, for a two-state system characterized by a folding/unfolding rate  $k_{fold/unfold}$ , the probability of lifetimes follows an exponential distribution (Fig. 4.11b):

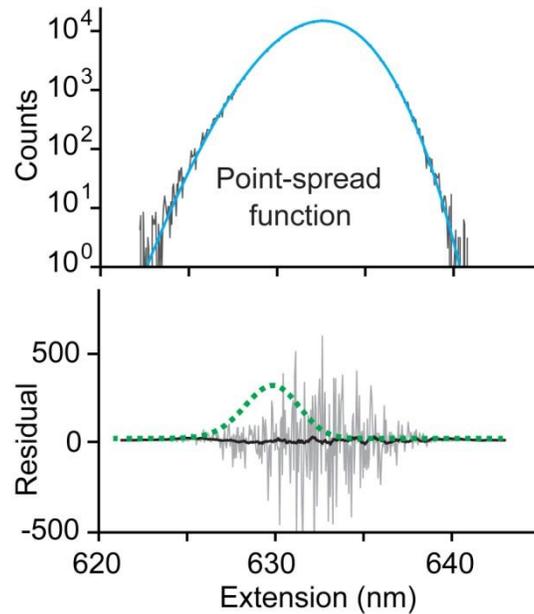
$$P_{unfolded/folded}(t) \propto \exp(-k_{fold/unfold} t). \quad (4.16)$$



**Figure 4.11 Lifetime determined from constant force measurements.** (a) A representative trajectory of a two-state folding system was separated by thresholding to determine the dwell times in each state. (b) The lifetime distributions of the folded (cyan) and unfolded state (pink) were well fitted with exponential functions (blue and red lines), yielding a folding rate of  $0.3 \text{ s}^{-1}$  and an unfolding rate of  $2.5 \text{ s}^{-1}$  at this specific force.

The folding/unfolding rate can either be determined from the exponential fitting of the lifetime distribution, or from the reciprocal of the mean lifetime of the unfolded/folded state. The position of the transition state along the reaction coordinate,  $\Delta x^\ddagger$ , can be then determined from the force-dependence of the kinetics

(Woodside et al., 2006b) using the Bell model, Eq. 3.7. Note that the Bell model assumes that the location of the transition state does not move when force is applied, an assumption that is known to be incorrect in general (Dudko et al., 2006); the more correct expression is Eq. 3.8, which assumes motion of the transition state. However, the assumption that the position of the barrier is force-independent is often borne out fairly well in constant-force studies, because only a relatively narrow range of forces is measured. As a consistency check,  $F_{1/2}$  can also be determined from the force at which folding and unfolding rates are equal and compared to the value obtained from the probability distribution  $P_u(F)$ .



**Figure 4.12 Point-spread function.** A histogram of the extension from a constant-force record of the reference construct lacking protein (upper graph, grey) shows the point-spread function (PSF) of the trap fit to Eq. 4.17 for the PSF (cyan). Counting noise in the residual (bottom graph, grey) from the histogram binning is smoothed in a 2.5-nm window (black). The green curve illustrates how a possible short-lived state would appear.

### 3. Point spread function and detection of short lived states

Special protocols have been developed in this thesis for detecting short-lived or rare intermediate states, which are not directly visible from the extension trajectories. One of the methods is to compare histograms of the molecular extension with the point-spread function (PSF) of the optical trap. The PSF is the distribution of extensions that would be expected for a construct with a fixed length, due to thermal fluctuations; it can be measured using a reference construct consisting of just DNA handles. On our instrument, extension records measured at constant force with a 50 kHz bandwidth for  $\sim 100$  s displayed distributions that were almost Gaussian but partly asymmetric (Fig. 4.12). For a harmonic potential with fixed stiffness, the PSF should be Gaussian (Gebhardt et al., 2010; Woodside et al., 2006a). The asymmetry we observe is due to the anharmonicity of the trapping potential used for the passive force clamp (Greenleaf et al., 2005): the amplitude of the Brownian fluctuations in the position of the bead is here sufficiently large that the bead explores a significant portion of the anharmonic portion of the potential well. The trap stiffness experienced by this bead therefore varies across the distribution of bead positions. At high displacements from the trap centre (corresponding to low molecular extensions) the bead visits regions of negative stiffness, thereby decreasing the effective stiffness of the system (traps plus molecule), whereas at low displacements (corresponding to high molecular extensions) the bead visits regions of positive stiffness, increasing the effective system stiffness. The low-extension side of the Gaussian distribution expected for a harmonic trap is therefore stretched out, whereas the high-extension side is compressed.

These effects can be modeled phenomenologically by describing the PSF as a pseudo-Gaussian function with an extension-dependent width parameter:

$$P(x) = \begin{cases} A \exp \left[ -\frac{(x-x_0)^2}{2(\sigma - c|x-x_0|)^2} \right], & x > x_0 \\ A \exp \left[ -\frac{(x-x_0)^2}{2(\sigma + c|x-x_0|)^2} \right], & x < x_0 \end{cases} \quad (4.17)$$

Here  $c$  represents the extension-dependent stretching or compression of the distribution width, and  $\sigma$  represents the extension-independent component of the width. This function fits the data very well, leaving effectively no residual other than counting noise (Fig. 4.12). When fitting extension histograms for a molecule of interest, residuals left after fitting the PSF can be checked for consistency with the shape of the trap PSF: any PSF-shaped residual would indicate the presence of an additional, rarely occupied state.

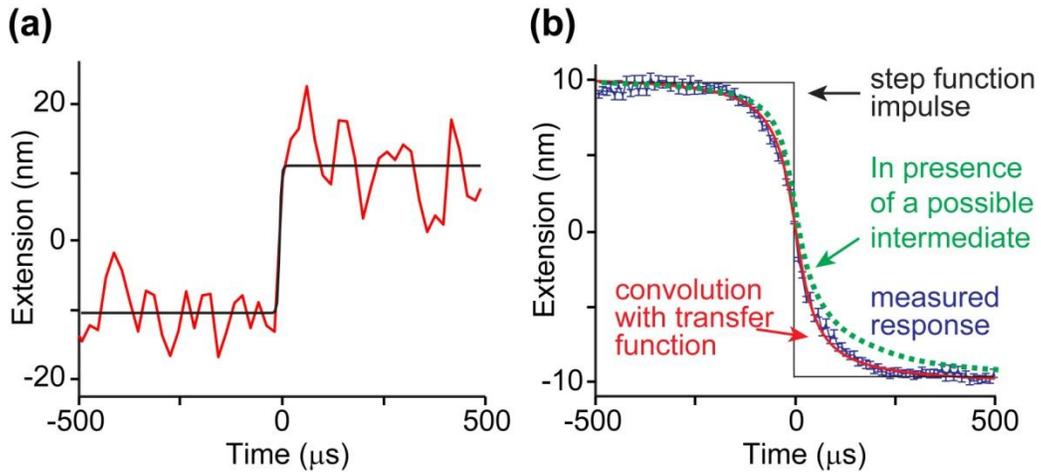
#### 4. Time response, transition paths, and detection of short-lived intermediate states

In principle the extension transition in constant force measurements records the actual transition path of the molecule during the folding/unfolding process. However in practice, the temporal response of the current optical tweezers systems is not fast enough to extract the real transition path signal from the molecule of interest. However, intermediate states, such as local minimums of the energy surface along the transition path, if there are any, might be stable enough to be observed during the brief transition time between adjacent states. In order to search for short-lived intermediate states which are not directly obvious within a single transition, say from N to U (or U to N), multiple transitions must be aligned and averaged to reduce the effects of Brownian motion and determine the shape (extension vs time) of the folding and unfolding transitions accurately. The transitions can be aligned by fitting each transition to the logistic function,

an analytic approximation to the Heaviside step function which empirically matches the shape of the transitions reasonably well:

$$X(t) = \frac{1}{1 + \exp[-\alpha(t - t_0)]}, \quad (4.18)$$

where  $X$  is the end-to-end extension,  $t_0$  is the time at the centre of the transition, and  $\alpha$  represents the slope of the transition (as  $\alpha \rightarrow \infty$ ,  $X(t) \rightarrow \Theta(t)$ , the Heaviside step function). The records are aligned on the  $t_0$  fit values for each transition (Fig. 4.13a), and then averaged.



**Figure 4.13 Instrument response function.** (a) Fitting the folding/unfolding transitions at constant force. The extension recorded during a single unfolding transition (red) at constant force, is fit by Eq. 4.18 (grey). (b) The average instrument response (blue) to a step impulse function (black) is well-approximated by the convolution of the impulse function with a Lorentzian transfer function (red). The green curve illustrates a possible presence of an intermediate state.

The instrument response can be determined from the response of the optical trap to a step function signal (Fig. 4.13b, blue) using the reference DNA handle construct.

Holding the reference construct between the traps, the traps can be moved suddenly ( $< 1$  μs), and the resulting motion of the reference construct measured with the detectors. Such

measurements reflect the intrinsic filtering effects of all aspects of the instrumentation (including the effects of the beads and handles); they can be used to determine the effective transfer function which smooth the actual extension change in the protein according to  $O(t) = X(t) \otimes H(t)$ , where  $O(t)$  is the observed extension record,  $X(t)$  is the actual extension time series of the protein itself, and  $H(t)$  is the transfer function of the instrument. We applied this method to our trap, averaging 200 measurements of the response of the optical trap detectors as described above for the unfolding/refolding transitions. Approximating the fast impulse used to measure the response as a true step function, the response is described well if the transfer function is Lorentzian (Fig. 4.13b, red); the time response of the trap is on the order of  $\sim 100 \mu\text{s}$ .

We note that the presence of a very short-lived intermediate state during the transition would change the shape of the average extension in Fig. 4.13b. It is possible to model the shape of the folding transition expected at constant force if an obligate intermediate I were present. By comparing the experimental data with simulated curves, the lifetime or the upper boundary of the lifetime of on-pathway intermediate states can be determined, which will be discussed in Chapter 7. This approach is similar to the method used to estimate transition times from single-molecule FRET trajectories (Chung et al., 2012).

The transition path time for protein folding being measured to date is on the order of a few  $\mu\text{s}$ , as found by computational simulation (Shaw et al., 2010) and experimental measurement using fluorescence (Chung et al., 2012). Since this is much faster than the time-response of the optical trap ( $\sim 100 \mu\text{s}$ ), the actual transition of the protein is not directly extractable. However, the transition path time can be extracted indirectly via an

alternative method using the parameters from the kinetic analysis of the FECs and constant force trajectories, which we will discuss in detail in Chapter 8.

## 5. Free Energy landscape reconstruction from equilibrium measurements

Free energy landscape profiles can be reconstructed based on equilibrium measurements as well as from FECs. The physical principle is that the free energy  $\Delta G(x)$  and the probability density  $P(x)$  of the states are related by the Boltzmann factor:

$$\Delta G_{\text{eqm}}(x) = -k_{\text{B}}T \ln(P(x)), \quad (4.19)$$

where  $x$  is the reaction coordinate.  $P(x)$ , the probability distribution of the molecular extension, is found by measuring the extension at high bandwidth ( $>50$  kHz) near  $F = F_{1/2}$ . This method of energy landscape reconstruction based on equilibrium method has been validated by Woodside et al. through quantitative comparisons of the landscape profiles measured for different DNA hairpin sequences to a parameter-free model of the sequence-dependent landscapes (Woodside et al., 2006a). This method has been proven to be accurate and has been applied recently to measure protein folding landscapes (Gebhardt et al., 2010), but it is technically demanding, requiring very high spatio-temporal resolution and stability. It is also difficult to apply for proteins with slow folding rates, because of insufficient statistics. Furthermore, for proteins sampling different folding pathways at the same time, the method is problematic due to the difficulty in isolating the desired pathway from the measurement. In this case, the previous Hummer-Szabo method may be more appropriate, because different pathways can be isolated kinetically.

## Appendix: Optical trapping measurements

### 1. Calibration

Prior to any measurements, the instrument was first calibrated following standard procedures. The stiffness of each trap was calibrated independently using three methods: power spectral analysis of the bead position fluctuations, variance of the bead position, and Stokes drag of the bead in the trap, as described previously (Neuman and Block, 2004). The stiffness was found by averaging the three results for many different beads. The positions of the beads in each trap were calibrated by raster scanning beads through the volume of the detector beam, mapping the detector voltages to the bead positions by interpolation.

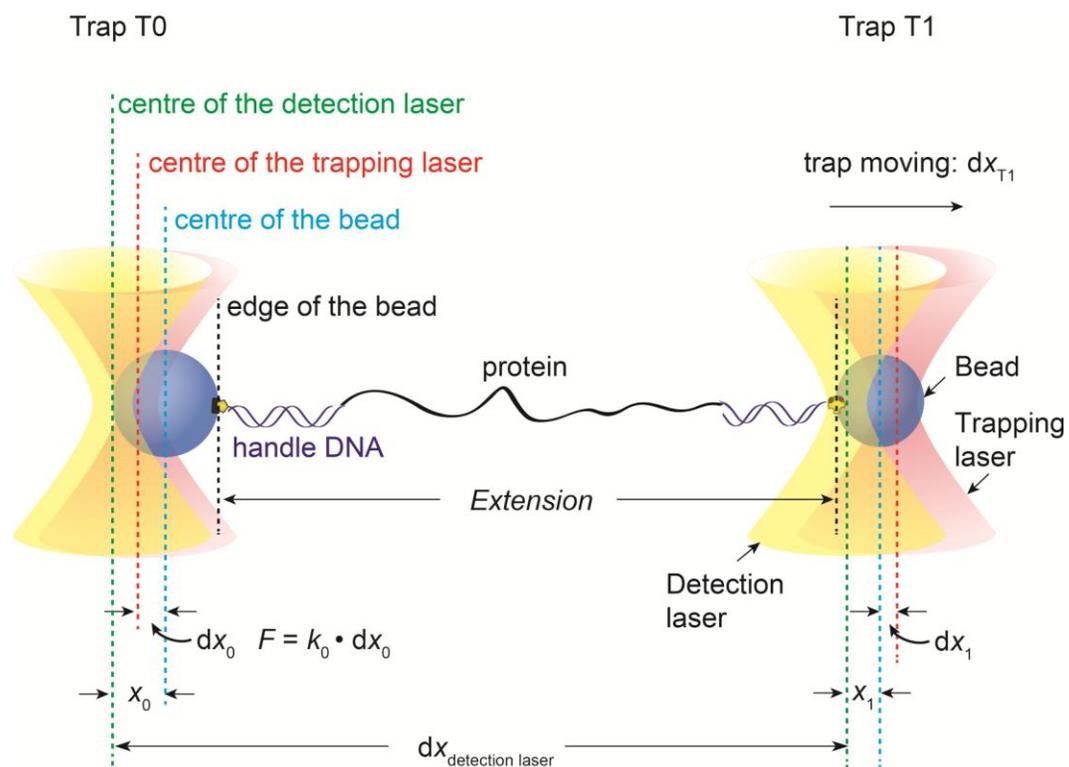


Figure 4.14 The distance and force calculations in SMFS measurements.

## 2. FEC measurements

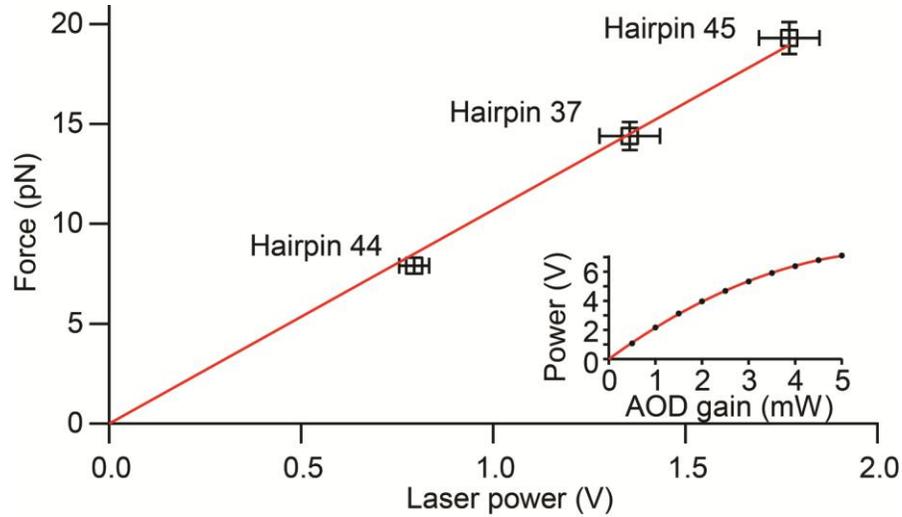
FEC measurements were performed by moving the two trapping beams apart or together at a constant speed by modulating the voltage on the EOD. Typically, the trap stiffnesses were 0.3 and 0.9 pN/nm. The speed of trap separation was typically 10–300 nm/s, generating a loading rate (rate of change of the applied force) ranging from ~1–50 pN/s. Data were sampled at 20 kHz, filtered online with an 8-pole 10 kHz Bessel filter, and averaged over each step.

For FECs, the real-time position was tracked for only one of the beads using the PSDs, following a method described previously (Abbondanzieri et al., 2005). The displacement of the bead with respect to the centre of the trapping laser (Fig. 4.14) is  $\mathbf{dx}_0 = \mathbf{x}_0 + \mathbf{dx}_{T0}$ , where  $\mathbf{x}_0$  is the position of the bead and  $\mathbf{dx}_{T0}$  is the position of the trap T0 in the detection space. Force was calculated based on  $\mathbf{F} = \mathbf{k}_0 \cdot \mathbf{dx}_0$ , where  $\mathbf{k}_0$  is the stiffness of the trap T0. The position of the bead in trap T1 was inferred from the trapping force  $\mathbf{F}$  measured in T0:  $\mathbf{x}_1 = \mathbf{dx}_1 - \mathbf{dx}_{T1} = \mathbf{F} / \mathbf{k}_1 - \mathbf{dx}_{T1}$ , in which  $\mathbf{dx}_{T1}$  is the position of the trap T1 and  $\mathbf{k}_1$  is the stiffness of T1. The extension (Fig. 4.14) of the molecule was then calculated from:

$$\begin{aligned} \textit{Extension} &= \mathbf{dx}_{\text{detection laser}} + \mathbf{x}_1 - \mathbf{x}_0 - r_{\text{bead0}} - r_{\text{bead1}} \\ &= \mathbf{dx}_{\text{detection laser}} + ((\mathbf{k}_0 \cdot (\mathbf{x}_0 + \mathbf{dx}_{T0})) / \mathbf{k}_1 - \mathbf{dx}_{T1}) - \mathbf{x}_0 - r_{\text{bead0}} - r_{\text{bead1}} \end{aligned}$$

where  $\mathbf{dx}_{\text{detection laser}}$  is the separation between the centre of the two detection lasers, which can be calibrated in advance,  $r_{\text{bead0}}$  (410 nm) and  $r_{\text{bead1}}$  (300 nm) are the radius of the two beads,  $\mathbf{k}_0$  and  $\mathbf{k}_1$  are the calibrated stiffnesses of the traps,  $\mathbf{dx}_{T1}$  is the real-time motion of

the trap T1 (the control parameter of the measurement). The only measured quantity is thus  $x_0$ .



**Figure 4.15 Force calibration using DNA hairpins for constant force measurements.** The AOD amplitude corresponding to the  $F_{1/2}$  of each of the DNA hairpin was measured at constant force for three different molecules and averaged. The nonlinear relation between the AOD amplitude and the laser power was also characterized to be  $Power = 2.3339 \cdot aodgain - 0.18333 \cdot aodgain^2$ . Finally, the  $F_{1/2}$  of each hairpin from previous studies (Woodside et al., 2006b) was plotted against the laser power and a linear relation was obtained.

### 3. Constant force measurements

Constant force measurements were made using a passive force clamp (Greenleaf et al., 2005), following a previously-described approach (Abbondanzieri et al., 2005; Greenleaf et al., 2008). Data were typically sampled at 50 kHz and filtered online with a 25 kHz 8-pole Bessel filter. The stiffnesses of the traps were usually set to be 0 and 0.3 pN/nm. The position of beads in both T0 and T1 ( $x_0$  and  $x_1$ ) were measured in real time independently. The extension of the molecule (Fig. 4.14) was then calculated based on:

$$Extension = dx_{\text{detection laser}} + x_1 - x_0 - r_{\text{bead0}} - r_{\text{bead1}}.$$

The extension in constant force measurements thus represents the differential signal of the traps, in which the relative motions between the two traps were eliminated. Because the forces inferred from the calculated stiffness of the traps contained systematic errors in the constant force measurements, we also calibrated the force (Fig. 4.15) by comparing to the folding of standard DNA hairpins which were found previously to fold/unfold over forces ranging from 5 to 20 pN (Woodside et al., 2006a; Woodside et al., 2006b).

## **5. SMFS of the *add* adenine riboswitch relates folding to regulatory mechanism\***

This chapter demonstrates SMFS can be used to measure the hierarchical folding/misfolding and map the energy landscape, using a relatively well understood RNA riboswitch molecule as an example. Riboswitches regulate gene expression via ligand binding to an aptamer domain which induces conformational changes in a regulatory expression platform. By unfolding and refolding single *add* adenine riboswitch molecules in an optical trap, an integrated picture of the folding was developed and related to the regulatory mechanism (Neupane et al., 2011).

### **5.1 Introduction of the *add* adenine riboswitch**

Riboswitches are gene regulatory elements in mRNA which modulate gene activity by sensing the concentration of associated small metabolites (Baird and Ferré-D'Amaré, 2010; Coppins et al., 2007; Roth and Breaker, 2009). Found in a wide range of organisms, they are often bipartite in structure, consisting of an aptamer domain which senses the target ligand and an expression platform which modulates gene expression. Ligand binding stabilises the aptamer domain through changes in secondary and/or tertiary interactions, typically switching the structure of the expression platform between two alternatives and thereby altering transcription, translation, splicing, or other processes

---

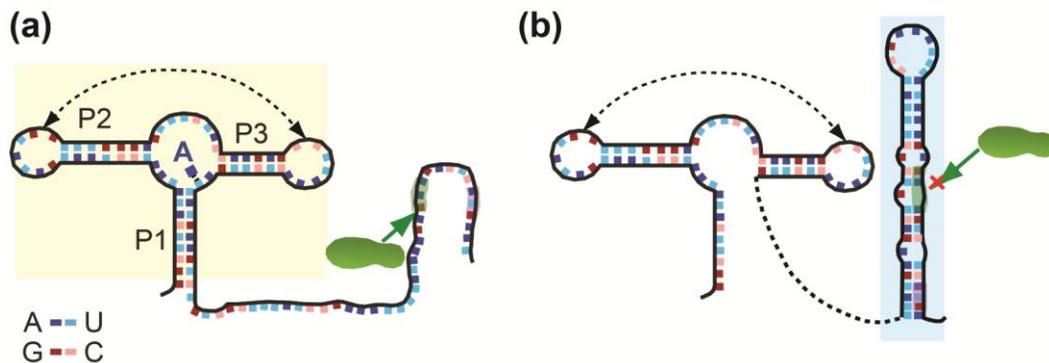
\* A version of this chapter has been published. Neupane, K., Yu, H., Foster, D. A. N., Wang, F. and Woodside, M. T. Single-molecule force spectroscopy of the *add* adenine riboswitch relates folding to regulatory mechanism. *Nucl Acids Res* **39**, 7677-7687 (2011). In this work, H.Y. built the instrument with D.A.N.F., performed experiments and analyzed data with K.N.. RNA riboswitches were made by F.W. and D.A.N.F.

during expression (Baird and Ferré-D'Amaré, 2010; Coppins et al., 2007; Roth and Breaker, 2009).

The purine riboswitches are among the smallest and structurally simplest riboswitches, yet collectively they still demonstrate much of the variety of mechanisms used by more complex riboswitches (Kim and Breaker, 2008). For example, the *xpt* guanine riboswitch controls expression via transcription termination (Mandal et al., 2003), the *pbuE* adenine riboswitch through transcription anti-termination (Mandal and Breaker, 2004), and the *add* adenine riboswitch through translation activation (Serganov et al., 2004). Despite such different mechanisms, the aptamers of these riboswitches all have very similar structures: a “tuning-fork” architecture wherein the purine binding site is located at a specific residue in a pocket formed at the junction of three helices, two of which are hairpins interacting via kissing loops and aligned on top of the third helix (Fig. 5.1) (Batey et al., 2004; Serganov et al., 2004). Comparative studies of these riboswitches thus provide an opportunity to investigate the molecular features involved in their different regulatory mechanisms (Lemay et al., 2011).

Previous studies have investigated various different features of the purine riboswitch aptamers, such as the ligand specificity (Mandal and Breaker, 2004; Noeske et al., 2005) and its structural basis (Gilbert et al., 2006; Noeske et al., 2005; Rieder et al., 2007), the kinetics of the loop-loop interaction (Lemay et al., 2006), the  $Mg^{2+}$ -dependence of the folding (Buck et al., 2010), the folding energy landscape (Greenleaf et al., 2008), the structural changes upon ligand binding (Gilbert et al., 2006; Noeske et al., 2007; Ottink et al., 2007; Stoddard et al., 2008), and the ligand binding/dissociation rates and energies, for both the aptamer and the full-length riboswitch (Lemay et al., 2011; Rieder et al., 2007; Wickiser et al., 2005a). The relation of aptamer folding to regulatory

mechanism has been less well studied, however, whether in purine riboswitches or more generally. The *pbuE* riboswitch has been shown, along with the FMN riboswitch, to function via kinetically-controlled folding of the aptamer in competition with ligand binding and transcription (Greenleaf et al., 2008; Lemay et al., 2011; Lemay et al., 2006; Wickiser et al., 2005a), while the folding kinetics of the preQ<sub>1</sub> riboswitch aptamer suggest that transcription regulation is achieved instead by ligand-induced thermodynamic changes (Rieder et al., 2010). The *add* riboswitch has been confirmed to control translation *in vitro* and *in vivo*, with expression levels indicating that regulation does not require transcription-translation coupling (Lemay et al., 2011). To date, however, little has been done to characterise the interaction between an aptamer and its corresponding expression platform, a key question for understanding and manipulating riboswitch function (Zhang et al., 2010).



**Figure 5.1 The structure of *add* adenine riboswitch.** The *add* riboswitches control translation by adopting different conformations. (a) In “ON” state, the aptamer (yellow) is fully folded and the expression platform is unfolded. The adenine (A) binding site is in the P1 loop. The ribosome binding site is exposed for the ribosome (green) to bind. The gene is on. (b) In “OFF” state, the expression platform (cyan) is fully folded and the aptamer is partially unfolded. The ribosome binding site is buried and the gene is off. The sequences of the aptamer, expression platform and the full length riboswitch used in this study are listed in Table 5.1. Figure courtesy of Krishna Neupane.

We addressed these issues by using SMFS to observe the folding and unfolding trajectories of individual *add* riboswitch molecules from *Vibrio vulnificus* held under tension in a high-resolution optical trap. We first built an integrated picture of folding in the aptamer alone, similar to previous force spectroscopy work on the closely-related *pbuE* adenine riboswitch aptamer (Greenleaf et al., 2008). Since these riboswitches have very similar aptamer structures but work by completely different regulatory mechanisms—translation activation (*add*) rather than transcription anti-termination (*pbuE*)—we can begin to discern which aspects of the folding arise from shared structural features and which may be important for the different regulatory mechanisms. We then extended these measurements to the complete riboswitch, probing the interaction between aptamer and expression platform and thereby demonstrating that the regulation is likely controlled by the folding thermodynamics.

## 5.2 RNA constructs

RNA constructs were created by inserting the sequence for the *add* aptamer alone, expression platform alone, or full-length riboswitch into the pMLuc-1 plasmid. A 77-bp DNA fragment coding for the *add* adenine riboswitch aptamer and 3 flanking nucleotides on each side (see Table 5.1) was inserted between the BamH I and Spe I sites of a pMLuc-1 plasmid (Novagen). A 2,226-bp, linear, double stranded (ds) DNA transcription template was amplified by PCR from this plasmid, with a T7 promoter placed in the upstream primer. The completed template consisted of 1,289 basepairs (bp) of linker sequence upstream of the aptamer, followed by the aptamer itself, and then an additional 842 bp of linker downstream of the aptamer. These 2,208 bp of DNA were transcribed

with MEGAscript high yield transcription kit (Ambion). The resulting transcription template containing the riboswitch (or its constituent parts) flanked by linker regions on either side was amplified by PCR and transcribed *in vitro*. Two single stranded (ss) DNA handles designed to be complementary with the upstream and downstream linker sequences flanking the *add* A-riboswitch aptamer were created by asymmetric PCR (Saiki et al., 1986): primary PCR was first performed with both forward and reverse primers for each handle on its own, then the purified ds PCR products were used as templates in asymmetric PCR reaction with only one primer present. The 3' end of the DNA handle complementary to the upstream (5' end) linker section of the transcript was labelled with dig-ddUTP using terminal transferase (Roche). The DNA handle complementary to the downstream (3') linker of the transcript was functionalised with biotin on the 5' end of the PCR primer.

The RNA transcript and the two ssDNA handles complementary to the linker sequences on either side of the aptamer were annealed by gradually increasing the temperature to 85°C, then stepping down to 62 and 52°C for 90 minutes each in a thermal cycler. The RNA-DNA hybrid thereby created consisted of a 1,289-nt RNA-DNA duplex handle labelled with digoxigenin at one end, 77 nt of RNA containing the aptamer, and a 842-bp RNA-DNA duplex handle labelled with biotin at the other end. The aptamer is separated from the duplex handles by 3 nt of unstructured ssRNA, to uncouple interactions between the handles and the aptamer. Constructs containing the expression platform alone and the full-length riboswitch were created in the same way, using either a 58-bp DNA insert coding for the *add* adenine riboswitch expression platform, or a 117-bp insert coding for the full-length riboswitch, in both cases with flanking nucleotides on each side (Table 5.1). The RNA constructs were prepared by Feng Wang.

**Table 5.1 Sequences of RNA molecules measured: *add* riboswitch aptamer (cyan), expression platform (green), full-length riboswitch (red represents the overlapping between the aptamer and the expression platform), and oligomer sequences complementary to portions of the aptamer.**

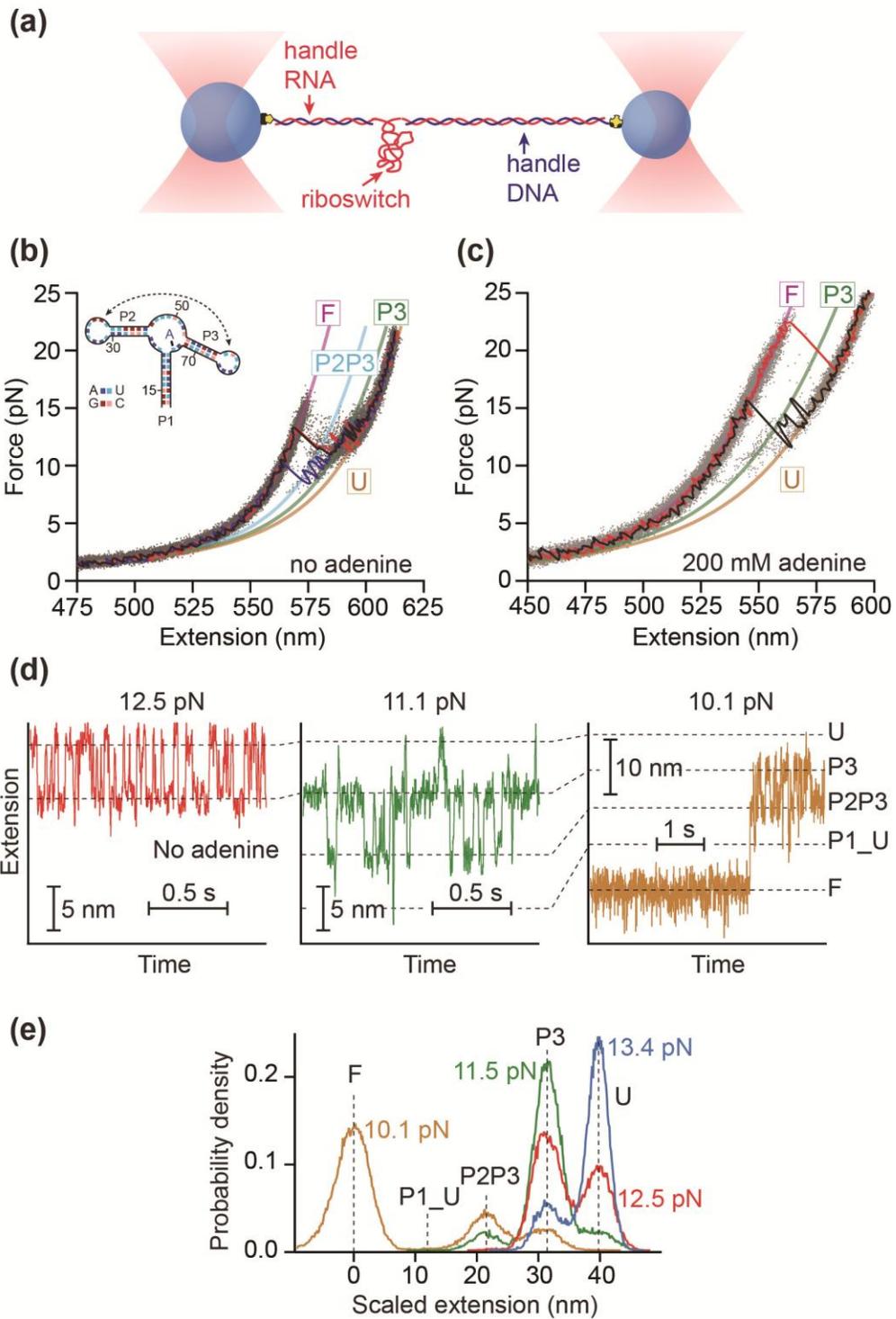
<b>Aptamer sequence</b>	5'-CGCGGCTTCATATAATCCTAATGATATGGTTTGGGAGTTTC TACCAAGAGCCTTAAACTCTTGATTATGAAGTC <i>TGT</i> -3' The single-stranded linkers to the duplex handles are listed in italics.	
<b>Expression platform</b>	5'-CCATTATGAAGTCTGTCGCTTTATCCGAAATTTTATAAAGA GAA GACTCATGAATCCC-3'	
<b>Full-length sequence</b>	5'-CGCGGCTTCATATAATCCTAATGATATGGTTTGGGAGTTTC TACCAAGAGCCTTAAACTCTTGAT <i>T</i> ATGAAGTC <i>T</i> GTCGCTTT ATCCGAA ATTTTATAAAGAGAAGACTCATAAATC-3'	
<b>Oligomers complementary to aptamer</b>		
complementary to junction J2/3 between P2 and P3 (46-53)	5'-GTAGAAAC-3'	
complementary to 5' strand of P1 (13-22)	5'-ATATGAAGCC-3'	
complementary to 5' strand and loop of P2 (25-38)	5'-CCATATTATTAGGA-3'	
complementary to loop and 3' strand of P3 (59-72)	5'-CAAGAATTTAAGGC-3'	

The final hybrid constructs were diluted to a concentration of 160 pM and mixed with equal volumes of 600-nm and 820-nm diameter polystyrene beads (Bangs) at concentrations of 250 pM. The smaller beads were coated with avidin DN (Vector Labs), and the larger beads were coated with polyclonal anti-digoxigenin (Roche). The bead-construct mixture was incubated for 1 hour at room temperature, then diluted 100-fold into RNase-free measuring buffer: 50 mM MOPS pH 7.5, 130 mM KCl, 4mM MgCl<sub>2</sub>, and 200 U/mL RNase inhibitor (SUPERase•In, Ambion). An oxygen scavenging system consisting of 40 U/mL glucose oxidase (Sigma), 185 U/mL catalase (EMD), and 250 mM

D-glucose (Sigma) was also included in the buffer. The diluted dumbbells were placed in a flow chamber prepared on a microscope slide, and inserted into the optical trap.

### 5.3 Folding intermediates of the aptamer

First the aptamer alone was measured, using FECs and trajectories at constant force to identify states in the folding pathway. The optical trapping assay is illustrated schematically in Fig. 5.2a. Three FECs measured in the absence of adenine (Fig. 5.2b: black, blue, and red lines) display typical features: the force,  $F$ , at first rises nonlinearly with extension due to the elasticity of the double-stranded (ds) handles (Liphardt et al., 2001; Smith et al., 1996), until characteristic “sawtooth” patterns occur when the length increases abruptly due to unfolding of different parts of the aptamer (Liphardt et al., 2001). Four states are seen with different contour lengths. The change in the contour length associated with each unfolding transition,  $\Delta L_c$ , was found by fitting the aggregated FEC data from hundreds of pulls on the same molecule (Fig. 5.2b: grey dots) to two extensible WLC models (Wang et al., 1997) in series, one for the ds handles, and a second for the unfolded ss RNA (Eq. 4.3) (Seol et al., 2004). The average values of  $\Delta L_c$  determined from more than 5,200 FECs measured on 4 different molecules were:  $\Delta L_c = 39.6 \pm 0.5$  nm from the fully folded state (labelled “F”, purple line) to the fully unfolded state (labelled “U”, yellow line),  $\Delta L_c = 19 \pm 1$  nm from F to the shorter intermediate (labelled “P2P3”, cyan line), and  $\Delta L_c = 30.3 \pm 0.5$  nm from F to the longer intermediate (labelled “P3”, green line). All uncertainties quoted in this thesis represent standard error on the mean (s.e.m.).



**Figure 5.2 Force spectroscopy of *add* aptamer alone.** (a) RNA containing the riboswitch aptamer flanked by two kilobase-long handle sequences was annealed to DNA strands complementary to the handles and attached to beads held in optical traps. (b) FECs in the absence of adenine reveal 2 intermediate states corresponding to the unfolding of everything except hairpins P2 and P3, followed by hairpin P2. Three FECs (black, blue, red) are plotted above the aggregated data from 700 FECs (grey dots). Worm-like chain fits are shown for the four states: F—fully-folded (purple); U—fully-unfolded (yellow); P3—P3 folded (green); and P2P3—both P2 and P3 folded (cyan). (c) Adenine binding resulted in similar behavior but higher unfolding forces. State P2P3 was observed less frequently if at all. (d) The extension as a function of time at different levels of constant force in the absence of adenine reveals 5 distinct states corresponding to the major structural features: fully-unfolded at the largest extension (“U”), then P3 folded (“P3”), then both P2 and P3 folded but no loop-loop interaction (“P2P3”), then the P2-P3 kissing loop complex (labelled “P1\_U”, P1 unfolded), and finally fully-folded (“F”). Only short segments of the full records are displayed. (e) Histograms of the full extension records show force-dependent occupancies of the states. The P1\_U state always has low occupancy.

We related these states to the unfolding of distinct structural elements in the aptamer (Fig. 5.2b, inset) via the  $\Delta L_c$  values expected from these structures:

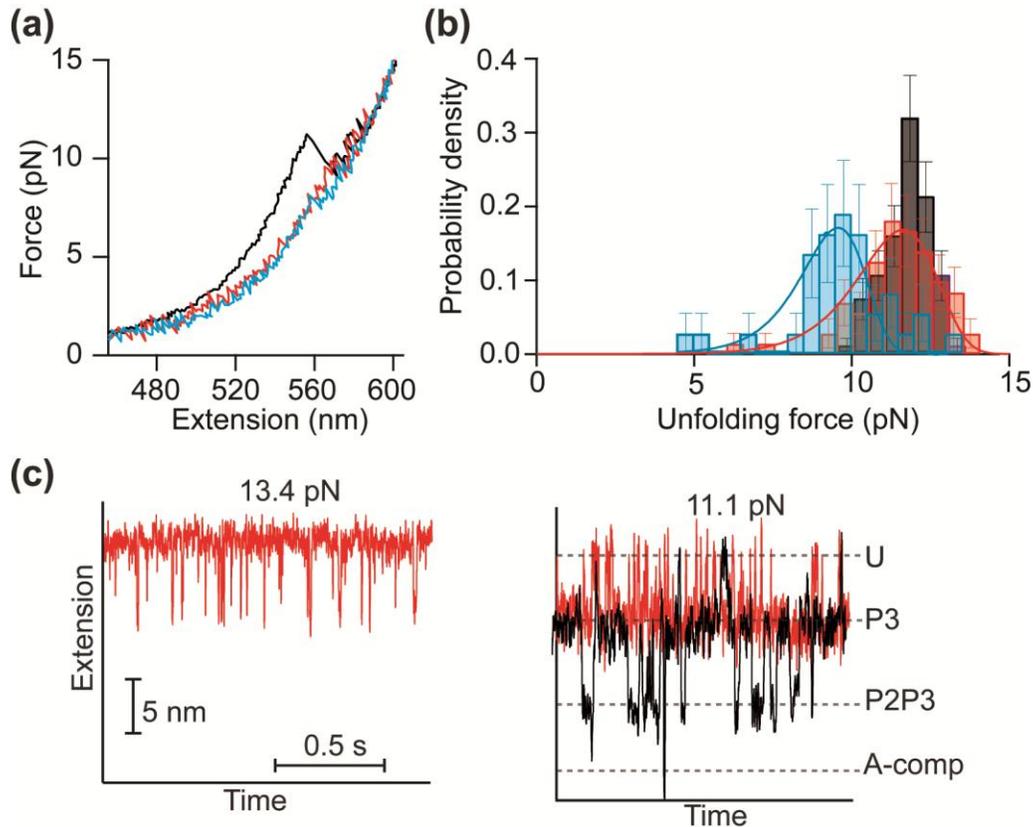
$\Delta L_c = N_{nt} \cdot L_c^{nt} - \Delta n_h \cdot d_h$ , where  $N_{nt}$  is the number of nucleotides unfolded,  $L_c^{nt} = 0.59$  nm is the contour length/nt for ssRNA (Saenger, 1984),  $\Delta n_h$  is the number of RNA helices removed during unfolding, and  $d_h = 2.2$  nm is the diameter of the A-form dsRNA helix (Saenger, 1984). The  $\Delta L_c$  values expected for different structures are listed in Table 5.2, alongside the measured values. We found excellent agreement with the expected value for full unfolding (F to U), confirming that the aptamer is being completely unfolded. P1 unfolding by itself was not observed, but the intermediate labelled P2P3 matched the result expected for unfolding P1 and the loop-loop interaction (*i.e.* everything but P2 and P3). The last intermediate (labelled P3) was in good agreement with the expectation for unfolding everything except P3. Notably, whereas state F never reformed once it was unfolded in a given FEC, multiple transitions between states P2P3, P3, and U were often observed (Fig. 5.2b), indicating much faster kinetics for hairpins P2 and P3 than for helix P1 and the junction.

**Table 5.2. Aptamer unfolding contour length changes.** The expected contour length changes from unfolding different structural components of the aptamer are listed, along with the values obtained from analysis of FECs in the presence and absence of adenine. Uncertainties represent the standard error.

Transition	F to U	F to P1_U	F to P2P3	F to P2	F to P3
$\Delta L_c$ expected	39.7 nm (71 nt)	10.6 nm (18 nt)	20.5 nm (31 nt)	29.5 nm (50 nt)	30.7 nm (52 nt)
$\Delta L_c$ measured (no adenine)	$39.6 \pm 0.5$ nm		$19 \pm 1$ nm		$30.3 \pm 0.5$ nm
$\Delta L_c$ measured (with adenine)	$40.1 \pm 0.8$ nm		$20.6 \pm 0.8$ nm		$30.4 \pm 0.4$ nm

However, since hairpins P2 and P3 differ by only 2 nt in length, additional confirmation for this state was obtained using an anti-sense DNA oligomer complementary to the 5' stem and loop of hairpin P2 to block P2 folding, and another complementary to the loop and 3' stem of hairpin P3 to block P3 folding (sequences listed in Table 5.1). FECs measured in the presence of 10  $\mu$ M of the P2-blocking oligomer (Fig. 5.3a, red) show that indeed the only unfolding transition that occurs is the one at the highest forces which had been previously identified by the extension change as P3. The contour length change for this remaining transition corresponds to  $19.5 \pm 0.5$  nt from 550 FECs. Since 19 nt are expected for P3, these results confirm the identification of the P3 folding transition. In contrast, FECs measured in the presence of 10  $\mu$ M of the P3-blocking oligomer (Fig. 5.3a, blue) show a single unfolding event but with a longer contour length change:  $22 \pm 1$  nt determined from 165 FECs, matching the 21 nucleotides expected for P2 unfolding. Just as importantly, the unfolding force distribution for P2 (i.e. P3 blocked: Fig. 5.3b, blue) was significantly different from the distribution for P3 (i.e. P2 blocked: Fig. 5.3b, red). We compared the distribution of unfolding forces for the

transition attributed to P3 in Fig. 5.1 (Fig. 5.3b, black) to the distributions expected for P2 unfolding (Fig. 5.3b, blue) or P3 unfolding (Fig. 5.3b, red). It clearly matches the distribution expected for P3 unfolding, and is inconsistent with the distribution for unfolding P2, further confirming that the identification of the transitions in Fig. 5.2 is correct.



**Figure 5.3 Confirmation of hairpin folding identification using anti-sense oligos.** (a) FECs under normal conditions (black) and with a DNA oligomer bound to the loop and 5' stem of hairpin P2 (red) and an oligo bound to the loop and 3' stem of hairpin P3 (cyan). The normal high-force unfolding transition (black) is the same as the transition with P3 unfolding (red), and P2 unfolding (cyan) is distinct from P3 unfolding (red) as it occurs at lower force. (b) Unfolding force distributions for P2 unfolding (cyan), P3 unfolding (red), and unfolding without blocking oligomers (black). The high-force unfolding distribution without oligomers (black) overlaps with P3 unfolding (red). P2 unfolding is at lower force, distinct from the P3 unfolding. (c) Under constant force, again only the high force transition is seen when the oligomer is bound (red), confirming that this transition involves P3 folding.

Similar behavior was observed when measurements were repeated in the presence of saturating adenine concentration ( $200 \mu\text{M} \gg K_D \sim 600 \text{ nM}$  (Rieder et al., 2007)), as shown in Fig. 5.2c. In this case, the aptamer was held at  $F \sim 0$  for 5 s between FECs to ensure that adenine binding occurred (Greenleaf et al., 2008). The principal change due to adenine binding was an increase in the typical unfolding force for the fully-folded state, and a concomitant reduction in the population of the intermediate states; P2P3 was often not observed. The average contour length changes between states, measured from over 3,000 FECs on 8 different molecules, were found to be identical to the results above within experimental uncertainty, indicating the same intermediates as when adenine is absent (Table 5.2).

To characterise the intermediate states in greater detail, folding trajectories were measured at constant force using a passive force clamp (Greenleaf et al., 2005). The aptamer was first fully unfolded at high force then the force was reduced in discrete steps, measuring the molecular extension while maintaining a constant force after each step for up to 1 minute to observe equilibrium behavior. The force was stepped down until the aptamer was fully folded. In the absence of adenine, transitions between 5 different extensions were observed (Fig. 5.2d), indicating 5 separate structural elements forming in apparently sequential order. The populations of these states changed as the force was reduced, as seen from histograms of the extension (Fig. 5.2e), with the more-folded states (at shorter extensions) becoming more populated at lower forces. All five states could, however, be observed coexisting in equilibrium at  $F \sim 10 \text{ pN}$ .

Analogously to the FEC analysis, the states were identified by converting the extension changes between them into the number of nucleotides folded, using  $N_{\text{nt}} = [\Delta x(F) + \Delta n_{\text{h}} \cdot d_{\text{h}}] / \Delta x_{\text{nt}}(F)$ , where  $\Delta x_{\text{nt}}(F)$  is the extension/nt at a given force. The first

folding transition from the unfolded state, at  $F \sim 11\text{--}14$  pN (Fig. 5.2d, red), involved  $19.0 \pm 0.4$  nt, precisely the number expected for hairpin P3. The second transition (at  $F \sim 10\text{--}12$  pN) involved  $21.3 \pm 0.4$  nt, matching the expectation for P2. These are the same partially-folded intermediates seen in the FECs (P3 and P2P3). However, the constant force measurements were sufficiently precise so as to be able to distinguish between P2 and P3: the standard error on the mean number of nucleotides involved in the transitions was only 0.4 nt. Hence the 2-nt length difference between P2 and P3 represents 5 standard deviations, and misidentification is extremely unlikely. Nevertheless, to confirm the identification, once again we measured the folding at constant force using  $10 \mu\text{M}$  of an anti-sense DNA oligomer complementary to nucleotides 25-38 (the 5' arm and loop of hairpin P2; sequence listed in Table S1) to prevent formation of P2. In this case, we should observe P3 folding but none of the other states, since they all depend on the folding of P2. As seen in Fig. 5.3c, the oligomer indeed prevented all folding transitions except the one occurring at the highest forces, which was identified from both the force range and the extension change ( $19.8 \pm 0.5$  nt) as being P3 folding. The last two transitions displayed very different behavior. At  $F < 9\text{--}10$  pN, the aptamer folded into a long-lived state  $30.3 \pm 0.7$  nt shorter than P2P3 (Fig. 5.2d, yellow). Since only 31 nucleotides remain unfolded once P2 and P3 have formed, this must be the fully-folded state F. Before F formed, however, a rarely- and transiently-occupied state  $\sim 5$  nm shorter than P2P3 was typically seen (Fig. 5.2d, green), at the extension expected when the loop-loop interaction has formed to pre-organise the binding pocket and only P1 is unfolded ( $8.6 \pm 0.7$  nt of folding measured from P2P3, 8 nt expected). We identify this as the P1 unfolded state (labelled "P1\_U" in Fig. 5.2d). This state was the only one not observed in the FECs.

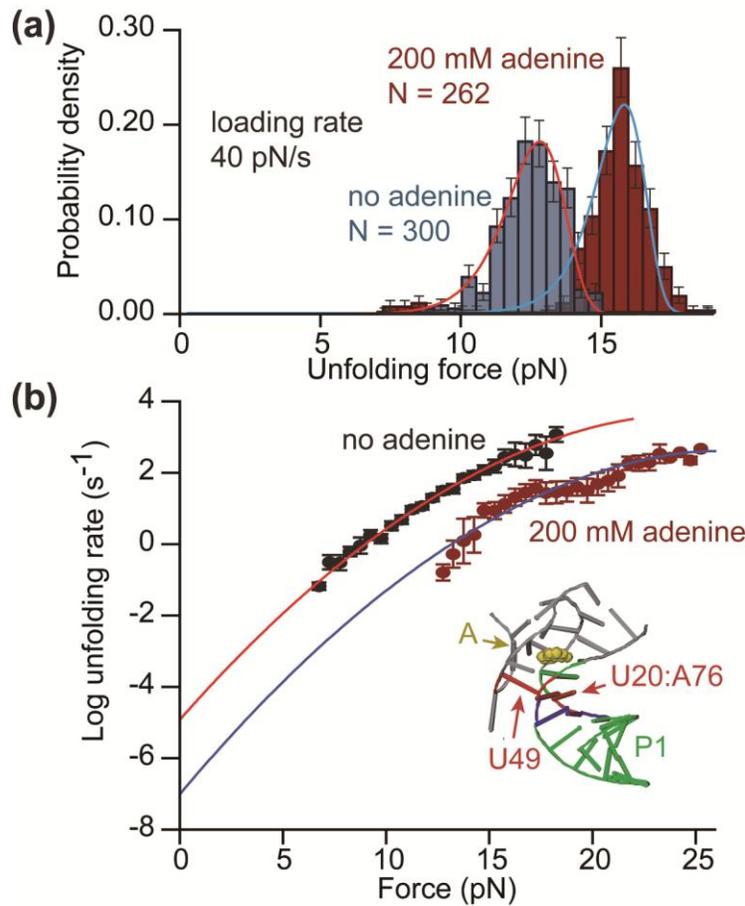
**Table 5.3. Constant force results.** “P3” represents folding of hairpin P3, “P2” represents folding of hairpin P2, “P1\_U” represents the P1-unfolded state, “NAF” represents complete folding in the absence of adenine, and “AF” represents complete folding in the presence of adenine.

Folding transition	$\Delta x$ (nm)	$\Delta x$ (nt)	$F_{1/2}$ (pN)	$\ln(k_{1/2})$ (s <sup>-1</sup> )	$\Delta G_{1/2}^\ddagger$ (kcal/mol)	$\Delta x_f^\ddagger$ (nm)	$\Delta x_u^\ddagger$ (nm)
P3	5.2 ± 0.2	19.0 ± 0.5	12.9 ± 0.2	4.2 ± 0.1	4.3 ± 0.1	2.7 ± 0.2	4.1 ± 0.3
P2	6.3 ± 0.2	21.3 ± 0.5	10.8 ± 0.2	4.1 ± 0.1	4.4 ± 0.1	2.5 ± 0.3	4.6 ± 0.4
P1_U	5.1 ± 0.2	8.6 ± 0.6	10.1 ± 0.2	4.3 ± 0.2	4.3 ± 0.1	2.1 ± 0.2	3.9 ± 0.2
NAF	7.7 ± 0.4	22 ± 1	10 ± 2	2 ± 1	6 ± 1	N/A	N/A
AF	7.7 ± 0.4	22 ± 1	N/A	N/A	N/A	N/A	N/A

When 200 μM adenine was added, the qualitative behavior of the aptamer was unchanged, with the same progression of states seen at similar forces. The principal differences involved helix P1: it folded at a slightly higher force, and most prominently, its lifetime was greatly increased. Indeed, once P1 folded in the presence of adenine, it was never observed to unfold again (on the typical timescale of ~1 min) without a large force increase. Adenine thus clearly shifts the equilibrium for P1 folding, moving the mid-point between folded and unfolded P1,  $F_{1/2}$ , to a higher force that is above the equilibrium for formation of the kissing loops and possibly even hairpins P2 and/or P3. As a result, with adenine bound it was not possible to observe equilibrium folding of P1 under constant force. The results of the constant force analysis are summarised in Table 5.3.

## 5.4 Energy landscape of the aptamer

Using the measurements above, we reconstructed the 5-state folding landscape of the aptamer. Because P1 folding was never in equilibrium in the presence of adenine, we first used the Jarzynski equality (Jarzynski, 1997) to determine from the FECs the free energy change between F and P2P3. The free energy for unfolding P1, the binding pocket junction, and the kissing loop complex was found to be  $10 \pm 1$  kcal/mol without adenine and  $18 \pm 2$  kcal/mol with 200  $\mu$ M adenine. The free energies of secondary structures can be calculated using *mfold* (Zuker, 2003), with an estimated 25% reduction to account for the different buffer ionic strength and temperature for our measurement conditions (Liphardt et al., 2001). Therefore, the prediction for the stability of P1 and the junction loops is  $-11.5$  kcal/mol at 37 °C and 1 M monovalent salt, which would be equivalent to  $\sim -8.6$  kcal/mol in our experiments. Comparing the free energy of the aptamer with and without tertiary contacts, our results indicate a stabilisation energy of  $\sim -1.5$  kcal/mol from tertiary interactions that form in the absence of adenine (principally the loop-loop complex), and a more substantial  $-8$  kcal/mol of additional stabilisation upon adenine binding. This is consistent with NMR results showing some tertiary interactions without adenine but considerable stabilisation of tertiary interactions upon adenine binding (Buck et al., 2007; Noeske et al., 2007), with a binding energy of  $\sim -8$  kcal/mol as calculated from dissociation constants obtained by fluorescence studies (Rieder et al., 2007).



**Figure 5.4 Unfolding force distributions and force-dependent rates determined from FECs. (a)** The unfolding force is higher with adenine bound (red) than without adenine (blue). Solid lines: fits to Eq. 4.4. Error bars show standard deviation. **(b)** Unfolding rate as a function of force, fit with Eq. 3.8. Error bars show s.e.m. The barrier to unfolding is located at the first base triple between P1 and the junction loop (inset, red).

We next found the height and location of the barrier for unfolding P1 by analysing the distributions of unfolding forces from the FECs. Representative unfolding force distributions measured at similar loading rates ( $\sim 40$  pN/s) with and without adenine (Fig. 5.4a) show a typical increase of several pN in unfolding force upon adenine binding. These distributions are well fit by the Dudko model (Eq. 4.4) (Dudko et al., 2006). Cumulative probability analysis yielded the lifetime of the folded state as a function of

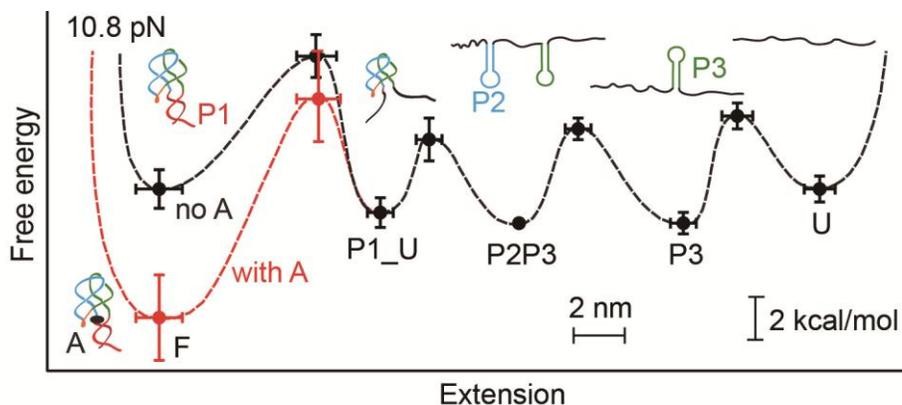
force and hence the unfolding rate  $k(F)$  (Dudko et al., 2008), which was well fitted by the same type of landscape model using Eq. 3.8 (Fig. 5.4b).

Distributions measured at loading rates ranging from 5–400 pN/s, with and without adenine, were analysed by both methods. We averaged the results obtained under the two limiting cases for the shape of the energy barrier ( $\nu = 1/2$ ,  $\nu = 2/3$ ), since this shape is unknown. The unfolding rate at zero force was found to be  $1.5 \times 10^{-5} \text{ s}^{-1}$  ( $\log k_{\text{unfold}} = -4.8 \pm 0.1$ ) without adenine and  $3 \times 10^{-7} \text{ s}^{-1}$  ( $\log k_{\text{unfold}} = -6.5 \pm 0.4$ ) with adenine bound. The barrier height was  $12.3 \pm 0.4 \text{ kcal/mol}$  without adenine, and  $16 \pm 2 \text{ kcal/mol}$  with adenine bound. The barrier position did not depend on adenine binding:

$\Delta x_{\text{N}}^{\ddagger,0} = 6.2 \pm 0.6 \text{ nm}$  with adenine and  $6.1 \pm 0.5 \text{ nm}$  without. Since ssRNA has an extension of 0.40–0.45 nm/nt in the force range 12–20 pN matching the most probable unfolding forces, this result places the transition state for unfolding at 7–8 bp along helix P1. A comparison to the crystal structure of the ligand-bound aptamer (Serganov et al., 2004) reveals that the transition state is located next to the first basepair in P1 involved in triplex interactions with the junction loops: nucleotides U20:A76 in P1 and U49 in the loop, shown in red in Fig. 5.4b (inset). These interactions therefore appear to act as structural keystones preventing unfolding.

Finally, these FEC results for P1 unfolding were combined with the constant force results for all the other transitions to reconstruct piecewise the energy landscape for the native folding pathway of the aptamer at a constant force of  $F = 10.8 \text{ pN}$ , with and without the binding of an adenine molecule (Fig. 5.5). The transitions other than P1 unfolding were analysed as sequential two-state processes. The relative positions and energies of the states were determined from constant-force extension histograms, whereas

the positions and heights of the energy barriers between states were determined from the force-dependent kinetics (Greenleaf et al., 2008; Woodside et al., 2006b).



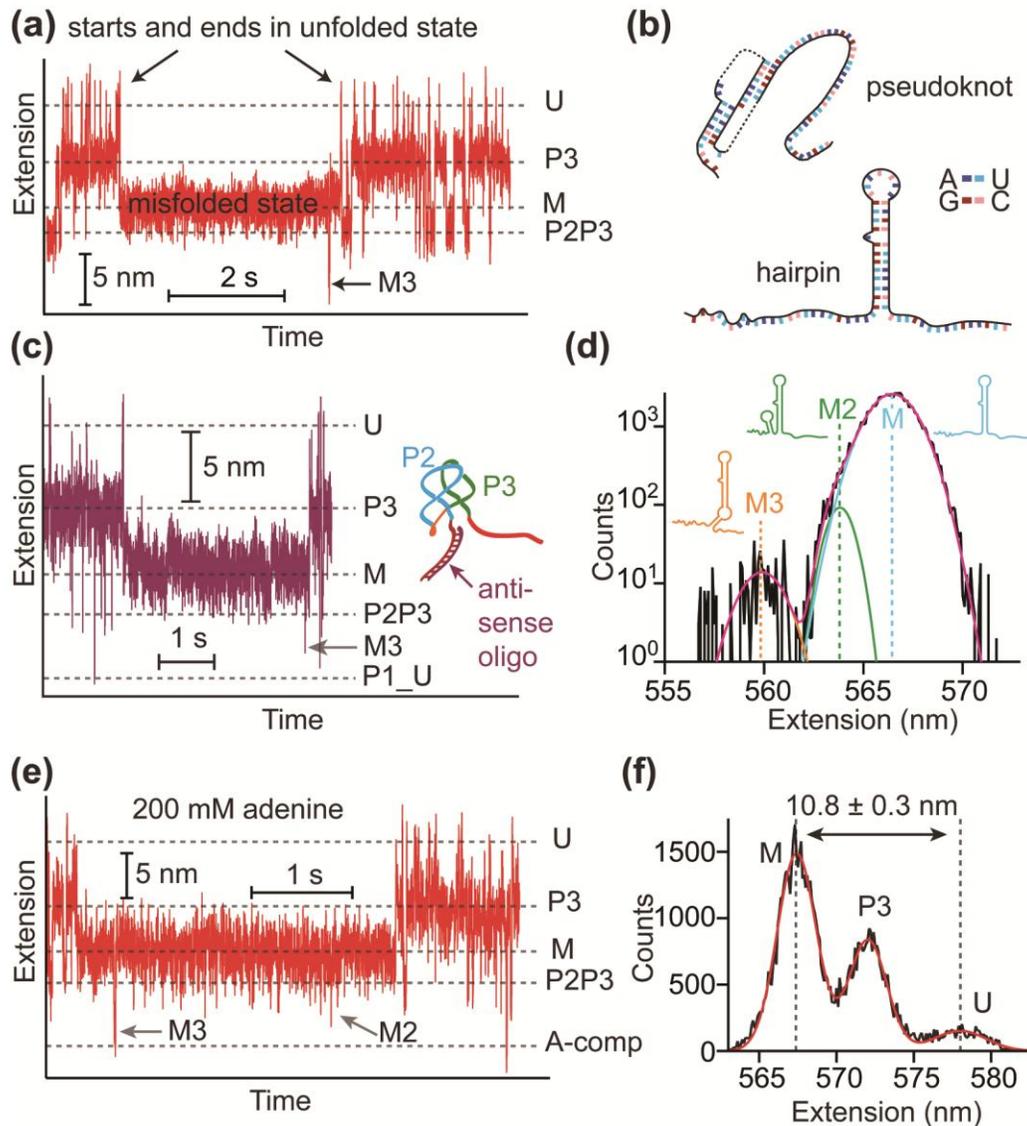
**Figure 5.5 Folding energy landscapes of the *add* aptamer.** The key features of the energy landscapes for the 5-state native folding pathway were reconstructed from piecewise two-state analyses of each transition. Energies and positions are plotted with reference to the P2P3 state. Error bars show s.e.m. Dotted lines indicate notional landscape shapes in the presence (red) and absence (black) of adenine. Cartoons indicate the structural features associated with each state.

## 5.5 Misfolded states in the aptamer

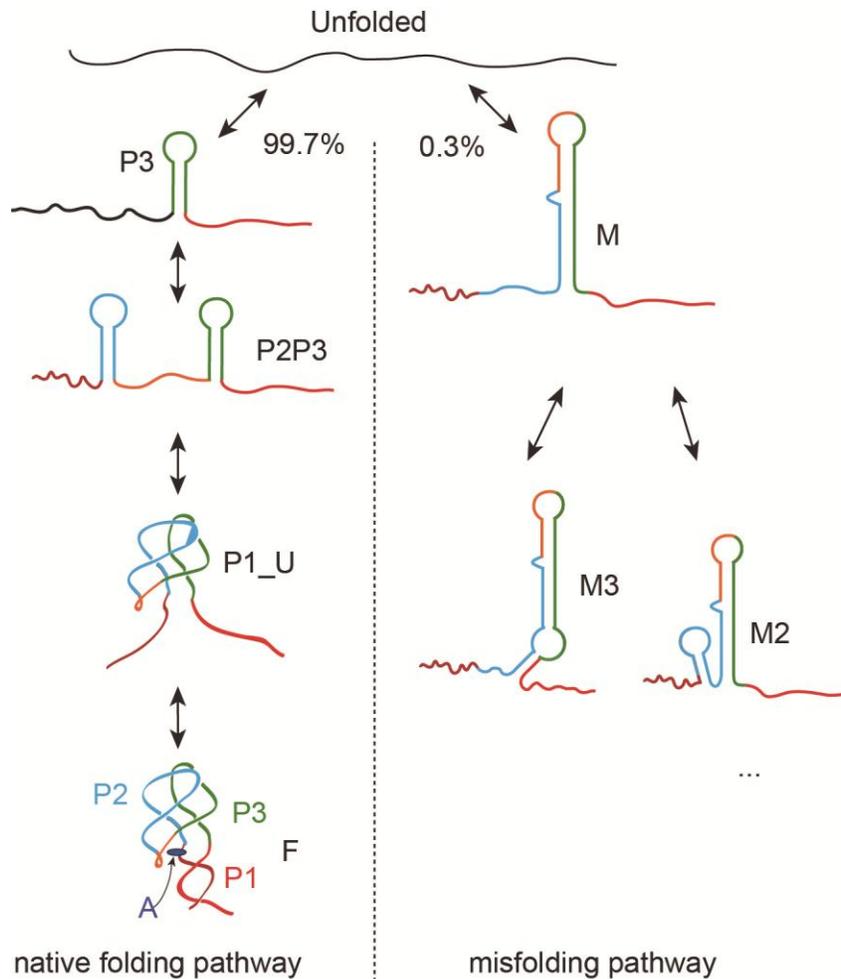
Intriguingly, an additional set of states with different extensions and distinct features (Fig. 5.6a) was observed occasionally at medium forces ( $\sim 10$  pN). Most prominent was a long-lived state labelled “M” in Fig. 5.6a. This state was observed for all molecules, independent of adenine concentration (Fig. 5.6e), occurring at a rate of  $\sim 1$   $\text{min}^{-1}$  with a lifetime of  $\sim 1$ – $10$  s. Within state M, additional transient folding events were also seen, such as the spike labelled “M3” in Fig. 5.6a, with lifetimes of  $\sim 10$  ms. The extension change from U to M, determined from extension histograms (Fig. 5.6f), corresponds to  $31 \pm 1$  nt of folded RNA. This does not match the length of any structural feature in the regular folding pathway of the aptamer. As the aptamer almost always ( $\sim 95\%$

of the time) entered and exited this state via the fully unfolded state, rather than any of the partially folded intermediates, we therefore attribute it to an off-pathway, “misfolded” structure.

A search of suboptimal structures using *mfold* (Zuker, 2003) and Hotknots (Andronescu et al., 2003) revealed two possible candidates for state M: a potential pseudoknot involving 29 nucleotides that would normally form hairpin P2 and the 5' end of P1, or a hairpin made from 32 nucleotides which would normally form the 3' stem of P2 through to the 5' stem of P3 (Fig. 5.6b). By using anti-sense oligos to block in turn the 5' stem of P1 and the 8-nt junction loop between P2 and P3 (J2/3), we found that M still forms when P1 is blocked (Fig. 5.6c) but not when J2/3 is blocked, indicating that M is the hairpin, not the pseudoknot. Evidence for at least two additional, transient folding transitions out of state M was found from the extension histograms (Fig. 5.6d): one to a state that is  $2.6 \pm 0.5$  nm shorter than M (labelled “M2”) and another to a state that is  $7.1 \pm 0.5$  nm shorter (“M3”). These distances are consistent with two of the extra misfolded states that may form along in concert with M (inset, Fig. 5.6d): an 11-nt hairpin causing an extension decrease of 2.5 nm (M2: nucleotides 24–34) and a 16-nt helix-bulge decreasing the extension by 6.5 nm (M3: nucleotides 30–36 and 69–77).



**Figure 5.6 Misfolded states.** (a) Occasionally a state with extension between P3 and P2P3 (labelled “M”) forms from the unfolded state. It does not lead to any of the other states on the pathway to the natively folded state F—it begins and ends at U—but it does lead occasionally to other misfolded states with different extensions, such as the state M3. (b) Two potential misfolded structures include a pseudoknot (nucleotides 17-45) and a hairpin (37-68). (c) With a DNA oligomer bound to the 5’ strand of P1, the misfolded state still occurs, indicating that misfolding does not involve the pseudoknot. The P1\_U state is still observed even when P1 is prevented from forming. (d) Extension histograms of the off-pathway, misfolded states reveal 3 distinct misfolded states which are well fit by Gaussian distributions. Insets: structures associated with each state. (e) The same misfolded states were observed with adenine present as without adenine. (f) The extension change from the unfolded state (U) to the misfolded state (M) was determined from extension histograms.



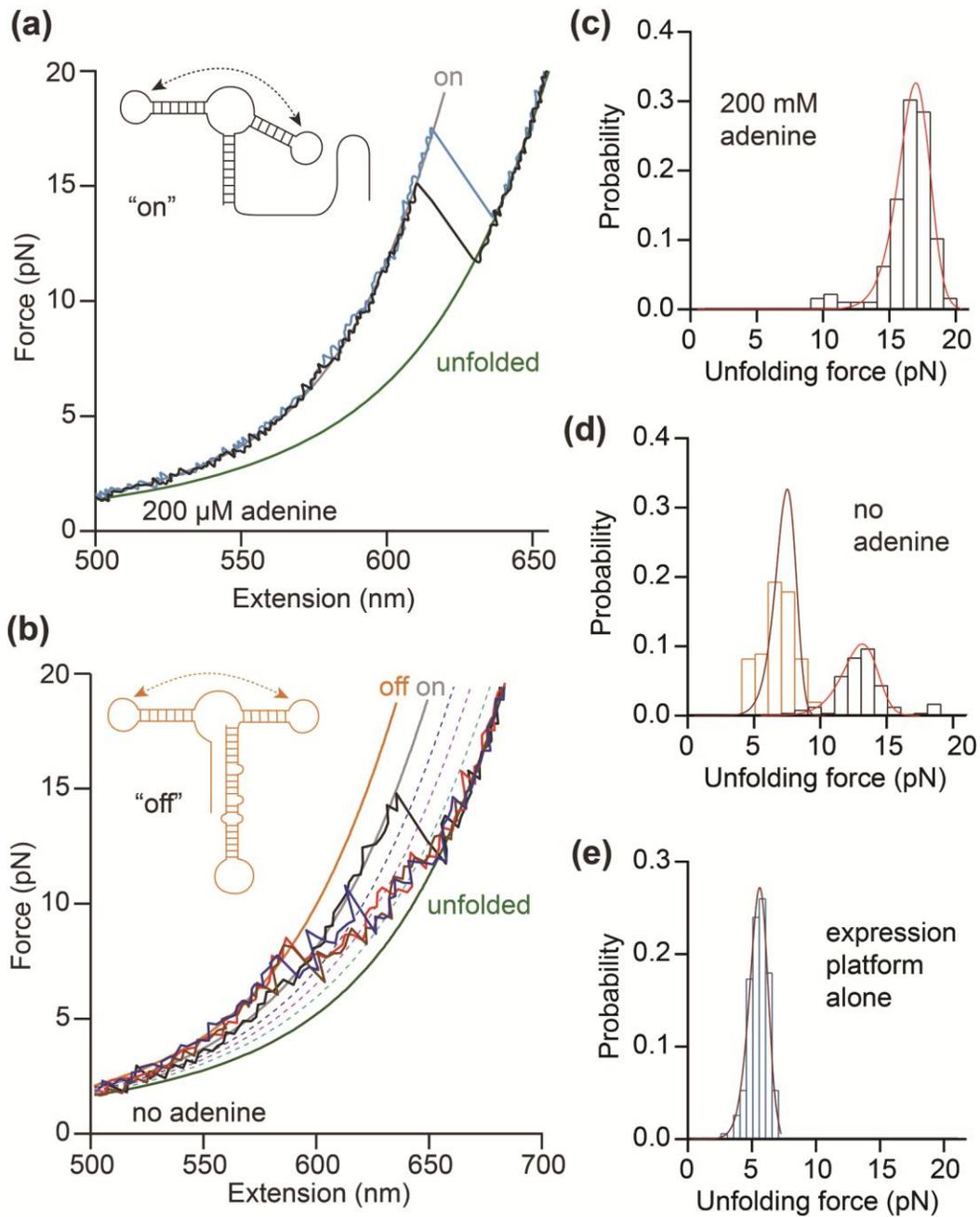
**Figure 5.7 Folding pathways of the aptamer.** The native pathway involves sequential formation of hairpins P3 and P2, then the P2-P3 kissing loop complex (P1\_U), and finally the fully-folded aptamer. 0.3% of the time the unfolded state folds into state M, preventing formation of the adenine binding pocket and leading to other off-pathway, misfolded states.

The misfolded state clearly prevents the adenine binding site from forming, which would negate the regulatory ability of the riboswitch. However, we do not expect state M to be biologically relevant: its folding rate is much slower than that of P2 and P3, and it forms only very rarely in our in-vitro assay (only ~ 0.3% of the folding transitions initiated from U go to M). It would likely form even more rarely co-transcriptionally, because P2 would be expected to fold first. Numerous RNAs are known to misfold into

non-native structures simply due to the rugged energy landscape generated by alternative patterns of basepairing (Russell, 2008), although few have been studied with single-molecule methods allowing individual trajectories to be followed (Li et al., 2007; Russell et al., 2002). Combining all the results above, we can build an integrated picture of the alternative pathways, each with multiple states (Fig. 5.7).

## 5.6 Full length riboswitch

In order to probe the interaction between the aptamer and the expression platform, we also made constructs containing the full riboswitch sequence, consisting of the aptamer plus the expression platform. FECs were measured by pulling the traps apart until the riboswitch unfolded, then rapidly bringing the traps together, waiting 5 s for folding and ligand binding, and repeating. FECs measured in the presence of 200  $\mu\text{M}$  adenine (Fig. 5.8a) look very similar to FECs of the adenine-bound aptamer alone: the riboswitch unfolds between two principal states at relatively high force, with an unfolding force distribution (Fig. 5.8c) that matches the result found for the adenine-bound aptamer (Fig. 5.4a, red). The contour length change between the two states, determined from WLC fits to 990 FECs from 2 molecules, was  $\Delta L_c = 40.7 \pm 0.5$  nm. This is very close to  $\Delta L_c$  for complete unfolding of the aptamer alone (39.7 nm), indicating that the aptamer is almost always folded, the expression platform is unstructured, and the riboswitch is in the “on” state for gene expression (Fig. 5.8a, inset).



**Figure 5.8 FECs of the full-length riboswitch.** (a) With adenine bound, the aptamer is folded (“on” state) and the contour length change is the same as for the aptamer alone. Solid lines: fits to “on” (grey) and unfolded (green) states. Inset: secondary structure in the “on” state. (b) Without adenine, the expression platform is folded (red, blue, brown) more often than the aptamer (black). Solid lines indicate fits to “on” (grey), “off” (orange), and unfolded (green) states. Dotted lines indicate partially-folded intermediates of the aptamer alone: P1\_U (blue), P2P3 (pink), and P3 (cyan). Inset: secondary structure

in the “off” state. **(c)** The unfolding force distribution of FECs measured with adenine present matches that expected for the adenine-bound aptamer. Red line: fit to Eq. 4.4. **(d)** The unfolding force distribution without adenine reveals two sub-populations. The higher-force peak (black) comes from curves initially in the “on” state and matches the distribution expected for the adenine-free aptamer (red line: fit to Eq. 4.4). The lower peak (orange) comes from curves initially in the “off” state and matches the distribution expected for the expression platform alone (brown line: fit to Eq. 4.4). **(e)** Unfolding force distribution of the expression platform alone (brown line: fit to Eq. 4.4). The most likely force is lower than for the “off” state in (d) due to a 15-fold lower loading rate.

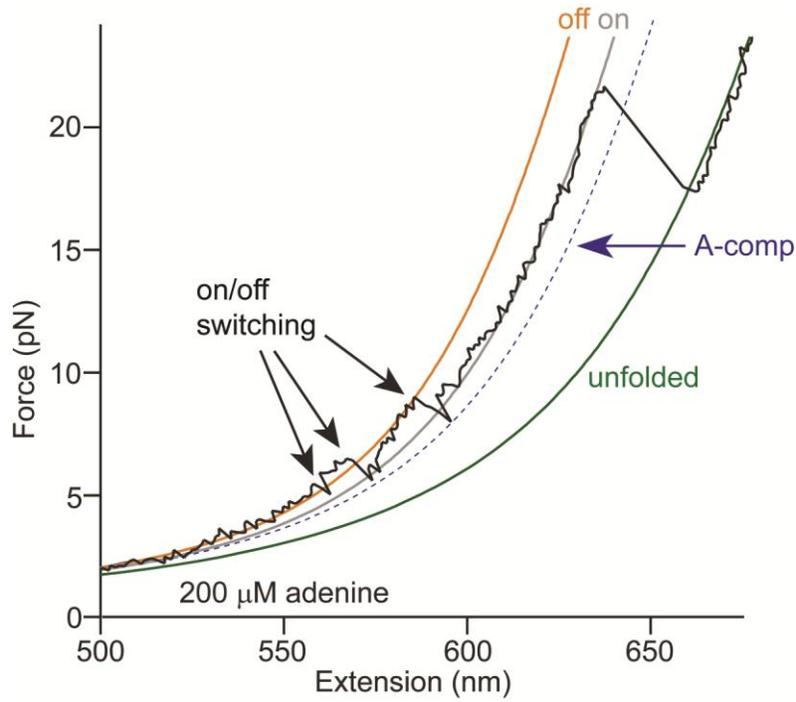
When no adenine is present, the signature of aptamer unfolding is still observed some of the time (Fig. 5.8b, black), but a majority of the FECs display qualitatively different behavior (Fig. 5.8b: red, blue, and brown). These two sub-populations of unfolding behavior are reflected very clearly in the unfolding force distribution, which reveals two distinct peaks for the force of the first unfolding event in each curve (Fig. 5.8d). The higher-force peak (Fig. 5.8d, black) results from the curves where the contour length change indicates the aptamer is initially folded and the riboswitch is “on” (Fig. 5.8b, black). Fits to this part of the distribution using Eq. 4.4 (Fig. 5.8d, red) return the parameters expected for the adenine-free aptamer (compare to Fig. 5.4a, blue), confirming the identification of the state as “on”. The riboswitch was observed to be “on” approximately 35% of the time, even though no adenine was present. This “leakage” rate under conditions that should turn the expression off is within the typical range of rates for similar regulatory elements such as transcription terminators (Nojima et al., 2005; Reynolds et al., 1992).

The lower-force peak is due to the other, majority class of FECs (Fig. 5.8b: red, blue, and brown), which lack the single prominent sawtooth feature seen for the aptamer alone and display instead a larger number of smaller, repeated unfolding/refolding transitions. In these curves, the first unfolding event occurs at considerably lower force (~ 5–8 pN) than for the “on” state, as seen from the unfolding force distribution (Fig. 5.8d,

orange). The contour length of these curves also starts off shorter than that of the “on” state: fitting the low-force region of these FECs to a WLC model (Fig. 5.8b, orange) reveals  $\Delta L_c = 52.2 \pm 0.9$  nm to the fully-unfolded state (Fig. 5.8b, green), measured from 1989 FECs on 4 molecules. This agrees well with the value of 53.4 nm expected for unfolding the full riboswitch in the “off” state, assuming that the expression platform is folded as well as P2, P3, and the loop-loop complex (Fig. 5.8b, inset). The low unfolding force observed is also consistent with unfolding the expression platform, since its three bulges and low G:C content reduce its mechanical stability (Woodside et al., 2006a; Woodside et al., 2006b). Indeed, the low-force peak of the distribution in Fig. 5.8d matches the unfolding force distribution measured using constructs that contain the expression platform alone, shown in Fig. 5.8e (1400 FECs on 2 molecules). With no adenine present, we observed the riboswitch to be in the “off” state  $\sim 65\%$  of the time. Interestingly, the riboswitch is in the “on” state  $\sim 2.5$  times more frequently when adenine is present than when it is not, matching well the  $\sim 3$ -fold increase in expression levels observed *in vitro* (Lemay et al., 2011).

In the FECs measured without adenine, the riboswitch is seen to fluctuate between 6 different states: the 5 states of the aptamer alone and the “off” state with the expression platform folded. Indeed, what looks like noise in the FECs between the “off” state and the unfolded state is consistent with frequent inter-conversion between the different partially-folded states of the aptamer, shown in Fig. 5.8b as dotted lines. However, the unfolding of the full riboswitch from the “off” state is not simply sequential, as for the aptamer alone, even though the unfolding force for the expression platform is several pN lower than that of any other structural element of the riboswitch. Instead, the

presence of two alternative, fully-folded states is observed directly, through switching between the “on” and “off” states.



**Figure 5.9 Switching and ligand binding.** A riboswitch, initially “off” in the presence of adenine, switches several times between “off” and “on”, eventually binding adenine and remaining in the thermodynamically more stable adenine-bound “on” state before unfolding completely.

Examples of this behavior are seen in Fig. 5.8b (brown curve at ~ 6 pN, blue curve at ~ 8–9 pN). It is even seen very occasionally with adenine present, as in Fig. 5.9. Here, the riboswitch is initially folded in the “off” state (without adenine bound), but at  $F \sim 5$  pN the contour length twice increases briefly to the “on” state value before returning to the “off” state, *i.e.* the folded expression platform is transiently replaced by a folded P1. A last switching event is seen at ~ 7 pN, when the “off” state unfolds to the contour length expected for the P1-unfolded state (*i.e.*, with repressor hairpin completely unfolded) before P1 folds to produce the “on” state one last time. Adenine clearly binds

to the aptamer soon thereafter, since the “on” state is stabilised and does not unfold until a high force characteristic of the adenine-bound state.

## 5.7 Discussion

These results provide new insight into the relation between folding and functional mechanism in the *add* riboswitch, both at the level of the aptamer alone as well as the full-length riboswitch, especially through comparison to previous results for the *pbuE* riboswitch (Greenleaf et al., 2008). While the two riboswitches have aptamers with very similar structures, they operate by quite different mechanisms: translation activation for *add* as opposed to transcription anti-termination for *pbuE*. Comparison of the behavior of the aptamers reveals many commonalities due to the similar structures, but also some important differences. In both cases, a total of 5 states was observed, and the aptamers were found to fold sequentially: first the two hairpins (P2 and P3) form, then the loop-loop interaction, and finally P1. Crucially, the helix P1—which acts as the mechanical switch—is the last part of the molecule to fold. In both cases, this helix is strongly stabilised by adenine binding, whereas the energetic and kinetic properties of the other states are not affected (Fig. 5.5). The reconstructed energy landscapes are in fact quantitatively very similar, due to the extensive structural and sequence similarities between the two aptamers.

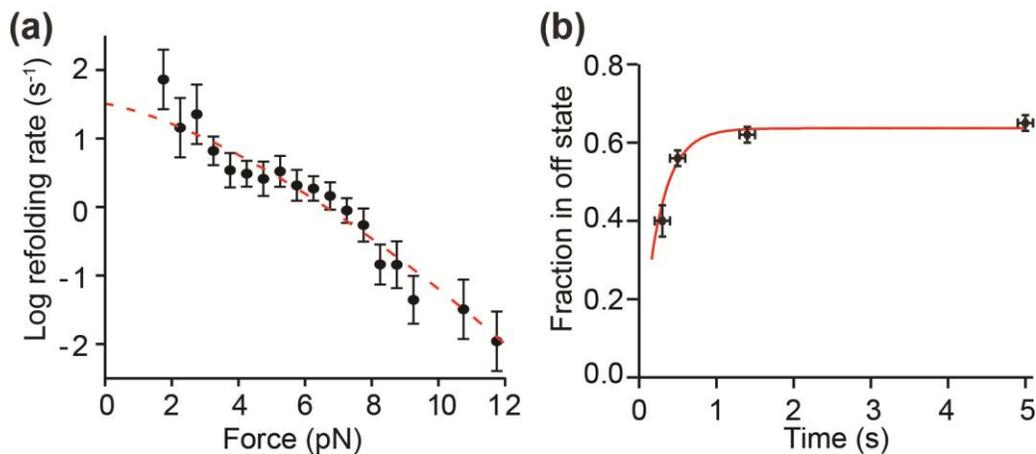
The landscapes in Fig. 5.5 also highlight some of the important differences between the aptamers. First, the order of folding of the two hairpin loops is reversed: P3 folds at higher force than P2 in the *add* aptamer, vice versa for the *pbuE* aptamer. This effect was predicted by simulations (Lin and Thirumalai, 2008) and is due to the different

relative stabilities of the helices P2 and P3 in the two different aptamers. More fundamental is the difference in stability of helix P1. In the *add* aptamer, P1 is similar in stability to P2 and P3 even in the absence of adenine ( $F_{1/2} \sim 10$  pN) and it has a low unfolding rate at  $F = 0$  ( $k_{\text{unfold}} \sim 10^{-5} \text{ s}^{-1}$ ). In contrast, P1 in the *pbuE* aptamer is quite unstable without adenine bound ( $F_{1/2} = 3$  pN) and has an unloaded unfolding rate  $k_{\text{off}}$  that is 10,000 times higher. As a result, the *pbuE* aptamer is much more susceptible to strand invasion by the expression platform, the process which is required to form the terminator hairpin and switch the gene off.

The different stabilities of P1 appear to be related directly to the different functional mechanisms of the riboswitches. For the *pbuE* riboswitch, the conformation with the terminator hairpin folded and aptamer unfolded is the most stable thermodynamically, regardless of whether or not adenine is bound to the aptamer. Both *in vitro* (Greenleaf et al., 2008; Lemay et al., 2006; Wickiser et al., 2005b) and *in vivo* (Lemay et al., 2011) measurements point to a mechanism whereby the outcome of riboswitch folding (and hence level of gene expression) is governed by a competition between the kinetics of aptamer folding, ligand binding, and transcription elongation. In such a situation, an unstable P1 is what is needed to permit strand invasion on a timescale competitive with transcription rates. Kinetic control of riboswitch folding has also been demonstrated for the FMN riboswitch (Wickiser et al., 2005b). Our measurements on the full-length *add* riboswitch, in contrast, show quite clearly that the thermodynamically stable conformation switches upon adenine binding: the state with the aptamer folded and expression platform unfolded (“on”) is dominant, and hence most stable, when adenine is bound, but the state with the expression platform folded and aptamer partially-unfolded is dominant when adenine is not bound. Results such as those in Fig. 5.9 encapsulate the

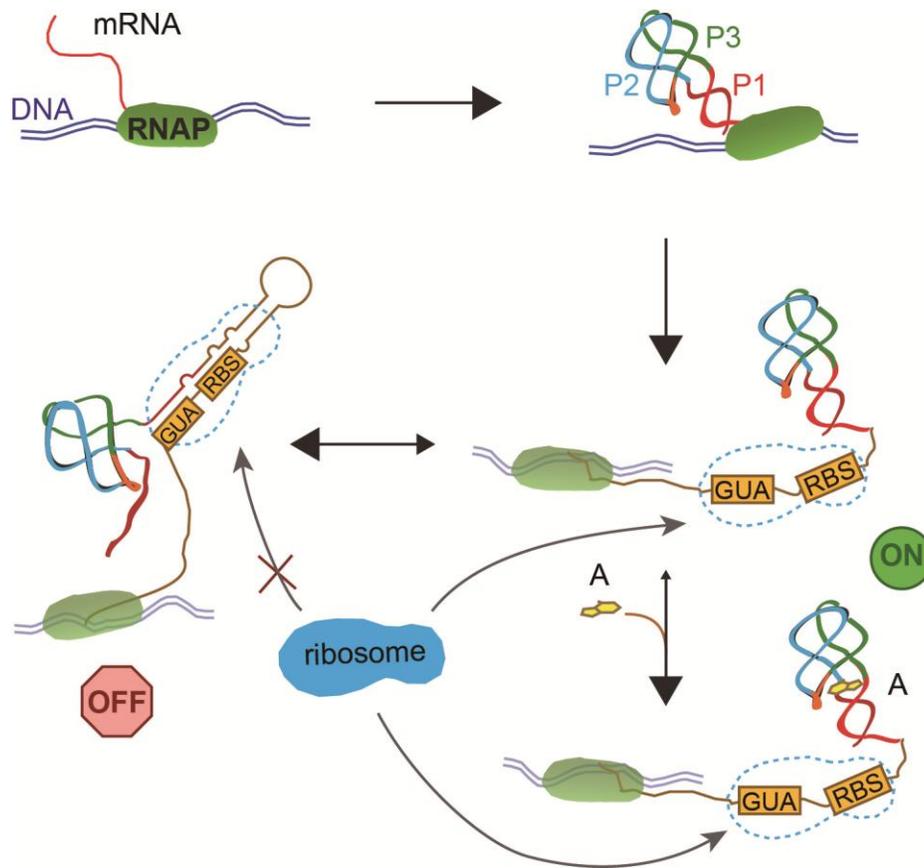
essence of the folding of this riboswitch: the RNA inter-converts from “off” to “on”, with the outcome ultimately determined by thermodynamic changes due to ligand binding.

It was previously suggested, based on the adenine-binding ability of the full-length riboswitch as well as *mfold* (Zuker, 2003) predictions of secondary structure energies for the “on” and “off” states, that the “off” state is more stable than the “on” state without adenine but still readily fluctuates into the “on” state in equilibrium (Lemay et al., 2011; Rieder et al., 2007). Our observations directly show that this is, in fact, the case. Given that the “off” state is occupied  $\sim 1.7$  times more often than the “on” state in the absence of adenine, the “off” state is more stable by  $\sim 0.3$  kcal/mol without adenine bound. This agrees within error with our estimate of the energy difference between the “on” and “off” states obtained by adding to secondary-structure energy calculations the energies we measured for tertiary interactions and ligand binding. The secondary structure alone is predicted to be more stable by 1.4 kcal/mol in the “off” state ( $-17.9$  kcal/mol) than in the “on” state ( $-16.5$  kcal/mol). In the absence of adenine, tertiary structure in the “on” state (mostly the loop-loop interaction) brings another  $-1.5$  kcal/mol of stability. However, since the loop-loop interaction can still form when P1 is prevented from folding (as seen by using the P1 blocking oligomer, Fig. 5.6c), the difference between “on” and “off” states remains unchanged. Since adenine binding provides an additional stability of  $-8$  kcal/mol, with adenine bound the “on” state is more stable by 6.6 kcal/mol, whereas without adenine the “off” state is more stable by 1.4 kcal/mol. Errors in these numbers are estimated at  $\sim 1$  kcal/mol.



**Figure 5.10 Folding kinetics and riboswitch mechanism.** (a) The aptamer refolding rate as a function of force without adenine is determined from the distribution of refolding forces when ramping the force down from denaturing values. Error bars show s.e.m. Red line: fit to Eq. 3.8. (b) The fraction of unfolding FECs in the off state as a function of refolding time at low force in the absence of adenine shows an exponential rise as the riboswitch structure equilibrates into the more stable “off” state.

In order to demonstrate that the regulatory mechanism of this riboswitch is controlled thermodynamically, we also measured the kinetics of the folding. We first determined the folding rate of the aptamer (the “on” state) at  $F = 0$  from refolding FECs measured in the absence of adenine while ramping the force from  $\sim 20$  pN (where the aptamer is completely unfolded) down to 0 pN. 1227 refolding FECs were measured, and the cumulative folding probabilities were analysed (Dudko et al., 2008) as in Fig. 5.4b. The average refolding times plotted as a function of force (Fig. 5.10a) were then extrapolated to zero force by fitting to Eq. 3.8, resulting in an unloaded folding rate of  $\sim 30$  s<sup>-1</sup>. Since transcription *in vivo* occurs at a typical rate of  $\sim 40$ – $50$  nt/s (Proshkin et al., 2010; Vogel and Jensen, 1994) and the expression platform contains 40 nt, this folding rate implies that the aptamer should fold completely, well before the repressor hairpin has even been partially transcribed, regardless of adenine binding. Hence the riboswitch always starts in the “on” state.



**Figure 5.11 Schematic of the riboswitch mechanism.** The aptamer folds rapidly before the expression platform is transcribed, regardless of adenine binding. If adenine binds to the aptamer, it stabilises the “on” state of the riboswitch (aptamer folded, ribosome binding site exposed). Without adenine binding, the “on” state is unstable and equilibrates into the “off” state (P1 unfolded, ribosome binding site sequestered). RBS: ribosome binding site. RNAP: RNA polymerase.

Since adenine binding locks the riboswitch into the “on” state with a large binding energy and low unfolding rate, then if the riboswitch is to act as a switch, clearly the “on” state must be able to equilibrate into the “off” state rapidly in the absence of adenine, before the ribosome binds the Shine-Dalgarno sequence. To determine this equilibration rate, 5,012 FECs were measured on 11 molecules while varying the time spent waiting at low force to permit equilibration, from 0.3–5 s. An exponential fit to the fraction of molecules found in the “off” state as a function of time (Fig. 5.10b) yielded an

equilibration time of  $\tau_{\text{eq}} = 0.2 \pm 0.1$  s. The ribosome binding time is not well known, but a lower bound (even for coupled transcription and translation) is provided by the time required to transcribe the mRNA footprint for the ribosome. Footprinting studies and crystal structures (Hüttenhofer and Noller, 1994; Yusupova et al., 2001) imply this involves another  $\sim 15$  nt downstream of the 3' end of the riboswitch expression platform, requiring  $\sim 0.3$  s at typical transcription rates (or  $1.5 \tau_{\text{eq}}$ ). Hence the “on” and “off” states should be substantially (if not completely) equilibrated by the time the ribosome binds, strongly supporting a thermodynamically-controlled regulatory mechanism.

The picture of the riboswitch mechanism suggested by our results is summarised in Fig. 5.11. During the transcription of the expression platform, the riboswitch folds into the “on” state. If adenine does not bind to the aptamer after the expression platform is fully transcribed, the “on” state is no longer thermodynamically stable and the riboswitch equilibrates rapidly into the lower-energy “off” state, repressing gene expression but still allowing the “on” state to be populated at an appreciable level. If adenine does bind to the “on” state, then it becomes greatly favoured thermodynamically and the riboswitch stays “on”.

## 6. Prion misfolding and diseases

### 6.1 The prion hypothesis

Prions are the only known infectious agent composed solely of proteins, without nucleic acid (Prusiner, 1998). The term “prion”, meaning proteinacious infectious particle, was originally coined to denote the causative agent for a group of progressive neurodegenerative disorders called transmissible spongiform encephalopathies (TSEs), also now known as the prion diseases, such as Creutzfeldt-Jakob disease (CJD) in humans, bovine spongiform encephalopathy in cattle, chronic wasting disease in deer and elk, and scrapie in sheep (Colby and Prusiner, 2011; Prusiner, 1998). The majority of human prion diseases are sporadic, caused by some undetermined origins. About 15% (Colby and Prusiner, 2011) run in families (Gerstmann-Sträussler-Scheinker syndrome (GSS), fatal familial insomnia (FFI) and familial CJD), coming from gene mutations. What makes the prionopathies special is that they are infectious. That is, they can be caused by exposure to contaminated material, such as prion-infected beef. Examples of infectious prion diseases include kuru, variant CJD and iatrogenic CJD. Although rare in humans (annual incidence of one or two per million (Zerr, 2013)), the unusual “protein-only” nature of prion diseases and their connection to food items like beef and venison have attracted intensive research and clinical interest.

According to the protein-only hypothesis (Griffith, 1967; Prusiner, 1982), the infectious misfolded isoform of PrP, which is known as PrP<sup>Sc</sup>, is able to replicate itself by recruiting the natively-folded PrP<sup>C</sup>, to form additional PrP<sup>Sc</sup>, thereby permitting transmission between individuals and species. This templated conformational change is thought to be the basic mechanism underlying the propagation of prion disease. PrP<sup>Sc</sup>

shares the same primary sequence with PrP<sup>C</sup>, but with different secondary and tertiary structures. The infectious information is stored in the conformation of the protein, rather than in its genetic material. The most recent evidence to support the protein-only hypothesis is the recent success to generate prion diseases using recombinant PrP in combination with other cofactors such as phospholipid and RNA, where the infectious components are highly controllable (Wang et al., 2010).

Beyond this general conceptual framework of the protein-only hypothesis, the microscopic details of prion propagation remain unresolved and controversial. The native structure of PrP<sup>C</sup> has been solved, but central problems that remain unsolved include: (1) what is the nature of PrP<sup>Sc</sup>; (2) what is the mechanism of the PrP<sup>C</sup>-PrP<sup>Sc</sup> conversion; (3) how does prion misfolding cause disease.

## 6.2 PrP<sup>C</sup> native folding

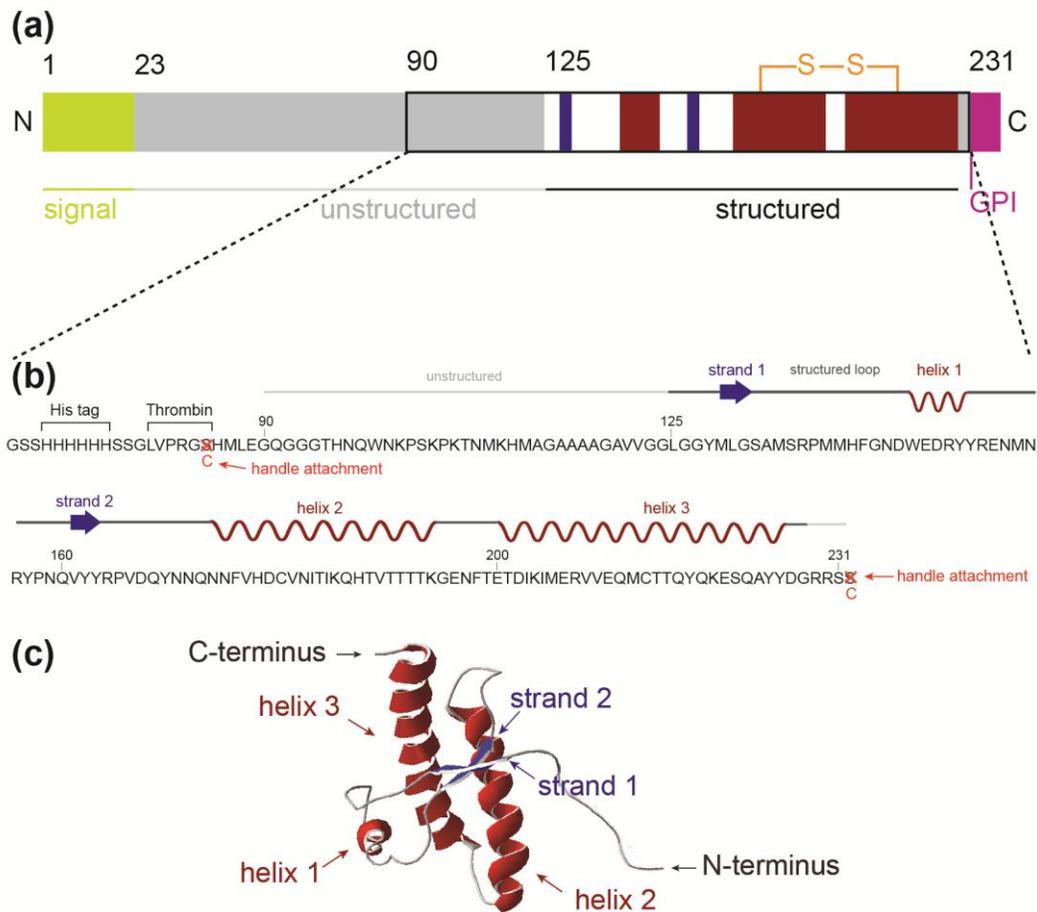
The knowledge of native folding structures, mechanisms, and pathways may provide clues for understanding the molecular origins of prion misfolding. PrP<sup>C</sup> is a 35-kDa membrane-associated protein. Its cellular function remains elusive and controversial (Westergard et al., 2007), although the high conservation of PrP<sup>C</sup> sequence and structure (Wuthrich and Riek, 2001) suggests the importance of PrP in evolution. PrP is encoded by the *Prnp* gene as a 254-residue nascent protein (Fig. 6.1a). During post-translational processing, a N-terminal 22-amino-acid signal peptide which targets PrP to the plasma membrane (Oesch et al., 1985) and the C-terminal 23-amino-acid peptide are removed. PrP can be variably glycosylated at two N-glycosylation sites and a glycosylphosphatidylinositol (GPI) anchor is attached to the C-terminus of the protein, as a

feature of cell surface proteins (Lawson et al., 2005; Rudd et al., 2002). The fully processed protein is shortened to 209 residues when bound to the plasma membrane by the GPI anchor (Stahl et al., 1987). Although these post-translational modifications have been shown to be related to pathogenic conversion of PrP (Priola and Lawson, 2001), they are not necessary prerequisites for PrP<sup>Sc</sup> formation, as suggested by the generation of infectious material using bacterially-expressed PrP<sup>C</sup> lacking glycosylation via protein misfolding cyclic amplification (PMCA) (Wang et al., 2010).

The native structure of PrP<sup>C</sup> (Wuthrich and Riek, 2001) has been solved by NMR as a monomer for a range of species: including mouse (Riek et al., 1996), Syrian hamster (James et al., 1997), bovine (Lopez Garcia et al., 2000), human (Zahn et al., 2000), other mammalian species (Christen et al., 2009; Gossert et al., 2005; Lysek et al., 2005; Perez et al., 2010; Wen et al., 2010), as well as chickens, turtles and frogs (Calzolari et al., 2005). PrP<sup>C</sup> has proven to be very hard to crystallize, but one crystal structure of PrP<sup>C</sup> has been solved, yielding a domain-swapped dimer (widely viewed as a crystal artifact) (Knaus et al., 2001). Additionally, a few structures of PrP<sup>C</sup>-antibody complexes have also been published (Antonyuk et al., 2009; Baral et al., 2012; Eghiaian et al., 2004). PrP<sup>C</sup> is composed of an unstructured N-terminal tail, comprising residues 23-125 in hamster PrP (which is the target of our work), and a globular C-terminal domain covering residues 126-228. The unstructured region contains multiple metal binding sites and can become structured upon binding Cu<sup>2+</sup> (Davies and Brown, 2008). The globular domain contains 3  $\alpha$ -helices and 2 short  $\beta$ -strands (Fig. 6.1c). There is a single disulfide bond between cysteine 179 and 214, which has been thought to be important in keeping the secondary structure of PrP intact. In particular, disulfide-free PrP was found to be incapable of folding properly in ensemble measurements (Maiti and Surewicz, 2001). However, from

our results, the disulfide bond is not, in fact, necessary for PrP<sup>C</sup> to fold into PrP<sup>C</sup> (chapter 7); likely the difference is that the single-molecule assay prevents aggregation as a competing pathway. Interestingly, work on the folding of other disulfide-containing proteins has found that native folding precedes the formation of disulfide bonds, which tends to happen at a later stage of the folding process under reduced conditions (Kosuri et al., 2012). Therefore, it is not unreasonable that PrP could fold natively without the disulfide bond if aggregation could be prevented.

While the native structure is known, the folding pathway to reach the native state remains disputed. A partially-folded on-pathway intermediate has long been proposed to mediate misfolding (Cohen et al., 1994), and partially-native structures are a feature of several models of PrP<sup>Sc</sup> (DeMarco and Daggett, 2004; Gossert et al., 2005). However, the experimental evidence for on-pathway intermediates is conflicting: some studies found only two-state folding, *e.g.* using fluorescence and CD spectroscopy (Robinson and Pinheiro, 2009), stopped-flow chemical denaturant (Wildegger et al., 1999) or fast temperature jump (Hart et al., 2009) at pH 7-8, whereas others were consistent with the presence of intermediates at both neutral and acidic pH (Apetri et al., 2006; Apetri and Surewicz, 2002; Jenkins et al., 2008; Kuwata et al., 2002; Martins et al., 2003). A more complicated pH-dependent folding pathway was also suggested in a few publications. Specifically, unfolding of PrP was observed to be two state at neutral pH, whereas as the pH is lowered, the unfolding pathway shifted to a three-state transition (Khan et al., 2010; Swietnicki et al., 1997).



**Figure 6.1 The structure and sequence of PrP.** (a) Domain structure of the full sequence of PrP. The fragment 90-231 that is protease-resistant in PrP<sup>Sc</sup> is denoted by the black box. (b) Sequence and secondary structure of SHaPrP(90–231) construct used in our study. The secondary structure and key functional areas (*e.g.* His tag, Cys residues for DNA handle attachments) are annotated on the amino acid sequence. (c) The native structure (James et al., 1997) consists of 2 short  $\beta$ -strands near the N-terminus (strands 1 and 2), 1 short  $\alpha$ -helix (helix 1), and two long  $\alpha$ -helices (helices 2 and 3) at the C-terminal end of the protein.

### 6.3 PrP misfolding and aggregation

PrP<sup>Sc</sup> shares the same primary sequence but adopts an alternative conformation as PrP<sup>C</sup>. Upon protease digestion, PrP<sup>Sc</sup> usually produces a protease-resistant 142-amino-acid core with approximately 67 amino acids from the N-terminus being cleaved, termed

as PrP 27-30 (McKinley et al., 1991). PrP 27-30 retains the infectivity and mice exclusively expressing the protease-resistant part 90-231 of PrP are still susceptible to prion disease (Fischer et al., 1996), indicating the N-terminal 23-89 segment is not necessary for prion propagation. Meanwhile, the principal structural reorganization during the conversion has been shown to involve the entire protease-resistant part of PrP molecule (Smirnovas et al., 2011). Therefore PrP 90-231 (Fig. 6.1b) is widely used to study the conformation transition between PrP<sup>C</sup> and PrP<sup>Sc</sup> (Gong et al., 2011; Grabenauer et al., 2010; Jenkins et al., 2009).

Other than the general fact that PrP<sup>Sc</sup> is a more  $\beta$ -rich isoform (Caughey and Raymond, 1991; Pan et al., 1993), much less is known about its structure and folding pathways compared with PrP<sup>C</sup>. The problem of understanding prion replication at a biophysical level can be reduced crudely to two principal questions: (1) what is the structure of PrP<sup>Sc</sup>? and (2) what is the mechanism of PrP<sup>C</sup>-PrP<sup>Sc</sup> conversion? A definitive structure for PrP<sup>Sc</sup> has proven elusive. PrP can aggregate into a wide variety of different forms, including various oligomers (Silveira et al., 2005; Simoneau et al., 2007) as well as amorphous aggregates (Pan et al., 1993) and amyloid fibrils (Prusiner et al., 1983). Soluble oligomers early in the aggregation pathway are in many cases believed to be the neurotoxic species (Silveira et al., 2005; Simoneau et al., 2007), rather than the mature fibrils (Colby and Prusiner, 2011). However controversy remains about which species are responsible for toxicity, which are infectious, and whether infectivity and toxicity are tied together.

Structural information about the PrP oligomers thought to be toxic is not available, instead various models have been proposed for the amyloid fibers into which PrP<sup>Sc</sup> aggregates. The structure of the misfolded aggregates have been characterized by

low resolution methods such as CD spectroscopy, AFM (Anderson et al., 2006), X-ray fiber diffraction (Nguyen et al., 1995; Wille et al., 2009), or electron microscopy image reconstruction (Govaerts et al., 2004). These have revealed fibrils typically tens of nanometers in diameter (Anderson et al., 2006; Nguyen et al., 1995) with extensive cross- $\beta$  structures (Govaerts et al., 2004; Wille et al., 2009). Atomic-resolution structures of a non-mammalian prion fibrillar state were also available from solid state NMR and X-ray crystallography. A left-handed  $\beta$ -helical structure has been found in amyloid fibrils of prion-forming domain of the HET-s protein (Wasmer et al., 2008), and a distinct cross- $\beta$  spiral structure has been found in amyloid fibrils formed from aggregation-prone peptides of PrP (Sawaya et al., 2007) and the yeast prion protein (Nelson et al., 2005). Based on these biophysical and biochemical evidences, a number of models have been proposed for prion fibrils (Cobb et al., 2007) as well as PrP<sup>Sc</sup> (DeMarco and Daggett, 2004; Govaerts et al., 2004). However, although fibrils are commonly found in prion diseases, amyloid formation is not an obligatory feature of the disease (Wille et al., 2000).

Uncertainty about the structure of PrP<sup>Sc</sup> has also made the question about PrP<sup>C</sup> to PrP<sup>Sc</sup> conversion difficult to address. Different models have been proposed about the mechanism of PrP<sup>C</sup>-PrP<sup>Sc</sup> conversion, including the seeded polymerization model (Come et al., 1993; Lansbury and Caughey, 1995), template-assisted replication models (Cohen et al., 1994; Cohen and Prusiner, 1998), or hybrids of these (Soto, 2005). In a seeded nucleation process, fluctuations of PrP<sup>C</sup> monomers produce a rare, non-native conformation able to slowly form an ordered, misfolded oligomer which then recruits and stabilizes additional monomers. However, in the template-assisted replication model, the conformation of PrP<sup>Sc</sup> is copied with the assistance of chaperone proteins, which haven't been identified to date (Colby and Prusiner, 2011). The clarification of different

conversion mechanisms will rely on the effective characterization of the PrP early aggregation process, *i.e.*, what happens when PrP first misfolds and aggregates.

We can re-frame the question in terms of what folding pathways are available to PrP, which of these lead to misfolding and aggregation, and how these pathways change in response to stimuli (such as protein-protein or ligand-protein interactions) that are known to affect conversion. Several studies have characterized oligomeric species involved in the early stages of prion amyloid formation at the ensemble level (Jain and Udgaonkar, 2008; Redecke et al., 2007; Sokolowski et al., 2003). However, since the initial events in conversion are likely rare, or transient, it is challenging to resolve key features of this process using traditional methods, owing to ensemble averaging effects. Conditions required for such rare or transient states to accumulate sufficiently to be detectable (*e.g.* the use of chemical denaturants to stabilize non-native states) may result in further aggregation, leading to results that are difficult to interpret. Furthermore, PrP can aggregate into a wide variety of different forms (Aguzzi et al., 2007), likely via many different pathways. The difficulties in observing intermediates and distinguishing different pathways can be solved by applying single-molecule techniques, to investigate the misfolding and conversion of PrP (Yu et al., 2013).

## **6.4 Mutations and prion diseases**

At least 30 point mutations in *Prnp* gene have been associated with prion diseases. These mutations are the origin behind the heritable forms of prion diseases such as familial CJD, GSS and FFI (Mead, 2006; van der Kamp and Daggett, 2009). In these cases, one or more mutations cause cells to favour producing PrP<sup>Sc</sup>, however the detailed

mechanisms by which the mutation cause disease are not clear. These pathogenic mutations must influence the protein structure somehow, resulting in an increased propensity to produce abnormal prions.

Various studies have probed the effects of mutations on the stability, folding, or cellular interactions of PrP (van der Kamp and Daggett, 2009). Some mutations were believed to affect stabilities of specific substructures or interactions (e.g. salt bridges, hydrogen bonds). For example, the E200K mutation, which is related to familial CJD, destabilizes the two large helices 2 and 3 (Gallo et al., 2005) and changes the surface electrostatic potential (Zhang et al., 2000) without affecting the overall stability of the protein (Swietnicki et al., 1998). D178N, a well-studied mutation involved in diseases such as familial CJD and FFI, affects important salt bridges and internal hydrogen-binding interactions inside the molecule. Meanwhile, the regions around the two native  $\beta$ -strands is also destabilized by this mutation, although the overall fold stays largely the same (Watanabe et al., 2008). Some mutations are believed to affect the overall folding of PrP<sup>C</sup> as well. The F198S mutation leaves a “hole” in the hydrophobic core of PrP, resulting in a significant decrease in thermodynamic stability (Liemann and Glockshuber, 1999) and an increased propensity for PrP<sup>Sc</sup> conversion (Vanik and Surewicz, 2002). The no-internal-Cys mutation (C179A/C214A) was originally designed to elucidate the role of the disulfide bridge in the stability and folding of PrP. Previous studies showed this mutation results in molten globule-like properties for PrP, with a partial loss of native  $\alpha$ -helical structure and increased propensity towards aggregation attributed to the absence of the disulfide bond (Maiti and Surewicz, 2001). However, our current results (chapters 7 and 8) contradict these earlier interpretations and suggest that removing the disulfide has little effect on the native folding of PrP, in the absence of aggregation.

There are several mutations that significantly destabilize the PrP<sup>C</sup> conformer in favor of an inferred folding intermediate (Apetri et al., 2004) that is postulated to be a PrP<sup>Sc</sup> precursor (Apetri et al., 2006; Apetri and Surewicz, 2002): for example F198S, Q217R, R208H, D178N, V180I, V210I. These mutations greatly increase the aggregation propensity, possibly by reducing the barrier for PrP<sup>Sc</sup> formation, which makes them possible candidates for elucidating the mechanism of prionopathies underlying familial prion diseases.

## 6.5 Single-molecule studies of PrP

Because of their distinct advantages in measuring energy landscapes and observing rare or transient species directly, single-molecule methods provide a powerful tool for probing the complex folding pathways found in prion misfolding, especially for characterizing early stages of PrP misfolding and aggregation process. In the single-molecule regime, individual PrP monomers, oligomers, or fibrils can be distinguished from one another as structural sub-populations and characterized independently. A variety of probes is available to monitor the structure, stability, and dynamics of the protein, based most commonly on fluorescence or force measurements, as introduced in chapter 3. FRET is a very sensitive probe of protein conformation and dynamics (Greenleaf et al., 2007), as shown by an elegant FRET study of the prion-determining NM domain of the yeast prion protein Sup35. These studies found that Sup35 takes on an ensemble of rapidly-interconverting structures (Mukhopadhyay et al., 2007). To our knowledge, however, single-molecule FRET has not yet been applied to PrP. Fluorescence probes can also monitor the formation and growth of oligomeric aggregates,

for example by measuring the size-dependent diffusion of oligomers via fluorescence correlation spectroscopy (FCS) or the inter-molecular FRET signals from differently-labeled proteins as they aggregate. FCS has been used to follow oligomerization in conjunction with ensemble methods, observing a multi-step process starting with  $\beta$ -structured dimers (Post et al., 1998). It has also been used to study the interaction of PrP with antibodies, with applications aimed at ultra-sensitive prion detection (Bieschke et al., 2000; Birkmann et al., 2006; Fujii et al., 2007).

Force spectroscopy, with its particular advantages for studying misfolding and aggregation, has started to be applied in prion research. The first force spectroscopy study of PrP used AFM to pull apart amyloid fibrils made from recombinant human PrP(90-231) (Ganchev et al., 2008). Fibrils were adsorbed to a mica surface, and the gold-coated AFM tip was bonded covalently to the thiol group of a cysteine engineered near the N terminus of the protein, thereby pulling on one of the PrP monomers comprising the fibril. Increasing the force on this monomer, part of the protein was found to stretch like an unstructured WLC polymer before the monomer was ultimately plucked out of the fibril core at ~100-200 pN of applied force. By measuring how the force for removing the monomer changed with the pulling speed of the AFM (Evans and Ritchie, 1997), the unfolding of the PrP monomer from the amyloid was shown to share certain characteristics typical of  $\beta$ -structured proteins, as opposed to the helical structure of PrP<sup>C</sup>. More interestingly, from the length of the unstructured WLC the authors were able to postulate the location of the edge of the  $\beta$ -sheet core of the amyloid. Their results contradicted the predictions of two popular structural models of PrP amyloid (DeMarco and Daggett, 2004; Gossert et al., 2005) which retain substantial native-like helical structure near the C terminus, but was consistent with a different model in which the

native C terminus is completely restructured (Cobb et al., 2007; Smirnovas et al., 2011). Most recently, nanopore force spectroscopy was used to explore the structural dynamics of the wild-type and a pathogenic mutant of PrP (Jetha et al., 2013). The current fluctuations were monitored as a single PrP molecule was captured electrophoretically through an organic  $\alpha$ -hemolysin nanometer-scale pore, allowing discrimination between the wild-type and the D178N mutant of PrP<sup>C</sup>. To describe the kinetics of the molecule inside the pore, a four-state model was proposed.

The most in-depth force-spectroscopy studies of the dynamics of PrP folding and misfolding to date, however, are those using optical tweezers that I describe in this thesis. The folding of isolated PrP monomers has been extensively characterized, revealing a two-state native folding pathway and three distinct misfolding pathways. By joining two PrP molecules together, the ability of interacting PrP molecules to form non-native structures was also studied. We found that the folding pathway of PrP in the context of individual dimers was dramatically altered from the normal native folding. The energy landscapes of natively-folded monomers and misfolded dimers were measured, revealing detailed mechanisms for the folding and misfolding of PrP. These results demonstrate the potential of single-molecule analysis to probe the detailed kinetics and conformational dynamics of misfolding and aggregation of PrP, as well as other disease-related proteins.

## 7. Direct observation of multiple misfolding pathways of PrP monomers<sup>\*</sup>

In this chapter, we applied SMFS to observe directly the misfolding of PrP monomers. By measuring folding trajectories of single PrP molecules held under tension in a high-resolution optical trap, we found that the PrP native folding pathway involves only two states, without evidence for partially-folded intermediates which have been proposed to mediate misfolding. Instead, frequent but fleeting transitions were observed into off-pathway intermediates. Three different misfolding pathways were detected, all starting from the unfolded state. Remarkably, the misfolding rate was even higher than the rate for native folding. A mutant PrP with higher aggregation propensity showed increased occupancy of some of the misfolded states, suggesting these states may act as intermediates during aggregation (Yu et al., 2012b).

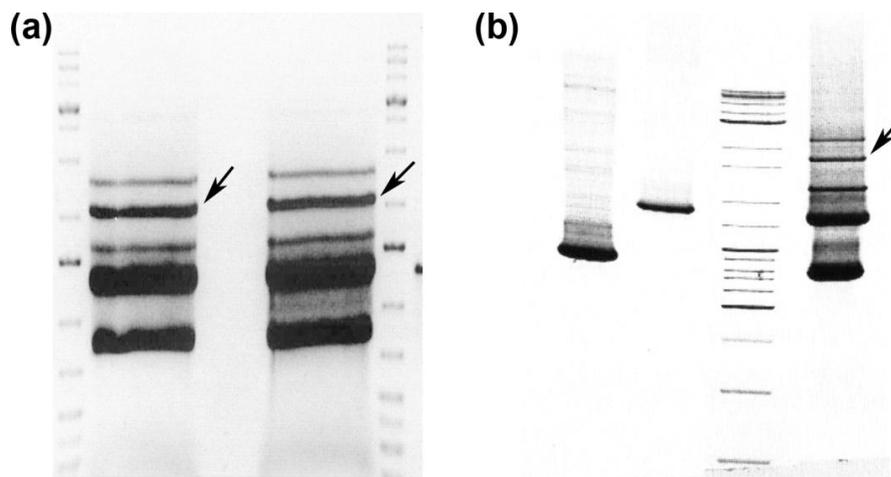
### 7.1 Sample preparation

In order to observe misfolding directly in single PrP molecules, we applied our standard SMFS assay (Fig. 4.5) by attaching kilobase-long DNA handles to the protease-resistant fragment (residues 90–231) of Syrian hamster PrP (SHaPrP), and then bound the handles specifically to polystyrene beads held in optical tweezers. Proteins were engineered with cysteine residues at each terminus and expressed in *E. coli*. After expression, cell pellets were resuspended in buffer A (6M GnHCl, 10mM Tris, 100mM

---

<sup>\*</sup> A version of this chapter has been published. Yu, H., Liu, X., Neupane, K., Gupta, A. N., Brigley, A. M., Solanki, A., Sosova, I. and Woodside, M. T. Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc Natl Acad Sci USA* **109**, 5283-5288 (2012). In this work, H.Y. built the instrument, performed experiments and analyzed data. Handle attachment protocols were mostly developed by X.L.. Proteins were provided by A.M.B., A.S. and I.S..

NaH<sub>2</sub>PO<sub>4</sub>, 10mM reduced glutathione, pH 8.0) and sonicated for 2 minutes. The lysate was centrifuged, filtered, and then purified by FPLC (GE Healthcare) using a Ni-NTA column. After rinsing the protein bound to the column with buffer A, PrP was refolded on the column using gradient exchange from buffer A to buffer B (10mM Tris, 100 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0). Impurities were removed with by rinsing on column with buffer C (10 mM Tris, 100 mM NaH<sub>2</sub>PO<sub>4</sub>, 50 mM imidazole, pH 8.0). The refolded PrP was then eluted with buffer D (10 mM Tris, 100 mM NaH<sub>2</sub>PO<sub>4</sub>, 500mM imidazole, pH 5.8), before dialysis into 50 mM sodium phosphate buffer, pH 7.0. Product purity and identity were assessed by SDS PAGE, Western blotting (Anti-prion(109-112) clone 3F4, Millipore), and folding into the native structure verified by CD spectroscopy. PrP expression and purification were done by Angela Brigley.



**Figure 7.1 Gel electrophoresis of the handle attachment reaction. (a)** 1% agarose gel of the handle attachment reaction. Four lanes represent: ladder, product, product, ladder. **(b)** 4-20% gradient SDS-PAGE gel of the protein-DNA attachment reaction. Four lanes represent: 798-bp handles, 1261-bp handles, the ladder and the final product. The five bands in the product lanes represent, from top to bottom, constructs containing two 1261-bp handles, 1261-bp and 798-bp handles, two 798-bp handles, 1261-bp handle alone, and 798-bp handle alone. The arrows indicate constructs containing one of each handle. This band contains a mixture of constructs with and without the protein bound between the handles; only a small portion consists of the protein-DNA chimera.

DNA handles were attached to the protein similarly to a previous protocol (Cecconi et al., 2008): the correctly-folded protein (typically  $\sim 20\mu\text{M}$ ) was first reduced with TCEP in a 100:1 molar ratio with PrP for 30 min at room temperature. It was then desalted by spin columns (Zeba, Thermo Scientific) to remove excess TCEP. The reduced protein was immediately activated with 2,2'-dithiodipyridine (Sigma-Aldrich) in a 100:1 molar ratio with PrP and incubated overnight at room temperature. The CD spectrum of the activated protein confirmed that it remained in the native fold. DTDP was removed by spin columns from the activated protein before reacting with sulfhydryl-labeled DNA handles of two different lengths prepared by PCR. One handle was 798-bp long and labeled at opposite termini by sulfhydryl and biotin, the other 1261-bp long and terminally labeled with sulfhydryl and digoxigenin. Two different lengths of the handles have been chosen for easy gel identification. The handles, typically  $\sim 20\mu\text{M}$  in concentration, were also reduced with 100-fold TCEP for 30 min at room temperature and then desalted by spin columns, after which activated proteins were added immediately and incubated at room temperature overnight. Reference constructs consisting only of the DNA handles without protein can be made by creating a disulfide bond between handles. The formation of the correct construct length was verified by gel electrophoresis (Fig. 7.1).

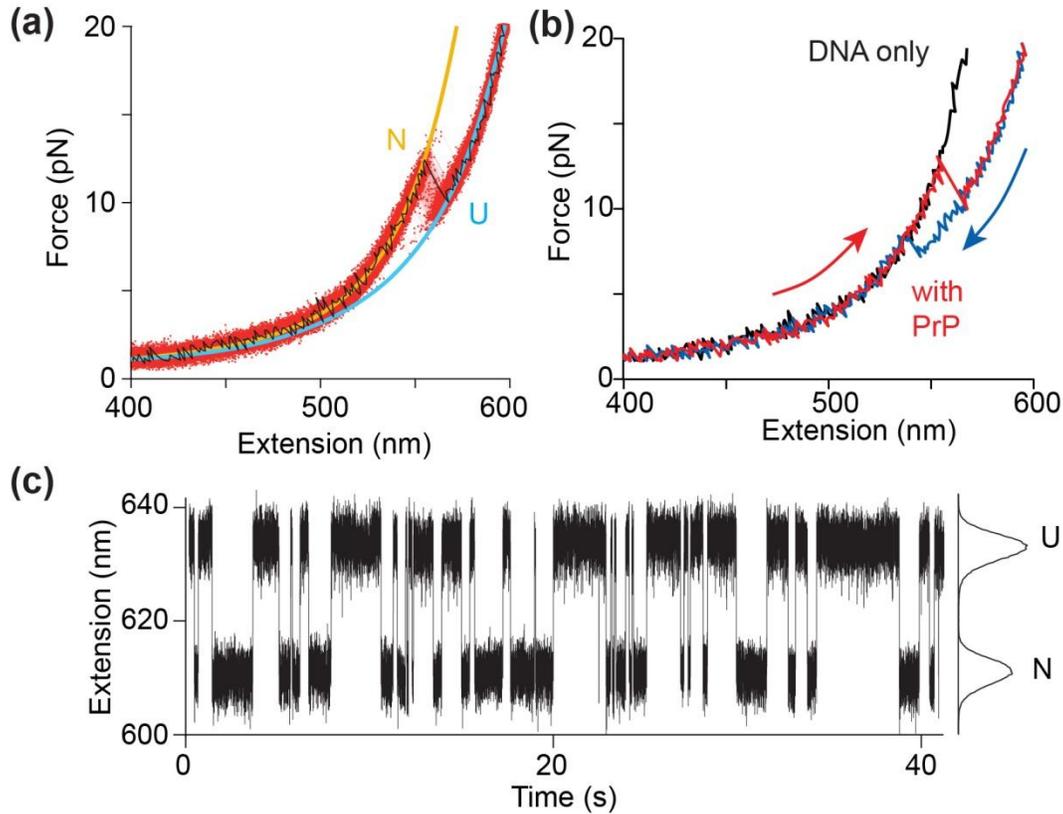
Protein-DNA constructs were incubated at  $\sim 100\text{ pM}$  with 250 pM polystyrene beads (600 nm diameter labelled with avidin, 800 nm diameter labelled with anti-digoxigenin), to form dumbbells (Greenleaf et al., 2008). Dumbbells were diluted to  $\sim 500\text{ fM}$  in 50 mM MOPS, pH 7.0, with 200 mM KCl and oxygen scavenging system (8 mU/ $\mu\text{L}$  glucose oxidase, 20 mU/ $\mu\text{L}$  catalase, 0.01% w/v D-glucose), before insertion into a sample cell for the optical tweezers. Using the tweezers to apply denaturing tension to a

single PrP molecule, we monitored the resulting dynamic structural changes in the protein by measuring the end-to-end extension of the molecule with high spatial and temporal resolution.

## 7.2 PrP native folding is two-state

FEC measurements were first performed by moving the traps apart at a constant rate to ramp up the force. The extension increased monotonically with force as the handles were stretched (Fig. 7.2a) until the protein unfolded at  $\sim 10$  pN, causing an abrupt increase in extension and concomitant drop in force indicative of an apparently two-state process (Borgia et al., 2008). Refolding curves, where the force was ramped down, also showed two-state behavior (Fig. 7.2b). 3250 FECs measured on 7 molecules displayed the same two-state behavior, without any distinguishable subpopulations (*e.g.* different contour length changes or unfolding forces). The contour length change,  $\Delta L_c$ , determined from WLC fits (Eq. 4.3) to the FECs (Fig. 7.2a), was  $\Delta L_c = 34.1 \pm 0.4$  nm (see appendix for a detailed protocol of WLC fitting). The number of amino acids (aa) unfolded,  $n_{aa}$ , was calculated from the  $\Delta L_c$ - $n_{aa}$  relation:  $n_{aa} = (\Delta L_c + \Delta d_T)/L_c^{aa}$ . Here we explicitly included the term  $\Delta d_T$  to take into account the fact that as the structure changes, the distance between the two termini of the structured parts of the protein (the points at which force is instantaneously being applied) changes, too. In the case of PrP,  $d_T = 0$  for the unfolded state (since there is no structured protein), whereas for the native state,  $d_T = 3.1$  nm is determined from the NMR structure of the protein (James et al., 1997). The crystallographic contour length of an amino acid,  $L_c^{aa} = 0.36$  nm/aa (Pauling and Corey, 1951), was used to convert  $\Delta L_c$  into the number of amino acids unfolded, rather than the slightly larger value reported from high-force AFM pulling experiments (Ainavarapu et

al., 2007). We found  $n_{aa} = 103 \pm 1$ , consistent with complete unfolding and refolding of the 104-aa native structure.

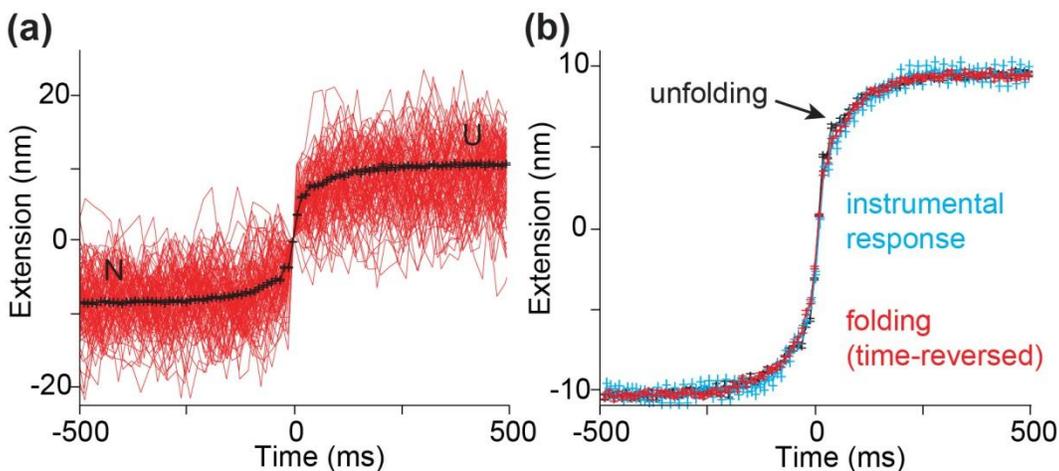


**Figure 7.2 Force spectroscopy measurements of PrP show apparent two-state folding.** (a) FECs of PrP. The handles stretch as the force rises monotonically until PrP unfolds at  $\sim 10$  pN, causing a discrete jump in the extension and force (black). By overlaying 100 unfolding FECs (red), the contour length change is found from worm-like chain fits to the folded (orange) and unfolded (cyan) states to be the value expected for unfolding of the native state. (b) The unfolding FEC of PrP (red) and refolding FEC (blue) both contain a single sawtooth feature demonstrating two-state behavior. Since the two curves overlap before and after the folding transition, the contour length change is the same in each case. For comparison, a FEC for the reference construct lacking protein (black) demonstrates only the wormlike chain elasticity of the DNA handles, without any unfolding transition. (c) With the force held constant at 9.1 pN (near  $F_{1/2}$  for PrP native folding) by a passive force clamp, the extension jumps between two values corresponding to the unfolded state (“U”) at higher extension and natively-folded state (“N”) at lower extension. The value of the molecular extension at a given force appears to vary from one molecule to the next due to variations in the size of the beads.

FECs probe the folding out of equilibrium, due to the changing force. To investigate the folding under equilibrium conditions, we also measured the extension of the molecule as a function of time while the force was held constant using a passive force clamp (Greenleaf et al., 2005). The extension jumped between two values corresponding to natively-folded (“N”) and unfolded (“U”) PrP as in the FECs, spending very little time between (Fig. 7.2c). No intermediates are immediately obvious in these data, despite the proven effectiveness of constant-force measurements for detecting them (Cecconi et al., 2005; Gebhardt et al., 2010; Woodside et al., 2006a). While the observation of two-state folding agrees with some previous ensemble measurements (Wildegger et al., 1999), others have inferred the existence of an intermediate (Apetri et al., 2006; Hart et al., 2009; Jenkins et al., 2008; Kuwata et al., 2002). To search for direct evidence of an intermediate that might be too rare or short-lived to be readily apparent at first glance, we examined extensive constant-force records (several hours in total) measured at high bandwidth. A passive force clamp was essential here, to avoid artifactual transients from feedback loop closure. Because these measurements were made under equilibrium conditions, the protein must have sampled all possible transitions between different conformations. As we demonstrated in chapter 4 and chapter 5, constant force measurement has a higher sensitivity towards observation of rare or short-lived states.

An intermediate state I on the pathway from U to N might be obligate (*e.g.* if there is only one pathway from U to N) or non-obligate (*e.g.* if there is more than one pathway from U to N), as illustrated in Fig. 4.8. Regardless of the nature of the intermediate, in measurements where the system is in thermodynamic equilibrium (such as constant-force records of long duration), all possible pathways between accessible states must be sampled. Hence any on-pathway state I must in general be observed in

three ways: (i) as a “step” between U and N as the molecule folds or unfolds; (ii) as a “spike” up from N, when N unfolds transiently into I and then refolds; and (iii) as a “spike” down from U, when U folds transiently into I which then unfolds again.

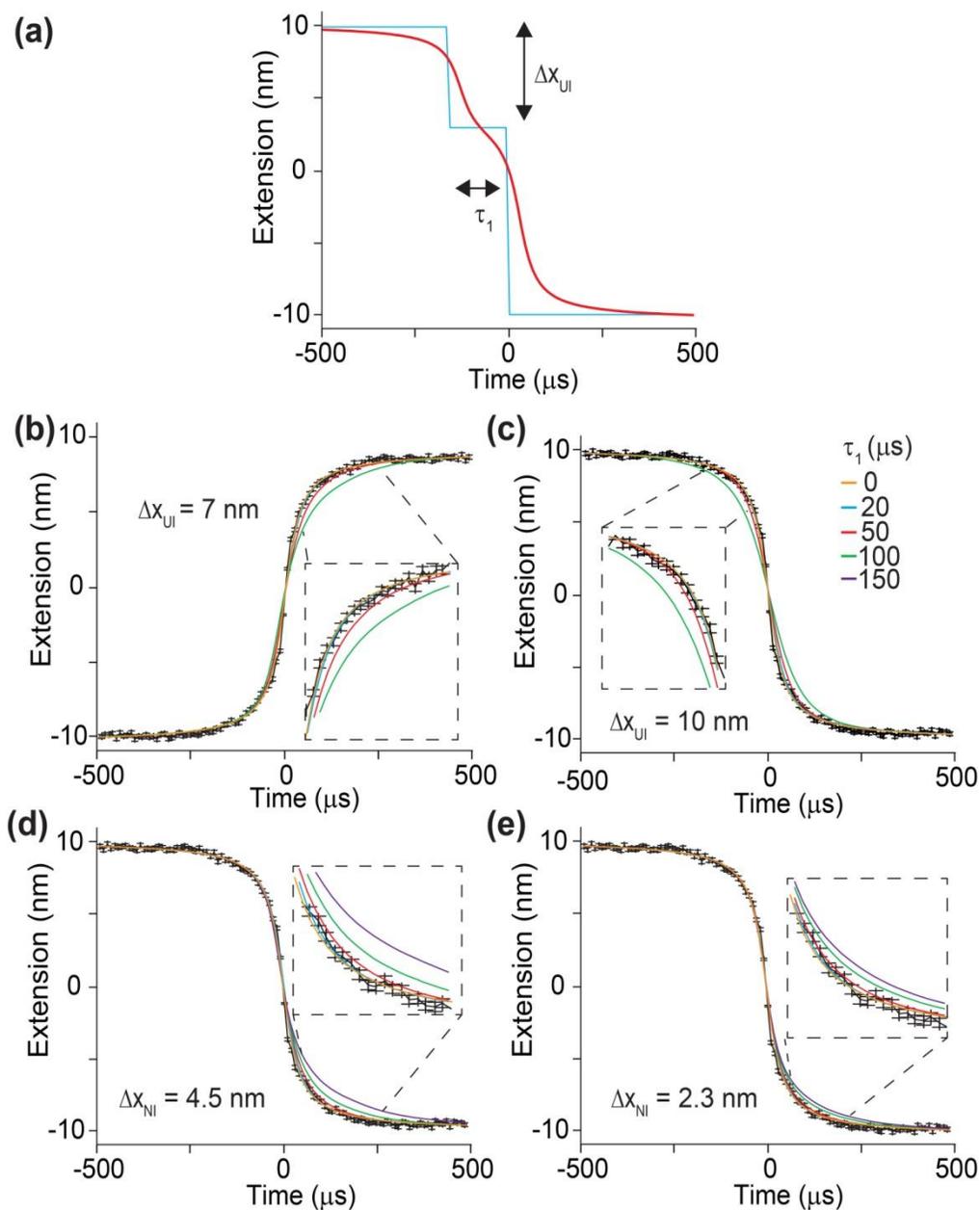


**Figure 7.3 On-pathway intermediates are not observed.** (a) To search for on-pathway intermediates, 100 separate transitions between folded (“N”) and unfolded (“U”) states taken from a single constant force record were aligned on the midpoint of the transitions (red) and averaged. The average of 3364 transitions is shown in black. (b) The average of the unfolding transitions (black) and the time-reversed average of 3318 folding transitions (red) are identical within experimental uncertainty, and the same as the average instrumental response function measured with a reference construct lacking protein (blue). An intermediate would cause the transitions to differ from the instrument response. Error bars: standard error.

We first searched for I during the brief time spent moving between N and U. We aligned all transitions on their midpoints (Fig. 7.3a, red) by fitting them to the logistic function (Fig. 4.13), and then averaged them to reduce noise: 3364 unfolding transitions were aligned and averaged (Fig. 7.3a, black), as were 3318 folding transitions (Fig. 7.3b, red). Since an intermediate state would slow down the transition from N to U (or U to N) compared to a simple two-state process, we next compared the measured transitions to the signal produced by a single step-like motion in the trap as would be expected for two-state folding, by measuring the motion of a DNA reference construct when the trap was

moved abruptly ( $< 1 \mu\text{s}$ ) by 20 nm, equal to the U–N extension change. Aligning and averaging 200 such response curves as above (Fig. 7.3b, blue), we observed that the folding and unfolding transitions (Fig. 7.3b, red and black respectively) have precisely the same shape as the instrument response to a step function, confirming that there is no observable intermediate.

To determine the measurement resolution and set an upper limit on the lifetime of any putative state I,  $\tau_I$ , we used the transfer function describing the instrument response (chapter 4.4.2) to simulate the shape of the average transition that would have resulted from an intermediate located at an extension  $\Delta x_{UI}$  from the unfolded state (Fig. 7.4a). To model the shape of the folding transition expected at constant force if an obligate intermediate I were present, we generated waveforms containing step functions from U to I and then from I to N. Step functions were used as an approximation for the fast ( $\mu\text{s}$ -scale) transition time for protein folding (Shaw et al., 2010), which is much faster than the time-response of the optical trap ( $\sim 100 \mu\text{s}$ ). The lifetime in state I was chosen randomly from an exponential distribution with mean value  $\tau$ . 200 such curves were generated, convolved with the Lorentzian transfer function to match the shape that would be expected in the measurement (illustrated in Fig. 7.4a), and then averaged as for the experimentally-measured transitions. This procedure simulated the average transition that would be measured for a given value of intermediate lifetime  $\tau$  and extension change between U and I,  $\Delta x_{UI}$ .

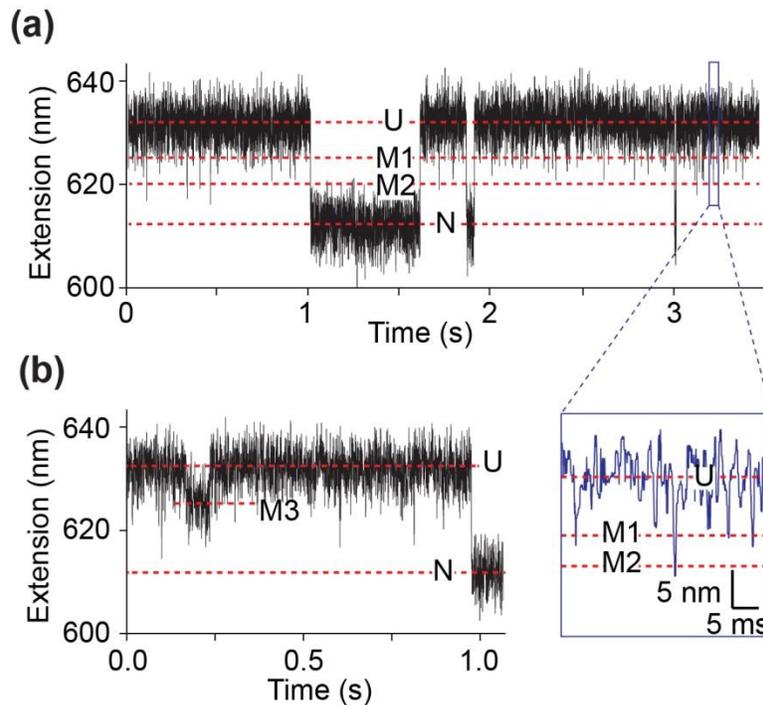


**Figure 7.4 Simulation of intermediate states.** (a) A step function from U to I and then from I to N was generated to simulate the presence of an intermediate a distance  $\Delta x_{UI}$  from U (cyan). This was convolved with the Lorentzian transfer function to add in the filtering effect of the instrument (red), and then averaged for exponentially distributed lifetimes with mean value  $\tau_1$ . (b)-(e) Simulated transitions involving obligate intermediate states with varying lifetimes  $\tau_1$ , located 7 nm from U (b), 10 nm from U (c), 4.5 nm from N (d), and 2.3 nm from N (e), respectively. Black: average measured transition; orange, cyan, red, green, purple: simulated transitions with  $\tau_1 = 0, 20, 50, 100, 150 \mu\text{s}$  respectively.

Fig. 7.4b compares simulations with  $\Delta x_{UI} = 7$  nm (the extension change from U to M1) and  $\tau = 0, 20, 50,$  and  $100 \mu\text{s}$  (respectively yellow, red, cyan, purple) to the experimental data (black). Fig. 7.4c repeats the calculation with  $\Delta x_{UI} = 10.0$  nm (the extension change from U to M2) and the same lifetimes. The experimental data are clearly inconsistent with an obligate intermediate having  $\Delta x_{UI}$  close to these values unless the intermediate lifetime is less than  $\sim 50\text{--}100 \mu\text{s}$ . Fig. 7.4d and e repeat the same calculations, this time simulating what would be expected for an on-pathway intermediate close to the native state, such as would occur if  $\beta$ -strand 1 were to detach (Hosszu et al., 2005). If all the N-terminal residues up to E146 were detached in state I (*i.e.*, everything up to helix 1), then the expected extension change between N and I,  $\Delta x_{NI}$ , would be  $\Delta x_{NI} = 4.5$  nm; if only the residues up to R136 detached, then it would be  $\Delta x_{NI} = 2.3$  nm. The area around position 139 is known from NMR experiments on bovine PrP to be relatively stable to chemical denaturation (Julien et al., 2009). Simulations with  $\tau = 0, 20, 50, 100,$  and  $150 \mu\text{s}$  (respectively yellow, red, cyan, purple, and green) in Fig. 7.4d and e show that the experimental data are inconsistent with such an intermediate unless its lifetime is less than  $\sim 100 \mu\text{s}$ .

The failure to observe any steps between U and N, whether directly in the records or indirectly via the average of the protein folding/unfolding transitions (as described above), indicates that an obligate on-pathway intermediate does not exist on the native folding pathway of PrP, unless its lifetime is less than the time resolution of the measurements (estimated from the simulations above at  $\sim 50\text{--}100 \mu\text{s}$ ). Even an intermediate located as little as 2 nm from U or N is ruled out. A rare, non-obligate intermediate is more difficult to rule out by observing the transitions between N and U: its effects might be lost when averaging the transitions because they are overwhelmed by

the transitions containing no intermediate. However, this would only be true if the intermediate lifetime is short compared to the averaging window needed to reduce the Brownian noise significantly. For example, states with lifetimes  $\sim 0.5$  ms or greater can be readily observed directly from extension records filtered at 0.5 ms. Still, no transitions from N to any state other than U could be discerned directly from the constant-force records (Fig. 7.5).



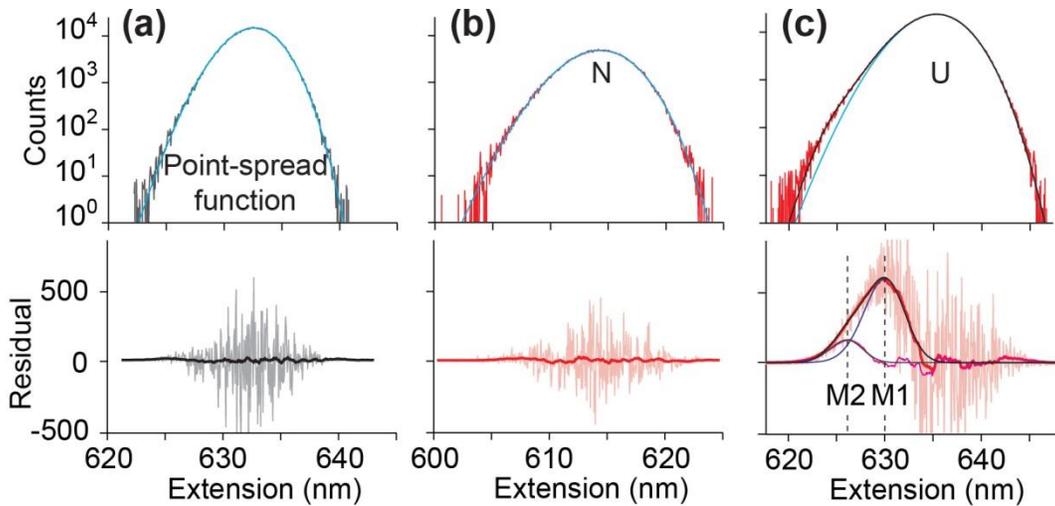
**Figure 7.5 Constant-force records reveal misfolded states.** (a) High-resolution extension records at a constant force of 9.5 pN reveal frequent transient “spikes” down from state U, reaching two different extension values which represent off-pathway intermediates (labeled “M1”, “M2”). Inset: magnified view of data in blue square. (b) Transitions from U to a longer-lived off-pathway state (“M3”) are observed very rarely. Force = 9.2 pN.

To place an upper bound on the occurrence of any rare, non-obligate intermediate, we analyzed the behavior in state N. First, we examined the constant-force records for “spikes” indicating transient unfolding from N to I, especially for evidence of a state I

that is well separated from N in distance and has a lifetime of  $\sim 0.5$  ms or longer.

Considering transient unfolding from N to I with  $\Delta x_{NI} \sim 10$  nm or more (*i.e.* a large extension change), we saw no evidence for such transitions at the level of 1 event per 300 s (upper bound based on the typical minimum record length), leading to an estimate of  $\sim 0.0002\%$  for the maximum occupancy of such a state.

Because this analysis would not reveal states with very short lifetimes, especially if they involve only a small distance change, we next examined the histograms of the extension values recorded when the protein was nominally in state N. Any partially-unfolded intermediate with a different extension than N would show up as an extra peak in the histogram, due to transitions from N to I and back. To determine the maximum possible occupancy of such a state consistent with our data, we measured the PSF of the trap, using the reference construct lacking protein (chapter 4.4.2, Fig. 7.6a). We then binned the extensions in the low-extension (nominally folded) state of the PrP trajectories and fit the resulting histogram to the PSF (Fig. 7.6b), looking for peaks corresponding to states with different extension. After fitting the extension histograms in state N to the PSF (as in Fig. 7.6), no peaks were observed in the residuals with  $\Delta x_{NI} < 10$  nm having an area greater than  $\sim 0.01\%$  of the total histogram area. The small peaks that did remain in the residual were almost certainly noise, as their location changed from one histogram to the next (Fig. 7.7a). Averaging over all the residuals to look for any peak occurring consistently at any given extension, we obtain the estimate of  $\sim 0.001\%$  for the maximum occupancy of a non-obligate intermediate. Since tension increases the lifetime and occupancy of extended states exponentially (Bustamante et al., 2004), any on-pathway intermediate would have to be exceedingly rare or short-lived at zero force.

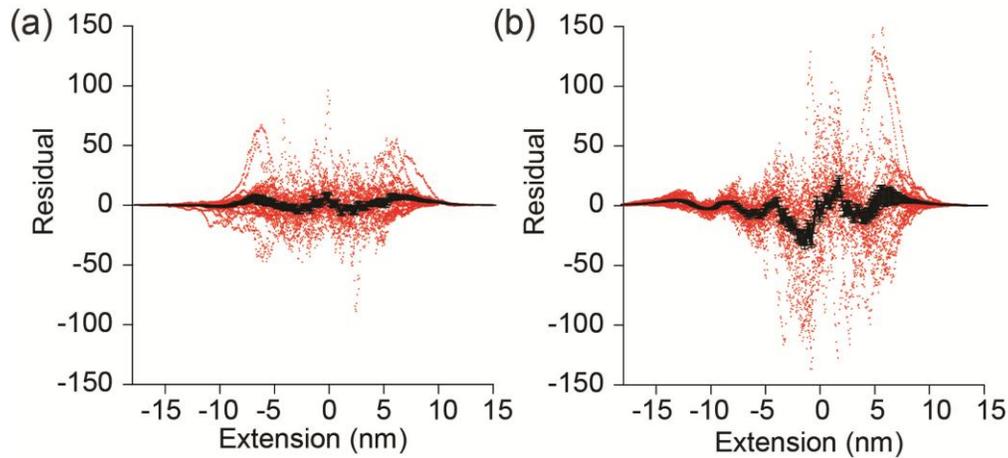


**Figure 7.6 Misfolded states form via the unfolded state only.** (a) A histogram of the extension from a constant-force record of the reference construct lacking protein (upper graph, grey) shows the point-spread function (PSF) of the trap fit to Eq. 4.17 for the PSF (cyan). Counting noise in the residual (bottom graph, grey) from the histogram binning is smoothed in a 2.5-nm window (black). (b) A histogram of the native-state extension at constant force (upper graph, red) fits well to the measured PSF of the trap (cyan). The residual (lower graph, pink) is smoothed as above (red) and is featureless, showing that N does not fold/unfold into anything other than U. (c) Fitting the unfolded-state extension histogram (upper graph, red) to the PSF (cyan) leaves a significant residual (lower graph, pink). Fitting the peak in the smoothed residual (lower graph, red) to the PSF (blue) again leaves a significant residual (purple), which is well fit by a third PSF (grey); the three PSFs together fit the full histogram completely (black). The residual fits yield the extension and occupancy of the misfolded states M1 and M2, which are entered only from U.

### 7.3 Identification of off-pathway misfolded states

Whereas no intermediates were observed on the native-pathway, the search for rare and short-lived events did reveal several states with extensions between U and N that were only entered and exited via U, and hence must be off-pathway. They were generally very short-lived ( $\sim 1$  ms or less), appearing as “spikes” pointing down from U in the constant-force extension records (Fig. 7.5a, inset). These spikes are not simply part of the extension distribution for U: in contrast to the native state distribution (Fig. 7.6b), a

histogram of the data in the high-extension (nominally-unfolded) state of the constant-force trajectory (Fig. 7.6c) reveals a small residual peak between N and U after subtracting the PSF, indicating the presence of extra states. Two additional states are needed to fit the residual: the first, “M1”, is located  $7.1 \pm 0.4$  nm from U and is occupied  $0.6 \pm 0.1\%$  of the time; the second, “M2”, is  $10.5 \pm 0.5$  nm from U and occupied  $0.11 \pm 0.02\%$  of the time. Very infrequently ( $\sim 6 \text{ hr}^{-1}$ , occupancy  $\sim 0.04\%$ ), an additional state (denoted “M3” in Fig. 7.5b) was observed  $4.9 \pm 0.2$  nm from U, with a lifetime  $\sim 50\text{--}100$  ms.



**Figure 7.7 Residuals for extension histogram fits.** (a) The residuals from fitting the low-extension peak of the extension histograms (“N” in Fig. 7.2c) to the PSF of the trap, as in Fig. 7.6b, are shown for 22 constant force records (red) containing over 29 million data points in total. All residuals were aligned on the peak of the original histogram. The amplitudes and locations of any peaks in the residuals fluctuate randomly. The average of all aligned residuals (black) is featureless, indicating that there is no other state present. (b) The residuals from fitting the high-extension peak of the extension histograms (as in Fig. 7.6c) to PSFs for states U, M1, and M2 are shown for 22 constant force records (red) containing over 86 million data points in total. Again, the average (black) is flat, indicating the excellence of the fits.

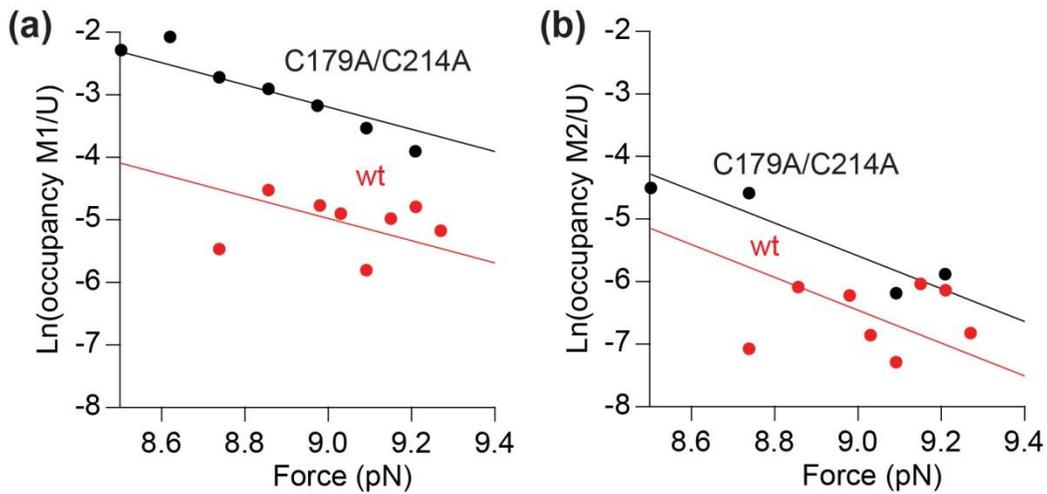
The free energy of the misfolded states can be determined directly from the ratio of occupancies of the states, using the Boltzmann probability (Bustamante et al., 2004):

$$\ln(P_M/P_U) = -k_B T (\Delta G_{fold} + \Delta G_{stretch} - F \cdot \Delta x_{MU}), \quad (7.1)$$

where  $P_M$  is the probability of finding state M,  $P_U$  is the probability of finding state U,  $k_B$  is the Boltzmann constant,  $\Delta G_{fold}$  is the folding equilibrium free energy of M relative to U,  $\Delta G_{stretch}$  is the free energy required to stretch out the unfolded amino acids under tension,  $F$  is the applied tension, and  $\Delta x_{MU}$  is the extension change between M and U.  $P_M/P_U$  was obtained from the extension histograms and plotted as a function of force for each misfolded state (Fig. 7.8). The plot was then fitted to Eq. 7.1, treating  $\Delta x_{MU}$  as a fixed parameter whose value was determined from the extension histograms. The stretching energy, estimated from the WLC model (Eq. 4.15), was then subtracted from the zero-force intercept of the fit to obtain the folding energy of the misfolded state at zero force. Finally, we find that M1 has ~ 40-50 aa folded and is  $3 \pm 1$  kcal/mol more stable than U at zero force, M2 has 60-70 aa folded and is  $5 \pm 2$  kcal/mol more stable than U, and M3 has 30-40 aa folded and is approximately as stable as U. Because these states do not lead to the natively-folded structure, they must represent misfolded conformations of PrP.

Kinetics of folding and misfolding can be extracted from our measurements as well. The average folding rate from U to N can be measured directly from the extension records. The average folding rate from U to M3 can also be estimated directly from the extension records. However, transitions into states M1 and M2 were difficult to identify directly from the extension records, as they were both very short-lived and had low occupancy. We therefore estimated the folding rates for M1 and M2 by dividing the total time spent in each state (determined from the area of the fits to the extension histograms as in Fig. 7.6c) by the lifetime of the states estimated from the extension records (as in Fig. 7.5a, inset) to determine the number of transitions into M1 or M2, and then dividing

the total time spent in the unfolded state by the number of transitions. The average lifetimes were estimated as being  $\sim 0.3\text{--}0.5$  ms for both states. Given the average occupancies for M1 and M2 (0.6% and 0.1%, respectively), this resulted in rates of  $\sim 15$   $\text{s}^{-1}$  and  $\sim 3$   $\text{s}^{-1}$ , respectively.

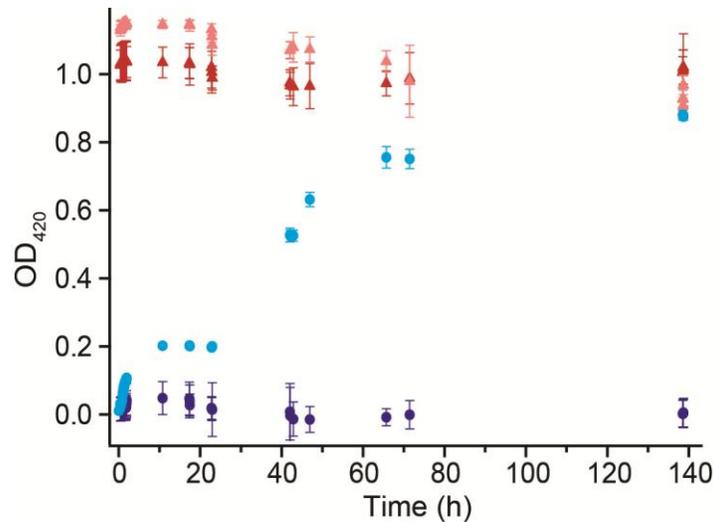


**Figure 7.8 Free energy of the misfolded states.** The misfolded:unfolded occupancy ratio for states M1 (a) and M2 (b) as a function of force were used to determine the misfolding free energies from fits to Eq. 7.1, for both wild-type PrP (red) and the C179A/C214A mutant (black). The slope of the fit, which is set by the extension change during misfolding, was treated as a fixed parameter determined from the extension histograms.

## 7.4 Mutation studies

To investigate whether these non-native states are related in some way to aggregation, we measured the folding of a mutant SHaPrP, C179A/C214A. This mutant aggregates much more readily than wild-type PrP (Fig. 7.9) and can form  $\beta$ -rich oligomers (Maiti and Surewicz, 2001). The aggregation of C179A/C214A PrP and wild-type PrP was measured using a turbidity assay, similar to a previously published method

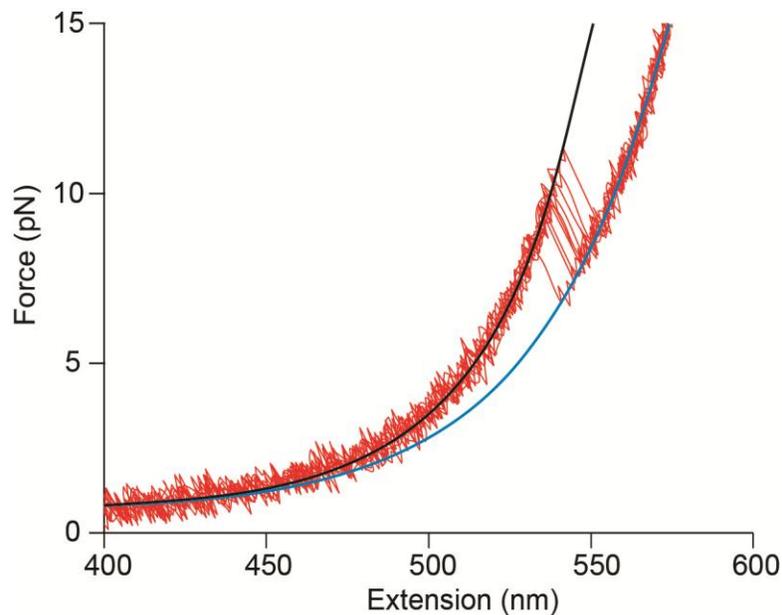
(Frankenfield et al., 2005). Briefly, 40  $\mu$ M protein in 10 mM phosphate buffer pH 7.0 was incubated for a week at 37°C while being agitated at 300 rpm, and the turbidity was measured periodically using a microplate reader via the optical density at 420 nm. The mutant aggregated during sample preparation before the first turbidity measurement was made (Fig. 7.9, red). The same turbidity was observed with and without reducing agent present (4 mM TCEP). Wild-type PrP aggregated more slowly: when reduced with TCEP, aggregation occurred over the course of 1 day, but without TCEP the protein was stable for more than one week. (Fig. 7.9, blue).



**Figure 7.9 Aggregation of wild-type and C179A/C214A PrP.** Turbidity measurements indicate that the C179A/C214A mutant aggregates very rapidly at pH 7, during sample preparation, whether oxidized (dark red) or reduced by TCEP (light red). In contrast, wild-type PrP aggregates at pH 7 over the course of a day when reduced (light blue), but does not aggregate over at least a week when not reduced (dark blue). Error bars represent s.e.m. from different replicates.

Because the geometry of the single-molecule assay isolates a single protein molecule, however, aggregation is prevented during the single-molecule measurements, allowing us to probe how the misfolding differs in a monomer that would otherwise aggregate rapidly. Force-extension curves (Fig. 7.10) revealed two states with

$\Delta L_c = 34 \pm 1$  nm, the value expected for the native structure, indicating that the native fold forms when the mutant is prevented from aggregating. Similar to the wild-type protein, misfolded states were observed in extension records at constant force (Fig. 7.11a), at the same extensions as M1–M3 and with similar kinetic properties. These states were therefore identified as states M1–M3. Notably, however, the occupancy of M1 was increased 4–5-fold and that of M2 by 2–3-fold in the mutant as compared to wild-type PrP (Fig. 7.11b and c), indicating that these states were respectively 1 and 0.5 kcal/mol more stable in the mutant (Fig. 7.8).



**Figure 7.10 Force-extension curves of C179A/C214A mutant.** 10 unfolding FECs (red) measured on the C179A/C214A mutant show two-state unfolding. WLC fits to the folded (black) and unfolded (blue) states reveal a contour length change of  $34 \pm 1$  nm, the result expected for unfolding of the native structure (34.3 nm).

Wild-type SHaPrP has a disulfide bond between C179 and C214 (James et al., 1997). This bond is present in our protein samples after purification and refolding, as shown by a Raman peak at  $533\text{ cm}^{-1}$  which is in the  $510\text{-}540\text{ cm}^{-1}$  range specific to S-S

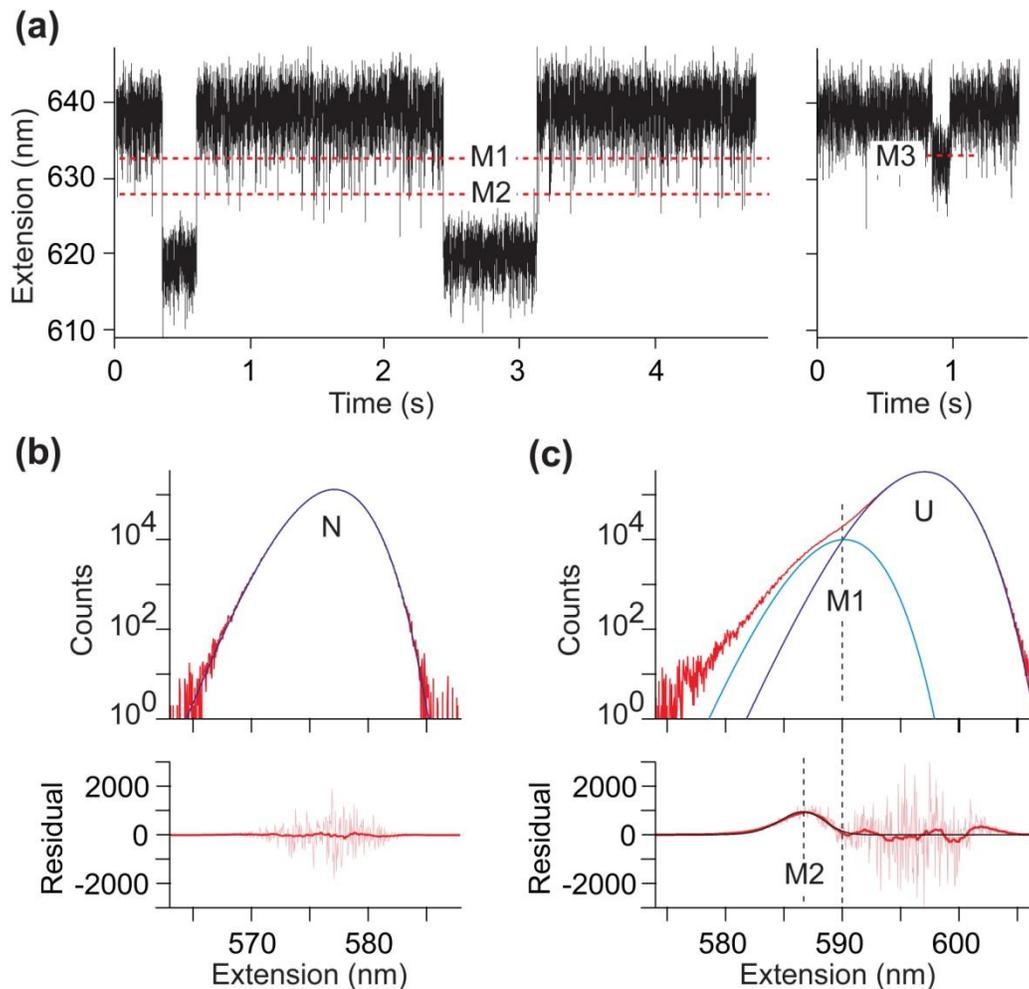
bonds (Spiro and Gaber, 1977). However, we observe no signature of the disulfide bond in the force spectroscopy measurements (Fig. 7.2): the full contour length of the structured domain unfolds in a single step at a force much lower than typically needed to break disulfide bonds, without any intermediate state at the distance expected from the NMR structure for unfolding everything up to the disulfide (22.2 nm). The C179A/C214A mutant adopted a similar native folding structure as the wild-type PrP, as suggested by a two-state folding behavior with the same contour length change from FEC measurements (Fig. 7.10). It therefore seems unlikely that the disulfide is present during the force spectroscopy measurements. It is possible the bond does not re-form properly after reduction to attach the DNA handles, or it might be broken while handling the beads at the outset of the measurement, when very large but uncalibrated forces are typically applied. The oxygen-scavenging system which protects against oxidative damage of the protein and DNA creates a non-oxidizing environment, reducing the likelihood of re-forming a disulfide bond.

## **7.5 PrP folding/misfolding pathways and implications on aggregation**

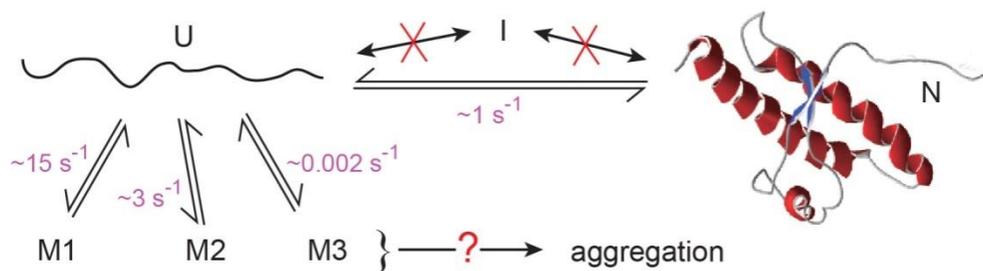
These results open a new window on the complex structural dynamics of protein misfolding, through direct observation of the conformational fluctuations of single PrP molecules. The picture of the network of folding pathways that emerges for PrP (Fig. 7.12) reveals some surprising details which shed light on PrP misfolding. First, we find that native folding is a two-state process. We see no evidence for any partially-folded intermediate on the native pathway, such as hypothesized to mediate misfolding (Cohen et al., 1994), despite previous reports (Apetri et al., 2006; Hart et al., 2009; Jenkins et al.,

2008; Kuwata et al., 2002). We cannot completely rule out on-pathway intermediates, since force probes are not sensitive to structural changes that leave the molecular extension unchanged (as in a proposed intermediate where helix 1 restructures (De Simone et al., 2007)); the chemical denaturants and low pH used in most misfolding studies might also produce intermediates not observable under the conditions here. Nevertheless, partially-native intermediates in which the structure changes near the termini, such as proposals with a detached  $\beta$ -strand 1 (Hosszu et al., 2005) or selectively disordered helices 2 and 3 (Kuwata et al., 2002), are inconsistent with our observations.

Instead of misfolding proceeding through an on-pathway intermediate, we find that non-native structures are only accessible from the unfolded state. Indeed, there is not just one misfolding pathway, but rather three of them, leading to different non-native structures. Remarkably, PrP explores these misfolding pathways more frequently than it does the native pathway: the rates for formation of M1 and M2 estimated from the constant-force trajectories,  $\sim 15 \text{ s}^{-1}$  and  $\sim 3 \text{ s}^{-1}$  respectively, are considerably higher than the rate for native folding at the same force,  $\sim 1 \text{ s}^{-1}$ . The relative rates imply that the vast majority ( $> 90\%$ ) of all attempts at structure formation lead to non-native structures rather than the native state, although the resulting misfolded states are not very stable for isolated PrP molecules under these conditions and are thus rarely occupied. One type of misfolding we did not observe, however, despite the amyloid-forming propensity of residues 106–126 (Kuwata et al., 2003), was structure within the natively-unstructured N-terminus of the protein: even transient structure formation by the N-terminus while the C-terminal domain was folded would have produced a detectable peak at short extensions in the residual to the N-state extension histogram fit (Fig. 7.6b, 7.11b).



**Figure 7.11 Mutant PrP shows increased misfolding.** (a) Constant force records of the C179A/C214A mutant show transient misfolding at the same extensions as M1, M2, and M3 for wild-type PrP, with similar lifetimes. Force = 9.2 pN (left graph), 9.7 pN (right graph). (b) The native-state histogram (upper graph, red) is well fit by the trap PSF (blue). The residual to this fit (lower graph, pink) is featureless after smoothing (red) to remove counting noise. (c) The unfolded-state histogram (upper graph, red) has a prominent shoulder. Fitting the peak and shoulder to one PSF for U (blue) and another at the location of M1 (cyan) leaves a significant peak in the smoothed residual (lower graph, red) at the location of M2, which is well fit by a third PSF (black). M1 and M2 are several times more prevalent in the mutant than the wild-type.



**Figure 7.12 Schematic diagram of folding/misfolding pathways.** SHaPrP does not pass through an intermediate on the native pathway, but the unfolded state leads to three different misfolded states at varied rates. Folding and misfolding rates at  $\sim 9$  pN are indicated. “U” = unfolded, “I” = intermediate, “N” = native, “M1”–“M3” = misfolded.

Our results suggest that the key state for misfolding may be the unfolded state (Hosszu et al., 1999), rather than the native state or an on-pathway intermediate. PrP could be unfolded *in vivo* during translocation or retro-translocation across the endoplasmic reticulum (ER) membrane (Ma et al., 2002), providing opportunities to misfold both in the ER and the cytosol. Interestingly, our measurements were made at neutral pH, similar to conditions in the cytosol, ER, and extracellular space. Previous work *in vitro* found that PrP does not readily misfold at neutral pH, but it does at low pH (Bjorndahl et al., 2011), supporting the hypothesis that PrP<sup>Sc</sup> develops in endosomes (Caughey et al., 2009). The misfolded states we found in single PrP molecules are so rarely occupied that they would be unlikely to be detected by ensemble methods, but the high misfolding rates we observe clearly indicate that PrP does indeed readily misfold at neutral pH.

The behavior of the C179A/C214A mutant provides a first look at how the observed misfolding pathways relate to aggregate formation. The  $\beta$ -rich oligomers that this mutant can form are similar to isoforms which have been investigated as possible intermediates for PrP<sup>Sc</sup> conversion (Bjorndahl et al., 2011). The fact that the misfolded

states M1 and M2 are stabilized in the mutant suggests that they could act as intermediates leading to oligomerization, with the mutation driving increased aggregation via enhanced occupancy of the misfolded intermediates. The existence of different misfolding pathways might possibly relate to the ability of PrP to form different oligomeric structures (Chakroun et al., 2010; Gerber et al., 2007), but additional measurements will clearly be needed to establish such a link. Additional studies will also be needed to address the question of how the misfolding of isolated PrP molecules relates to PrP<sup>Sc</sup> formation, *e.g.* by exploring the effects of mutations enhancing pathogenicity (van der Kamp and Daggett, 2009), probing the effects of different pH conditions, and observing aggregate formation directly. If M1–M3 are involved, however, the need to completely unfold the native state could support models where the C-terminal domain is significantly restructured in PrP<sup>Sc</sup> (Cobb et al., 2007), although partially-native structures (DeMarco and Daggett, 2004; Govaerts et al., 2004) might still form (despite being unstable in monomeric PrP) if they were stabilized during refolding into an amyloid.

A key feature of this study is the ability to observe very short-lived conformational fluctuations directly within the folding trajectories of single protein molecules, enabled by high time resolution and the capacity to resolve states with extremely low occupancies. The ability to map out the network of pathways that compete with native structure formation provides a powerful platform for investigating the molecular mechanisms of protein misfolding. Applying this approach to other proteins involved in misfolding disorders should allow new insight both into what features drive the unique behavior of PrP and into what commonalities exist between misfolding mechanisms for different proteins.

## Appendix: experimental details

Many FECs (typically 100 to 1000) were recorded at each loading rate. To remove slow drift over the course of the measurements, individual FECs measured on the same molecule under the same pulling conditions were aligned vertically using the low-force ( $F \sim 2\text{--}3$  pN) part of the data, where the FECs have very low slope, and horizontally using the high-force ( $F \sim 20$  pN) part of the data, where the FECs have very high slope. Typical drift correction was less than a few nm. All analysis was performed on aligned FECs.

DNA elasticity can be well described by the extensible WLC model (Eq 3.7) in the low force regime. The aligned FECs were separated into distinct branches corresponding to the different states. All the data from a given branch was then fitted using two WLC models in series, one for the handle and one for the protein:

$$F(x) = \frac{k_B T}{L_p^{DNA}} \left[ \frac{1}{4} \left( 1 - \frac{x}{L_c^{DNA}} + \frac{F}{K^{DNA}} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c^{DNA}} - \frac{F}{K^{DNA}} \right] + \frac{k_B T}{L_p^{protein}} \left[ \frac{1}{4} \left( 1 - \frac{x}{L_c^{protein}} + \frac{F}{K^{protein}} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c^{protein}} - \frac{F}{K^{protein}} \right] \quad (7.2)$$

Here  $L_p$  is the persistence length of the polymer (protein or DNA),  $L_c$  is the contour length,  $K$  is the stretch modulus, and  $k_B$  is the Boltzmann constant.

For a two-state system such as PrP, the concatenated FEC were separated into two branches: the first branch for the DNA handles with the folded protein and the second branch for the DNA handles with the unfolded protein. The first branch was fitted first by allowing the DNA parameters ( $L_p^{DNA}$ ,  $L_c^{DNA}$ ,  $K^{DNA}$ ) to float, but fixing the

parameters for the protein ( $L_p^{protein,1} = 0.65$  nm,  $L_c^{protein} = 15.12$  nm,  $K^{protein} = 2000$  pN), given that the natively-folded PrP contains 42 amino acids unstructured. A force offset was induced to compensate any offsets from calibration (since the force must be 0 at 0 extension), but no offset was applied in the fitting procedure to compensate for changes in bead radius. Typical fitting parameters for the DNA were  $L_p^{DNA} \sim 20 - 30$  nm,  $L_c^{DNA} \sim 550 - 600$  nm,  $K^{DNA} \sim 400 - 1500$  pN. The second branch was fitted using the same DNA parameters as found from the fits of the first branch, with  $L_c^{protein,2}$  being the only parameter allowed to vary in the fit. The contour length change for PrP unfolding was the given by  $\Delta L_c = L_c^{protein,2} - L_c^{protein,1}$ .

## 8. Energy landscape analysis of PrP native folding pathway: the diffusion constant, transition time and rates\*

Although simply mapping out the folding pathways for PrP is useful, a more detailed picture of the energy landscape would improve our understanding of the unusual behavior of PrP. In this chapter we discuss the energy landscape analysis of PrP native folding pathway from non-equilibrium force-extension measurements. After validating the Hummer-Szabo method for reconstructing landscape profiles (Gupta et al., 2011), we demonstrate the characterization of the energy landscape for native folding of PrP. The full landscape profile was reconstructed from force-extension curves, revealing a double-well potential with an extended, partially-unfolded transition state. Key parameters describing the landscape profile were recovered from different methods with good consistency, allowing the diffusion constant for barrier crossing and the transition path time across the barrier to be calculated (Neupane et al., 2012; Yu et al., 2012a). Finally, Kramers theory was used to predict the folding rates from the landscape profile,

---

\* A version of this chapter has been published in the following two papers. Gupta, A. N., Vincent, A., Neupane, K., Yu, H., Wang, F. and Woodside, M. T. Experimental validation of free energy landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat Phys* **7**, 631-634 (2011). In this work, H.Y. helped build the instrument, and performed experiments with K.N. The energy landscape reconstruction was done by A.N.G. and A.V..

Yu, H., Gupta, A. N., Liu, X., Neupane, K., Brigley, A. M., Sosova, I. and Woodside, M. T. Energy landscape analysis of the native folding pathway of the prion protein yields the diffusion constant, transition path time, and rates. *Proc Natl Acad Sci USA* **109**, 14452-14457 (2012). In this work, H.Y. helped build the instrument, performed experiments with X.L., and analyzed data with X.L. and A.N.G. The energy landscape reconstruction was mostly done by A.N.G. Proteins were provided by A.M.B. and I.S..

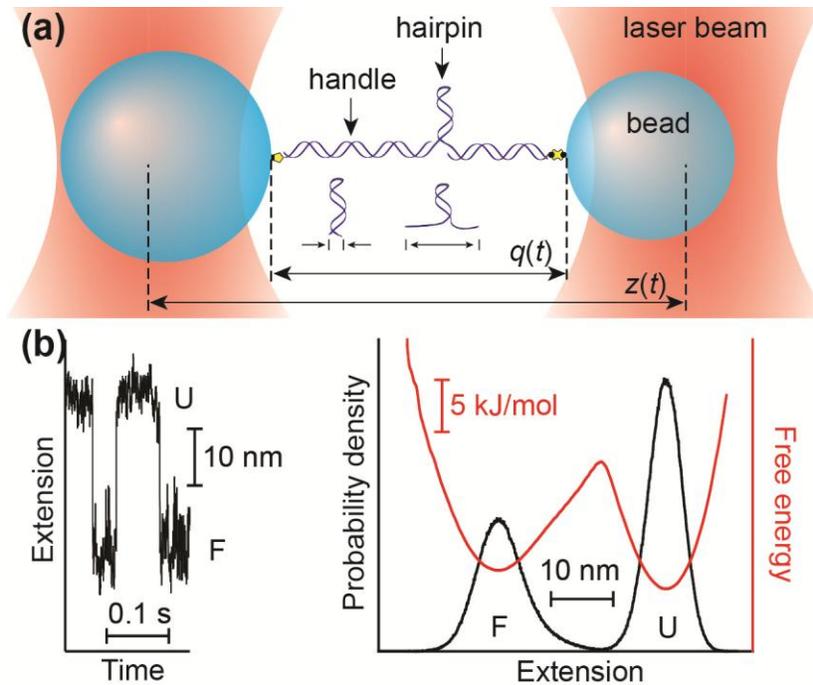
recovering the values observed experimentally both under tension and at zero force in ensemble experiments (Yu et al., 2012a).

## **8.1 Validation of the Hummer-Szabo method for energy landscape reconstruction**

As introduced in chapter 4, recent theoretical work by Hummer and Szabo (Hummer and Szabo, 2001) has shown that landscape profiles can be reconstructed from non-equilibrium single-molecule force spectroscopy measurements using an extension of the Jarzynski equality (Jarzynski, 1997). This method has been applied to simulations (Minh, 2006; Park and Schulten, 2004) and experiments (Harris et al., 2007; Liphardt et al., 2002) but never validated experimentally by comparison to known results. We tested it using force-extension measurements on DNA hairpins with distinct, sequence-dependent folding landscapes. Quantitative agreement was found between the landscape profiles obtained from the non-equilibrium reconstruction and those from equilibrium probability distributions (Woodside et al., 2006a). We also tested the method on a riboswitch aptamer with three partially-folded intermediate states, successfully reconstructing the landscape but finding some states difficult to resolve due to low occupancy or overlap of the potential wells (Gupta et al., 2011).

### **8.1.1 Two-state DNA hairpins**

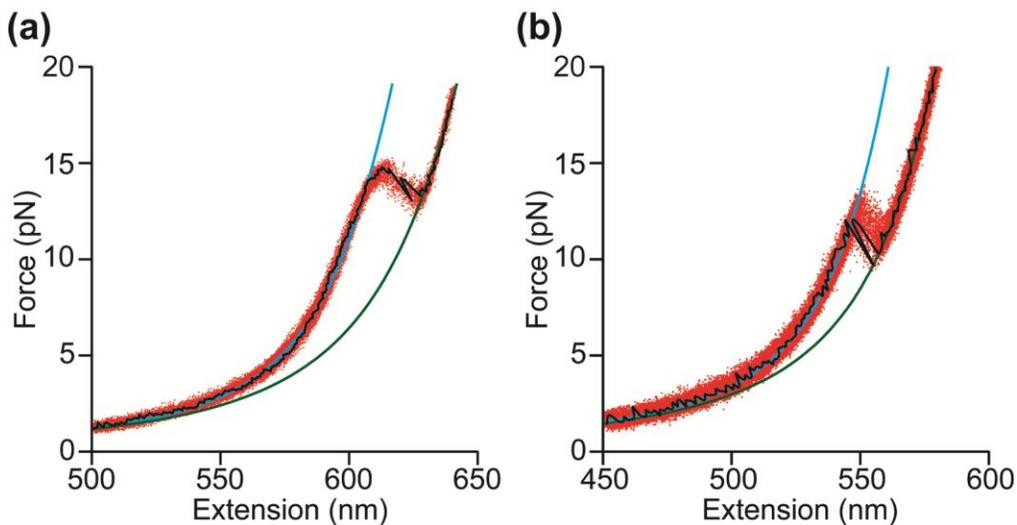
We first measured FECs for well-characterized DNA hairpins with different, sequence-dependent landscapes that were previously known from equilibrium measurements (Woodside et al., 2006a), and compared the non-equilibrium and equilibrium results.



**Figure 8.1 Optical trapping measurements of DNA hairpins.** (a) Schematic of hairpin construct in optical trap.  $z(t)$  represents the time-dependent trap separation,  $q(t)$  the molecular extension. (b) Measurements of extension at constant force can be used to generate extension probability distributions and hence the energy landscape profile. This is illustrated here for hairpin A.

Single DNA hairpins (consisting of a simple stem-loop structure) attached to kilobase-long handles of dsDNA were bound to polystyrene beads held in dual-beam optical tweezers (Fig. 8.1a). DNA hairpins attached to dsDNA handles were prepared as described previously (Woodside et al., 2006b). The resulting constructs contained a biotin-labelled 798-bp handle on the 3' end of the hairpin, a digoxigenin-labelled 1158-bp handle on the 5' end of the hairpin, and abasic sites separating the hairpin from each handle. The constructs were incubated with 600 nm and 820 nm diameter polystyrene beads labelled with avidin DN (Vector Labs) and anti-digoxigenin (Roche), respectively, to create dumbbells. Dumbbells were diluted to  $\sim 2$  pM in measuring buffer (50 mM MOPS pH 7.5, 200 mM KCl, and oxygen scavenging system: 40 U/mL glucose oxidase,

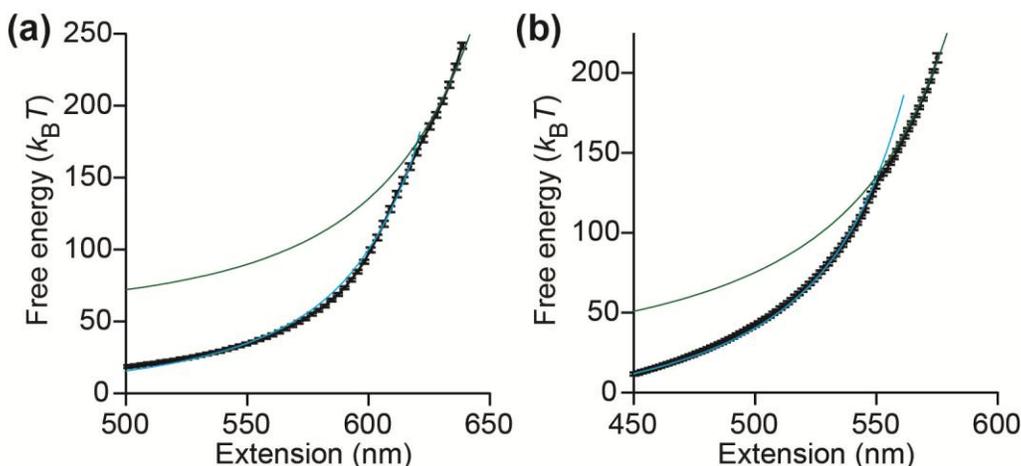
185 U/mL catalase, and 8.3 mg/mL glucose) and inserted into a sample chamber on a clean microscope slide in the optical trap. Hairpin A has the same sequence as 30R50/T4 (Woodside et al., 2006a) and hairpin B has the same sequence as 20TS06/T4 (Woodside et al., 2006a).



**Figure 8.2 FECs of DNA hairpins.** (a) FECs for hairpin A: (99 curves) with WLC fits to the folded (blue line) and unfolded (green line) states. (b) FECs for hairpin B (266 curves) with WLC fits.

Sets of FECs unfolding the hairpins were measured by moving the traps apart at a constant velocity and recording the force on the molecule as a function of extension until the hairpins unfolded. Two different hairpin stem sequences producing different energy landscape profiles were chosen to test the non-equilibrium landscape reconstruction: hairpin A, with a 30-bp stem sequence producing a barrier close to the unfolded state (Woodside et al., 2006a; Woodside et al., 2006b), and hairpin B, with a 20-bp stem sequence producing an energy barrier close to the folded state (Woodside et al., 2006a). A total of 1630 FECs was measured for hairpin A and 2918 for hairpin B, in 4 different sets for each hairpin. Within each set, FECs were aligned to remove the small

instrumental drift (typically a few nm or less) that occurred during the experiments. Individual FECs, displayed in black atop the aggregated data from a full set for each hairpin (Fig. 8.2), show the expected behavior: force increases nonlinearly with extension, due to the handle elasticity (Smith et al., 1996), until the hairpin unfolds cooperatively in a two-state manner, creating characteristic “sawtooth” patterns as the extension suddenly increases (Liphardt et al., 2001). The FECs are well fit by an extensible WLC polymer model (Wang et al., 1997) using two chains in series: one for the dsDNA of the handles and one for the ssDNA from the unfolded hairpin (Eq. 7.2).



**Figure 8.3 Free energy landscape reconstruction at zero force.** (a) Energy landscape of hairpin A reconstructed from measurements in Fig. 8.2. WLC energies for folded (blue) and unfolded (green) states extrapolated to zero extension reveal the free energy of folding. (b) Energy landscape for hairpin B reconstructed from Fig. 2b, with WLC fits. All error bars: s.e.m.

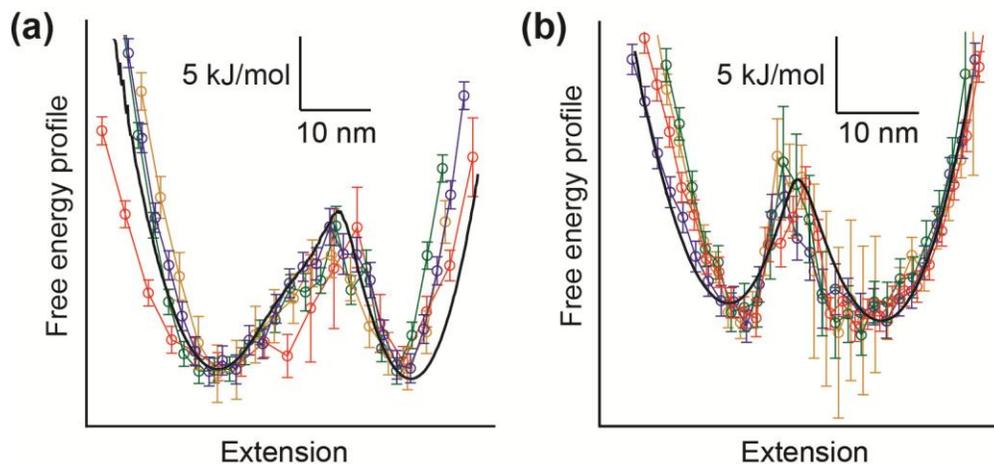
The unperturbed (zero-force) free energy profile,  $G_0(q)$ , was calculated from the FECs using the weighted histogram method (Hummer and Szabo, 2001) (Eq. 4.13) (Fig. 8.3, black).  $G_0(q)$  represents the energy profile for the entire molecule, including the handles as well as the hairpin. It is dominated by the stretching energy of the handles, as seen by fitting  $G_0(q)$  to the integral of the WLC fits from Fig. 8.2 (blue: folded state;

green: unfolded state). The kink in the profile near the crossing point of the two fits represents the unfolding of the hairpins, and hence is the region of interest. As a consistency check, we note that the energy differences between the two WLC fits at  $q = 0$ , respectively  $60 \pm 2 k_B T$  and  $36 \pm 2 k_B T$  for hairpins A and B, match the values for the unfolding free energy obtained from constant-force measurements ( $65 \pm 5 k_B T$  and  $35 \pm 2 k_B T$ , respectively (Woodside et al., 2006a; Woodside et al., 2006b)) as well as those calculated from modeling (Woodside et al., 2006a; Woodside et al., 2006b) ( $62 \pm 4 k_B T$  and  $38 \pm 3 k_B T$ ).

The profile  $G_0(q)$  is the fundamental result of the landscape reconstruction, but in order to validate the method we compared  $G_0(q)$  to the free energy profile,  $G_{\text{eqm}}(q)$ , determined from the inverse Boltzmann transform (Eq. 4.19) of extension histograms  $P(q)$  measured at constant force (Woodside et al., 2006a). The force was kept constant using a passive force clamp (Greenleaf et al., 2005) to avoid artifacts from active feedback loop closure.  $G_0(q)$  was transformed into a constant-force profile,  $G_F(q)$ , by subtracting the work done by the constant force  $F$ :  $G_F(q) = G_0(q) - Fq$ .

The results from each of the 4 sets of FECs for each hairpin are shown in Fig. 8.4 (open circles) along with  $G_{\text{eqm}}(q)$  (black lines). Error bars represent the standard error in the reconstructions, estimated by bootstrap analysis (Efron, 1979). Good agreement is found between the non-equilibrium and equilibrium reconstructions: for example, the barrier locations agree within one histogram bin, barrier heights are typically within one standard error of the equilibrium result, and the general shape of the profiles are very similar. The principal difference is in the shape of the potential wells of the folded and unfolded states: the wells in the non-equilibrium reconstruction on average have slightly

higher curvature, likely due to the fact that the effective trap stiffness used for the force-ramp measurements was higher than that for the equilibrium measurements.

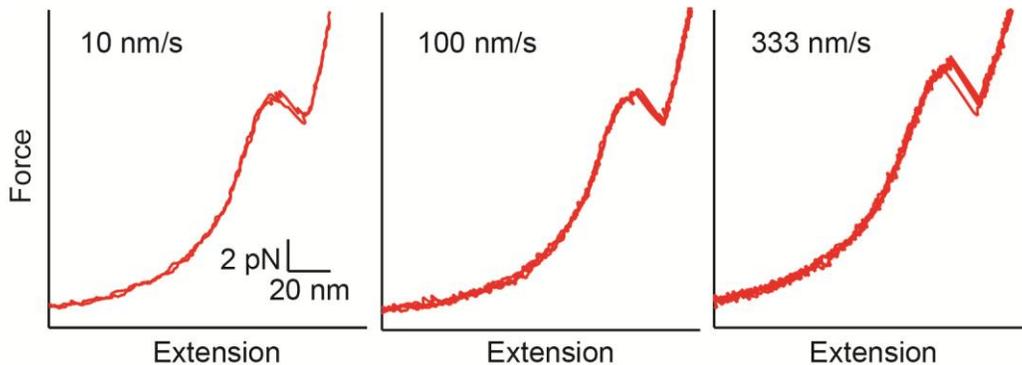


**Figure 8.4 Comparison of equilibrium and non-equilibrium reconstructions.** The energy landscapes near  $F_{1/2}$  for hairpin A (a) and hairpin B (b) reconstructed from FECs (open circles) agree with those reconstructed from constant-force measurement (black lines). The pulling rate for hairpin B was 500 nm/s (all curves); for hairpin A, it was 10 nm/s (green), 20 nm/s (yellow), 100 nm/s (blue), and 333 nm/s (red). Error bars: s.e.m. Standard error in equilibrium profiles at barrier peak: 0.3 kJ/mol (hairpin A), 0.4 kJ/mol (hairpin B).

These results validate the non-equilibrium method. We note that the four sets of data for hairpin A were measured at four different pulling rates (10, 20, 100 and 333 nm/s, with 99, 100, 138, and 1293 FECs respectively), producing dissipated energies of  $1.1 \pm 0.7$ ,  $1.4 \pm 0.7$ ,  $3.2 \pm 0.7$ , and  $4.9 \pm 0.3$   $k_B T$ , respectively. The FECs at 20 nm/s are shown in Fig. 8.2a; representative FECs at the other speeds are shown in Fig. 8.5. All FECs for hairpin B were measured at 500 nm/s, resulting in  $5.5 \pm 0.3$   $k_B T$  of dissipated energy. Little difference was observed in the reconstructions at different rates, although the errors were larger for the fastest rates, as would be expected due to the need to sample exponentially more curves when further away from equilibrium (Mossa et al., 2009). Significantly, a large number of FECs is not required to obtain a reasonable

reconstruction of the landscape profile, making this method more experimentally practical in many cases than the equilibrium method.

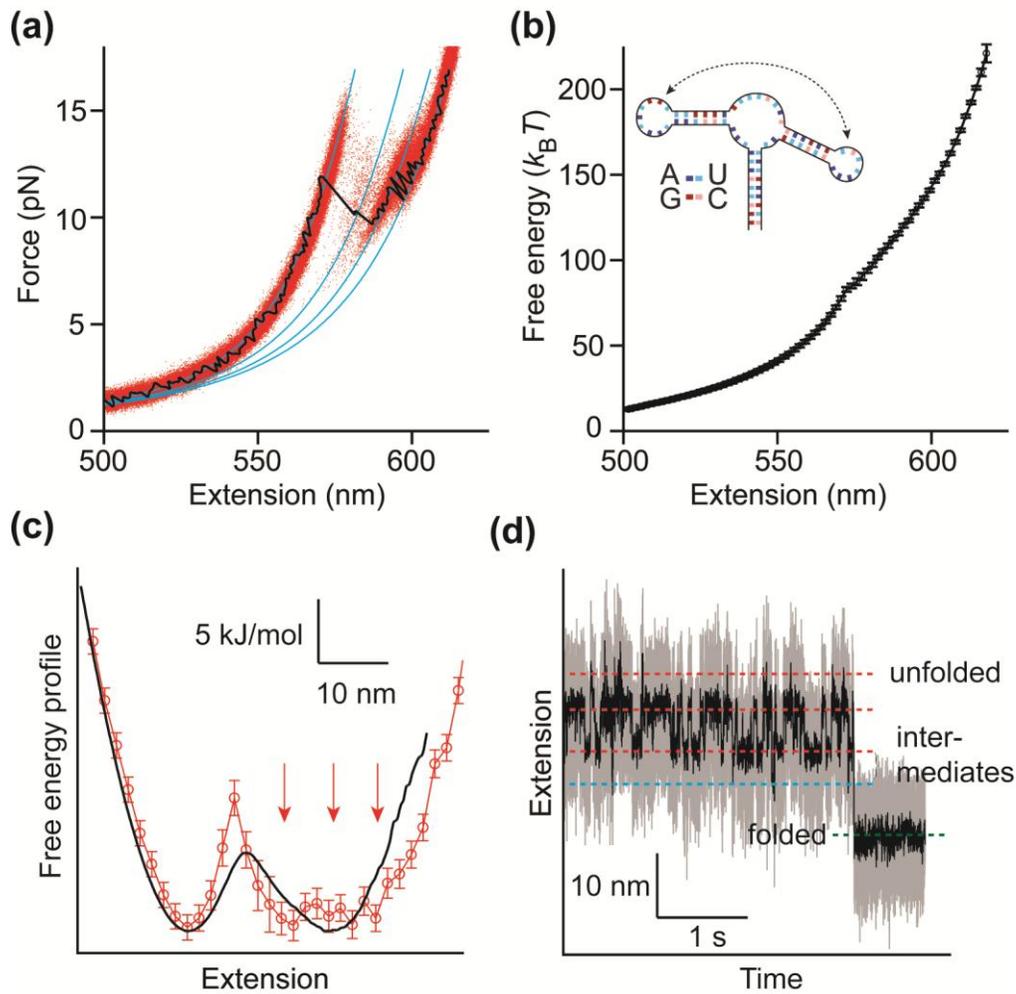
A key feature of these reconstructions is that they include the effects of the handles, which smooth out the sharp landscape features. To obtain the landscape of the hairpin alone, the handle effects would need to be removed by deconvolution (Woodside et al., 2006a). This effect is relevant to almost all SMFS data, and complicates the interpretation since the smoothing typically changes both the location and height of the energy barrier. It is especially important for analyzing data from tandem-repeat protein constructs, where the “handles” consist of unfolded protein: the handle length (and hence smoothing) changes as each subunit unfolds. In such experiments, the sequential unfolding transitions, which each involve different-length unfolded polypeptide handles, should be analyzed separately (Harris et al., 2007). The decoupling of the energy landscape of the molecule from the effect of the handle using deconvolution method will be discussed later in the chapter.



**Figure 8.5 FECs at different pulling rates.** Representative FECs of hairpin A measured at 10 nm/s, 100 nm/s, and 333 nm/s.

### 8.1.2 *add* adenine riboswitch aptamer with intermediate states

While two-state hairpins are a good model system for validating the method, many molecules unfold via intermediate states. To test whether the Hummer-Szabo method can capture multiple intermediates, we measured FECs of the *add* adenine riboswitch aptamer, which has 3 partially-folded intermediates similar to a previously-studied adenine riboswitch aptamer (Greenleaf et al., 2008). 700 FECs from a single measurement set (Fig. 8.6a) were analysed as for the hairpins to obtain  $G_0(q)$  (Fig. 8.6b). The resultant zero-force landscape was then tilted by force to compare with  $G_{\text{eqm}}(q)$  (Fig. 8.6c) as determined from constant-force data (Fig. 8.6d). Once again,  $G_0(q)$  matches  $G_{\text{eqm}}(q)$  reasonably well. Notably, while there appear to be only two potential wells, the high-extension well is much broader than for the two-state hairpins (Fig. 8.4): in fact this well arises from 3 distinct states, the unfolded state and two intermediates (indicated by red arrows), which overlap partially in extension and force. These states are visible in the filtered constant-force data (Fig. 8.6d, black, delineated by red lines), although only the two intermediates are highly populated at this force. One more intermediate can also be seen in the filtered constant-force data (Fig. 8.6d, blue line), but it is too rarely populated to appear in the equilibrium reconstruction. Since this state is not present in any of the FECs, it is also absent from the non-equilibrium reconstruction. These results demonstrate the ability to reconstruct multi-state landscapes, while also indicating some of the technical challenges involved in interpreting the reconstructions.

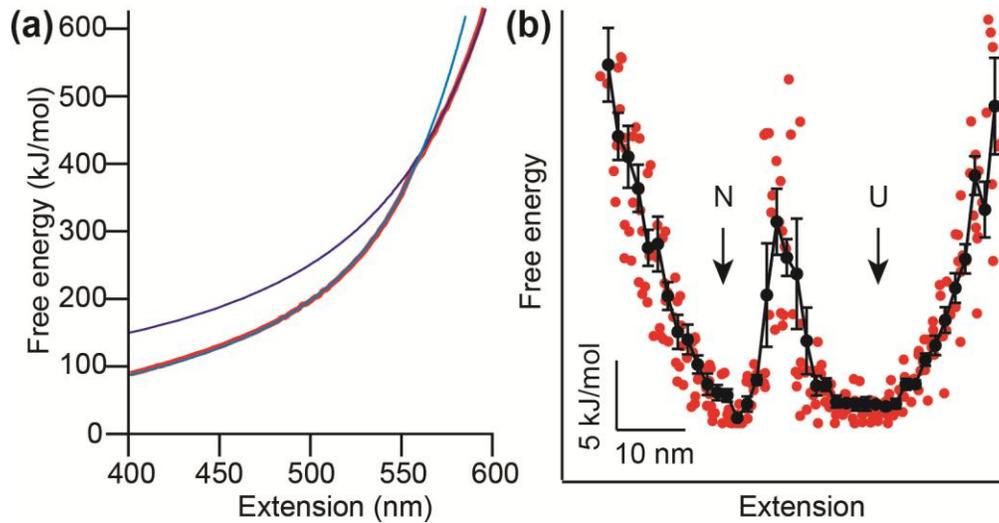


**Figure 8.6 Riboswitch aptamer landscape reconstruction.** (a) 700 FECs for *add* adenine riboswitch aptamer display multiple intermediate states (blue lines: WLC fits). (b) Free energy profile at zero force. Error bars: s.e.m. Inset: aptamer secondary structure. (c) Landscape at constant force reconstructed from FECs (open circles) agrees with that reconstructed from constant-force measurements (black line). The right-hand potential well contains contributions from three states (arrows) that are not well resolved. Error bars: s.e.m. (d) The aptamer extension at constant force (grey), median-filtered at 2.5 ms (black), shows 3 highly-occupied states and two low-occupancy states (dotted lines). Blue line: intermediate not seen in the FECs. The constant force trace was taken at 9.4 pN.

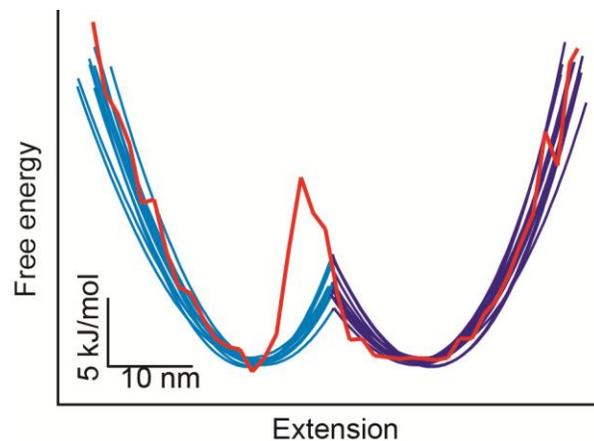
## 8.2 Energy landscape reconstruction of PrP native pathway

With the Hummer-Szabo formalism validated, we were able to reconstruct the energy landscape profile of the native folding pathway of PrP. In principle, landscape reconstructions based on high-bandwidth equilibrium measurements of the extension at constant force or trap separation typically provide higher spatial resolution (Gebhardt et al., 2010; Woodside et al., 2006a). However, equilibrium trajectories present interpretation difficulties for PrP, because in addition to native folding they also contain transitions into short-lived, non-native states accessible only from the unfolded state, which would distort the reconstructed landscape. In contrast, unfolding FECs always go first from N to U, with subsequent refolding into the misfolded states suppressed by virtue of the non-equilibrium conditions caused by the fast force-ramp rate. As a consequence, landscapes reconstructed from unfolding FECs should reflect primarily the properties of the native pathway, with minimal contribution from the non-native pathways.

The free energy profile at zero force was calculated from each set of FECs measured under the same conditions, as described previously. The resulting profiles are dominated by the stretching energy of the handles in the folded (Fig. 8.7a, cyan) and unfolded (Fig. 8.7a, blue) states, as expected. To see more clearly the potential wells for the native and unfolded states, the profiles were tilted at a force of  $F = 9.1$  pN (Fig. 8.7b, red: individual reconstructions, black: average). The barrier between the wells is indeed dominated by the protein, as seen by comparing the reconstructed landscape to the energy “profile” that consists solely of the WLC energies of the folded and unfolded states (Fig. 8.8).



**Figure 8.7 Free-energy profile reconstruction for PrP folding.** (a) The molecular free-energy profile at zero force reconstructed from a set of FECs by the Hummer-Szabo method (red) is dominated by the energies of the handles in the folded (cyan) and unfolded (blue) states. (b) Reconstructions from ten different sets (red) of FECs were tilted to 9.1 pN and averaged (black). Error bars show standard error.



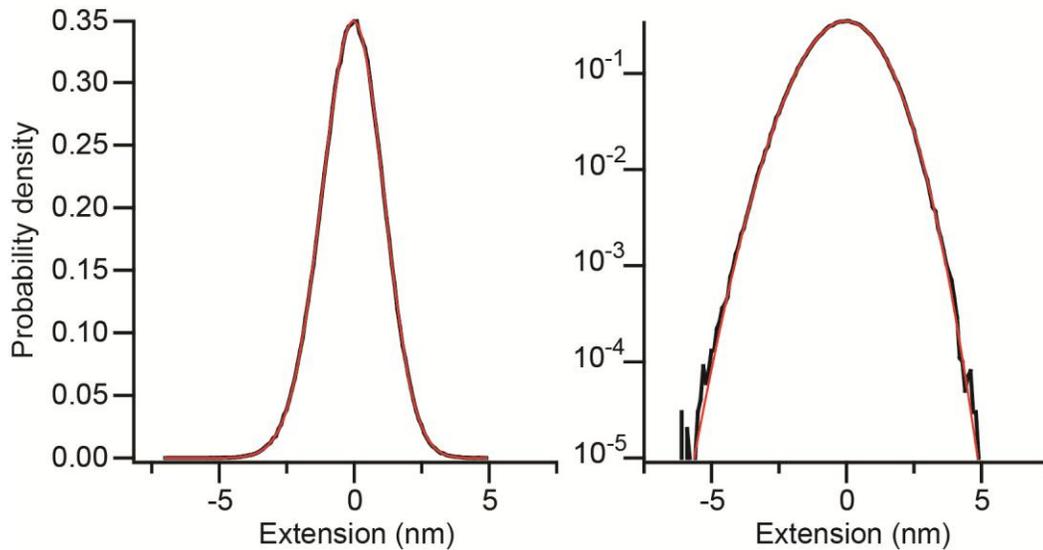
**Figure 8.8 Comparison of reconstructed landscape to WLC energies.** The outer walls of the potential wells of the reconstructed landscape (red) match the energies of the WLC fits to the folded (cyan) and unfolded (blue) states, but the barrier is quite different, indicating that the latter is dominated by the properties of the protein unfolding. The WLC stretching energies were found by integrating the FEC fits as in Fig. 8.7a, then tilting them under a force of  $F_{1/2}$  for comparison to the reconstructed landscape.

The reconstructed profile in Fig. 8.7b includes not only the intrinsic PrP free energy profile but also the effects of the compliant DNA handles and beads, which smooth the PrP landscape profile. The smoothing of the reconstructed free-energy profile due to elastic compliance was removed by pointwise non-linear iterative deconvolution as described previously (Gebhardt et al., 2010; Woodside et al., 2006a). The instrument PSF,  $S(x)$ , smoothed the true extension probability function to produce the measured extension probability,  $P(x)$ , hence the deconvolution was performed on  $P(x)$ . The free-energy profile,  $\Delta G(x)$ , was then recovered from  $P(x)$  by  $\Delta G(x) = -k_B T \ln(P(x))$ . The deconvolution was done pointwise because the PSF width varied with force, which itself varied with extension (Fig. 8.11a). An initial solution  $P^0(x)$  was first calculated from the average profile reconstruction at  $F_{1/2}$ . The true distribution function for the protein alone was then approached iteratively, with the  $(k + 1)$ th iteration at molecular extension  $a$  given by (Gebhardt et al., 2010; Woodside et al., 2006a):

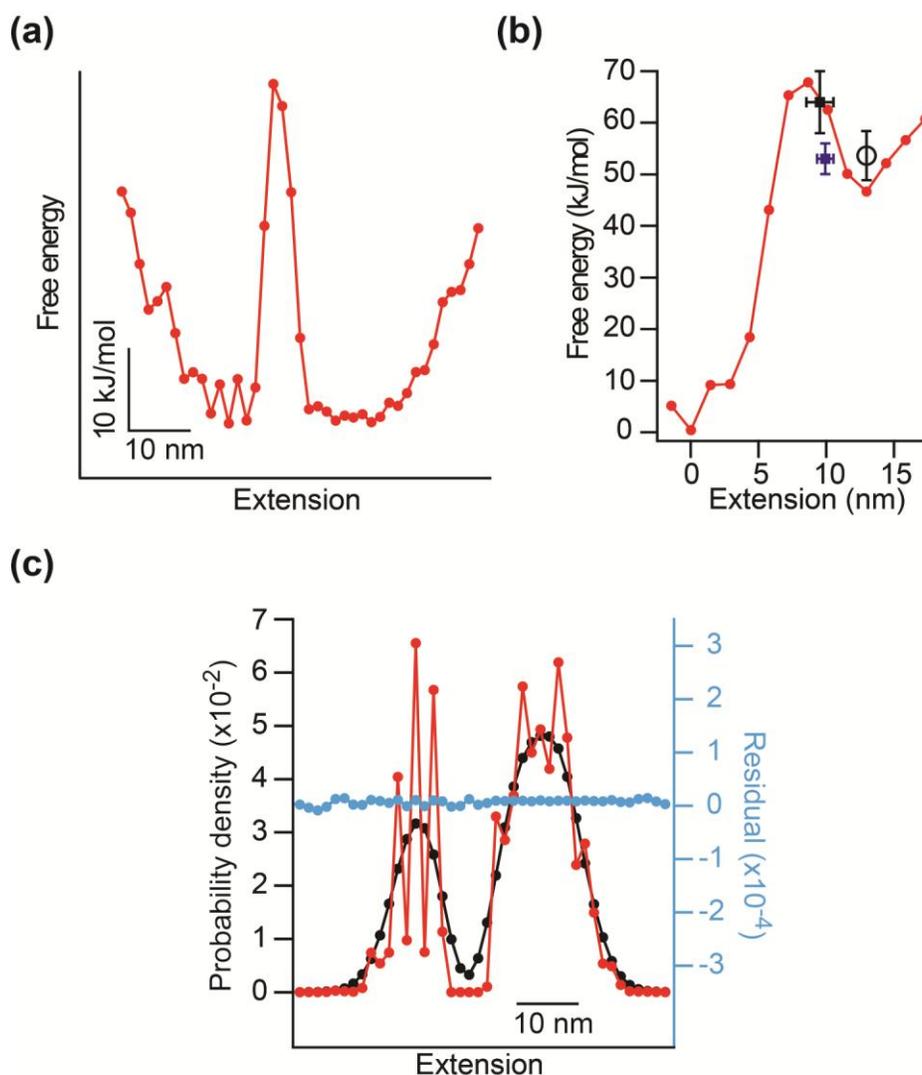
$$P^{k+1}(x) = P^k(x) + r[P^k(x)] \times \{P(x) - S^a(x) \otimes P^k(x)\}, \quad (8.1)$$

where  $S^a(x)$  is the PSF corresponding to the average force at extension  $a$ . The relaxation function  $r[P^k(x)] = r_0(1 - 2 \cdot |P^k(x) - 0.5|)$  constrained the solution to remain within the boundaries  $0 \leq P^k(x) \leq 1$  required for a physical probability function, with the amplitude  $r_0$  controlling the speed of convergence. Using  $r_0 = 2$ , the solution converged after approximately 40,000 iterations. To reduce artifactual fluctuations in the deconvolution due to measurement noise,  $P^0(x)$  was smoothed in a 3 nm window;  $S^a(x)$  was smoothed in the same way to compensate for this additional smoothing (Jansson, 1997). The centre of the low-extension probability distribution peak, giving the location of the folded state, was used as the origin for the profile at zero force. The energy required to stretch the

unfolded protein under tension was found from integrating the protein-only WLC curve and subtracted from the profile at zero force. This approach has been applied previously to landscape profiles obtained via equilibrium measurements (Gebhardt et al., 2010; Woodside et al., 2006a), but not yet using the Hummer-Szabo approach (Harris et al., 2007). The PSF of the instrument was measured from a construct consisting of beads and handles alone, without protein (Fig. 8.9), to perform an iterative nonlinear deconvolution of the landscape profile. The deconvolved landscapes at 9.1 pN and at zero force are shown in Fig. 8.10; the associated extension probability distribution and residual error are shown in Fig. 8.10c.



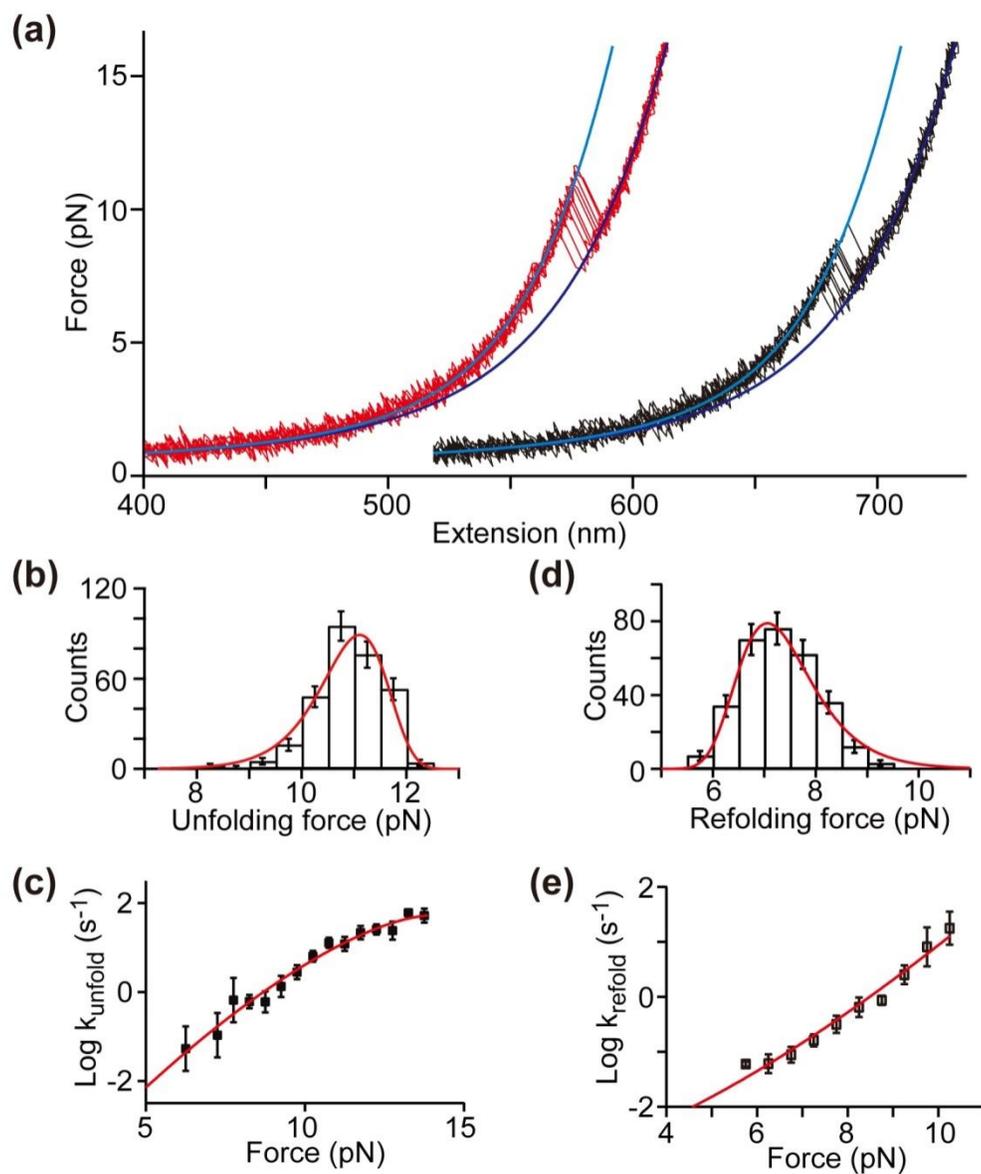
**Figure 8.9 Point-spread function for landscape reconstruction.** A typical point-spread function (at 11 pN) (black), measured under the same conditions as the FECs using the reference construct consisting of DNA handles only, shows a quasi-Gaussian profile, well-fit by Eq. 4.17 (red). The log scale graph (right) shows the slight asymmetry in the point-spread function.



**Figure 8.10 Energy landscape deconvolution.** (a) The free-energy profile at  $F_{1/2}$  after deconvolution. (b) The deconvolved free-energy profile tilted to zero force (red) agrees with the barrier height and location from the kinetic FEC analysis (black: unfolding, blue: refolding) and the free energy difference from the Crooks analysis (black circle). (c) The probability density of the extension at  $F_{1/2}$  after deconvolution (red) of the initial solution (black). There is little residual error (cyan). We note that the width of the folded state distribution is only somewhat reduced after deconvolution. This may reflect structural dynamics of the nominally-folded state, such as fluctuations arising from the relatively unstable  $\beta$ -strand 1 (Julien et al., 2009).

### **8.3 Energy landscape analysis yields the diffusion constant, transition path time and rates**

Energy landscape reconstructions like those above can recover the full profile, but they require large quantities (up to several thousands of FECs) of high quality data. As discussed earlier in chapters 2 and 4, energy landscapes can also be characterized from FEC measurements using the distribution of unfolding and refolding forces, although in a cruder, model-dependent way. We checked the consistency between the direct reconstruction and the landscape theory results. Meanwhile, kinetic properties can be predicted directly from the landscape profile using Kramers theory for diffusive barrier crossing, including the folding rates, the diffusion constant and the transition time for crossing the barrier. Kramers theory has been widely applied to interpret the timescales for protein conformational dynamics, but it has not been used to calculate protein folding rates and transition times directly from experimentally-measured free-energy profiles. Here, Kramers theory was used to predict the folding rates from the landscape profile, recovering the values observed experimentally both under tension and at zero force in ensemble experiments.



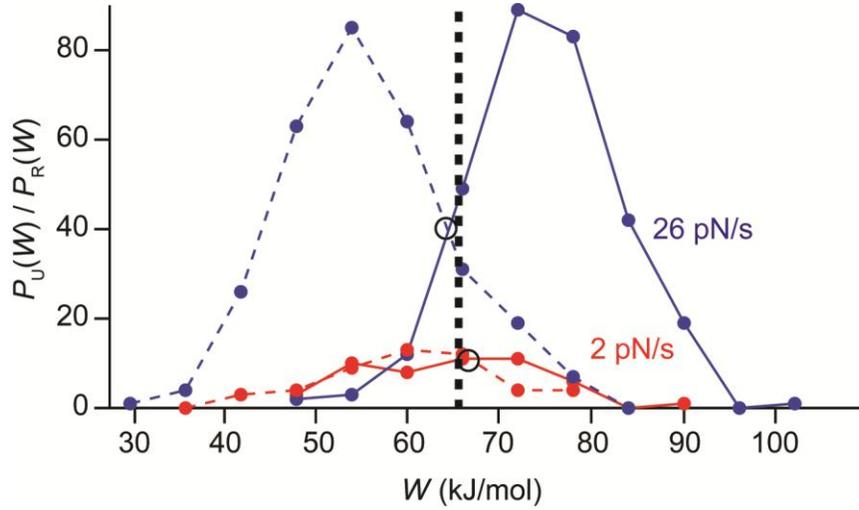
**Figure 8.11 Force-extension curves of PrP unfolding.** (a) Ramping up the force produces a FEC: the DNA handles stretch until the protein unfolds abruptly, here in a two-state process. 10 representative unfolding FECs (red) and refolding FECs (black) are shown (curves offset for clarity). Curved lines represent wormlike chain fits to the folded (cyan) and unfolded (blue) states. (b) The distribution of unfolding forces fits well to Eq. 4.4, yielding the barrier height for unfolding and location along the reaction coordinate. (c) The unfolding rate as a function of force obtained from the FECs (Dudko et al., 2008) fits well to Eq. 3.8, yielding the same parameters. (d, e) Similar fits for the refolding rate and refolding force distribution yield the barrier height and location for refolding. Fit results are given in text.

FEC measurements similar to those described above were performed: the traps were moved apart to ramp up the force until PrP unfolded, and moved close to ramp down the force to refold the protein. Similar behaviors were seen as the molecule unfolded and refolded, with the same change in contour length, as expected (Fig. 8.11a). The distribution of unfolding forces,  $p(F)$ , was fitted with Dudko model (Dudko et al., 2006), Eq. 4.4, to recover the key features of the folding landscape:  $k_{\text{unfold}}$ ,  $\Delta x_{\text{N}}^{\ddagger}$  and  $\Delta G_{\text{N}}^{\ddagger}$  (Fig. 8.11b). We also calculated the force-dependent unfolding rates (Fig. 8.11c) from the survival times during the pulling measurements (Dudko et al., 2008), and fit them to Eq. 3.8. Based on these two sets of fits, we found that  $\Delta x_{\text{N}}^{\ddagger,0} = 9 \pm 1$  nm from the native state,  $\log k_{\text{unfold}} = -6 \pm 1$  s<sup>-1</sup>, and  $\Delta G_{\text{N}}^{\ddagger,0} = 64 \pm 6$  kJ/mol from the native state. The refolding force distributions (Fig. 8.11d) and force-dependent refolding rates (Fig. 8.11e) were fit to Eq. 4.5 and Eq. 3.8, analogously to the data from the unfolding FECs, yielding  $k_{\text{fold}}$ ,  $\Delta x_{\text{U}}^{\ddagger,0}$  and  $\Delta G_{\text{U}}^{\ddagger,0}$ . The results from 1,062 refolding curves were  $\Delta x_{\text{U}}^{\ddagger,0} = 3.0 \pm 0.6$  nm,  $k_{\text{fold}} = 3.9 \pm 0.2$  s<sup>-1</sup>, and  $\Delta G_{\text{U}}^{\ddagger,0} = 5 \pm 3$  kJ/mol. Based on the unfolding and refolding FECs, from the Crooks theorem, the free energy difference between the native and the unfolded state was estimated to be  $56 \pm 5$  kJ/mol (Fig. 8.12), in good agreement with the result from the energy landscape reconstruction (Fig. 8.10b).

Using these landscape parameters, we evaluated the transition path time from Eq. 2.2, assuming that  $\kappa_b \approx \kappa_w$  (Chung et al., 2009) and making use of the fact that the Kramers prefactor can be expressed in terms of the fitting parameters as  $k_0 = k \exp(\Delta G^{\ddagger}/k_{\text{B}}T)$  (Eq. 2.1). The result for the unfolding transition was  $\tau_{\text{tp}} = 4 \times 10^{0 \pm 1}$   $\mu\text{s}$ , where the uncertainty arises mainly from the dependence of  $k_0$  on  $\Delta G^{\ddagger}$  and hence is in the exponent. For the refolding transition, the landscape parameters implied a similar value,  $\tau_{\text{tp}} = 5 \times 10^{0 \pm 0.3}$   $\mu\text{s}$ . Given that  $\tau_{\text{tp}}$  is expected to be the same regardless of the direction of

the transition (Chaudhury and Makarov, 2010), we averaged the results to obtain

$$\tau_{\text{tp}} = 5 \times 10^{0 \pm 0.3} \mu\text{s}.$$

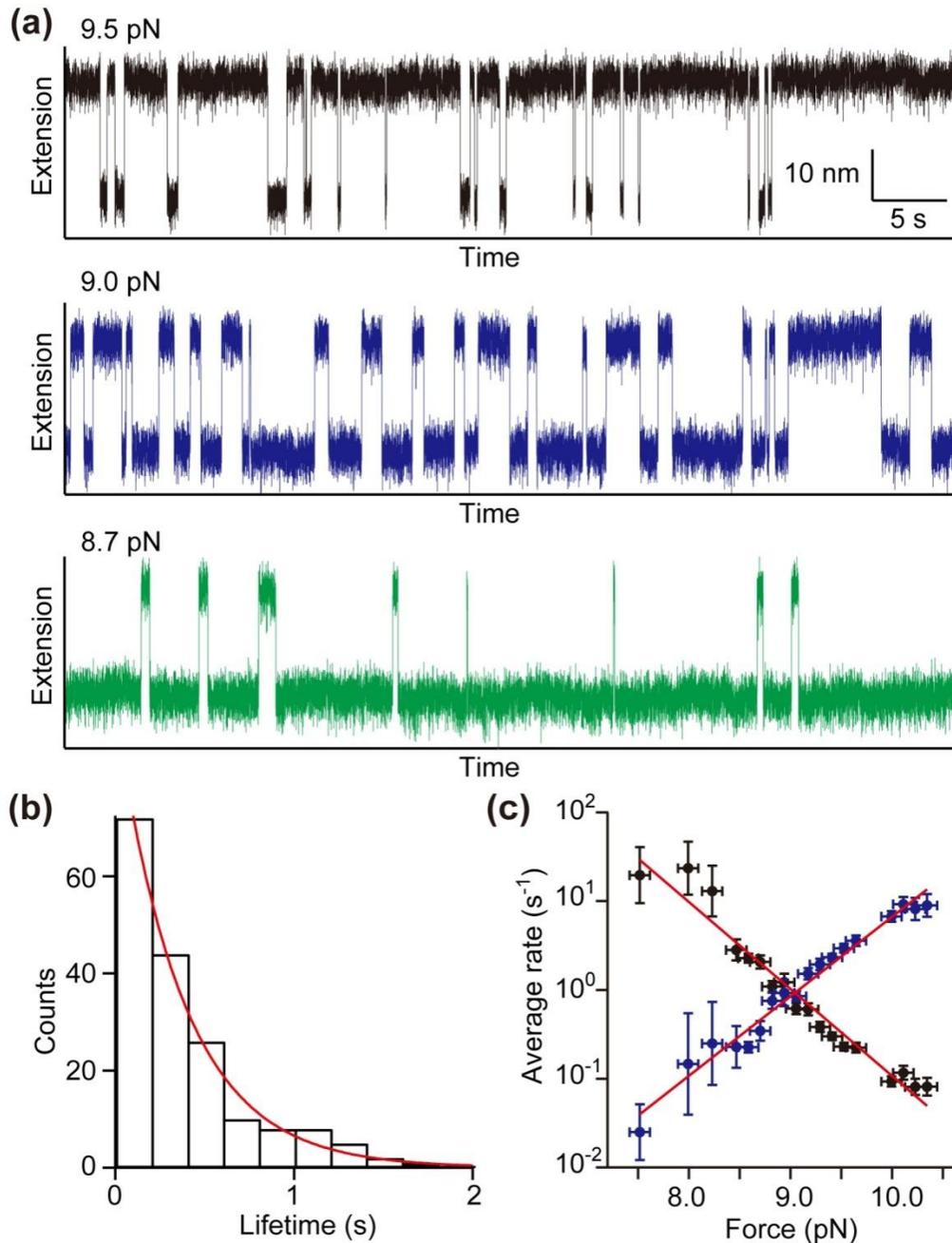


**Figure 8.12 Free energy estimation from the Crooks theorem.** Representative work distributions at different loading rates for unfolding (solid lines) and refolding (dashed lines) show the crossing point around 66 kJ/mol.

The landscape parameters from the FEC analysis also allowed the diffusion constant for barrier crossing to be determined by making some assumptions regarding the shape of the energy profile. For example, based on Eq. 2.1, under the assumption of linear-cubic potential profiles,

$$D = \frac{\pi}{3} \left[ \frac{k(\Delta x^\ddagger)^2}{\Delta G^\ddagger/k_B T} \right] \exp\left(\frac{\Delta G^\ddagger}{k_B T}\right). \quad (8.2)$$

As for  $\tau_{\text{tp}}$ , the values were calculated independently for unfolding and refolding. For the unfolding transition,  $D = 6 \times 10^{-13 \pm 1} \text{ m}^2/\text{s}$ , whereas for refolding, essentially the same result was found,  $D = 3 \times 10^{-13 \pm 0.6} \text{ m}^2/\text{s}$ . Since the values are again expected to be the same, we averaged them to obtain  $D = 4 \times 10^{-13 \pm 0.5} \text{ m}^2/\text{s}$ .

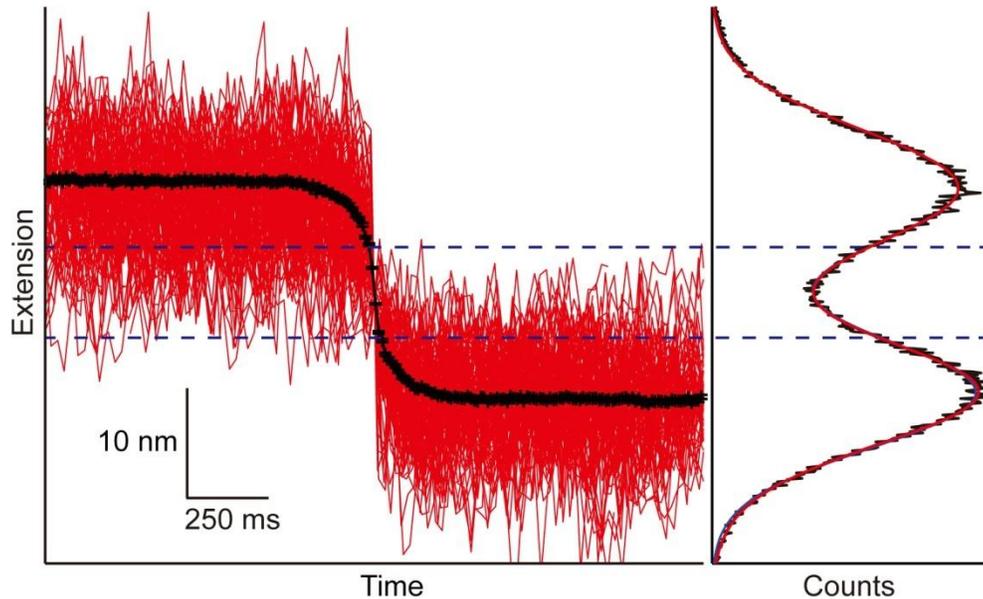


**Figure 8.13 Constant-force trajectories of PrP folding.** (a) Extension records at constant force for three different forces show abrupt jumps as PrP unfolds/refolds in a two-state process. (b) The distribution of lifetimes for N and U are single-exponential. Here the distribution and exponential fit are shown for N at 9.4 pN. (c) The rates for folding (black) and unfolding (blue) vary exponentially with force, crossing at force  $F_{1/2}$ . Error bars show standard error.

The reconstructed profile indicates a two-state system with a barrier roughly midway between the native and unfolded states under 9 pN of tension. As an independent test of whether this overall picture is correct, we analyzed equilibrium measurements of the extension at constant force, in the range  $\sim 8\text{--}10$  pN. Representative records at different forces (Fig. 8.13a) were filtered to remove the short-lived off-pathway states and the lifetimes of the two states were determined by threshold analysis (Woodside et al., 2006b). The lifetimes were single-exponentially distributed at each force (Fig. 8.13b), as expected for a transition with a single rate-determining barrier. The folding and unfolding rates varied exponentially with force (Fig. 8.13c) and were well fit, over the limited force range sampled, by the Bell formula (Bell, 1978) Eq. 3.7. These fits yielded a barrier  $10.4 \pm 0.6$  nm from the folded state at  $\sim 8\text{--}10$  pN, in good agreement with the location of the barrier in Fig. 8.7b. The rates from constant force measurements also agreed reasonably well with the rates obtained from the FECs (Fig. 8.11). The force producing equal folding and unfolding rates,  $F_{1/2}$ , was  $9.1 \pm 0.1$  pN.

However, for the above analysis of the energy barrier, the diffusion constant and the transition path time assumed a specific shape for the landscape profile (*i.e.*, linear-cubic). To verify that this assumption was reasonable, we repeated the calculation based on the reconstructed energy landscape profile. However, attempts to calculate  $D$  and  $\tau_{\text{tp}}$  from the undeconvolved profile (Fig. 8.13b) produce incorrect results: using the curvatures of the wells and barrier from Fig. 8.7b and the folding/unfolding rates at 9 pN,  $D = 6 \times 10^{-15 \pm 0.4}$  m<sup>2</sup>/s and hence  $\tau_{\text{tp}} = 3 \times 10^{2 \pm 0.4}$   $\mu\text{s}$ . Both of these values are 2 orders of magnitude slower than those obtained from the FEC kinetic analysis above;  $\tau_{\text{tp}}$  is also 1 order slower than the upper limit of  $\sim 40$   $\mu\text{s}$  found by direct examination of the trajectories (Fig. 8.14). This slower diffusion behavior can be explained by inducing an

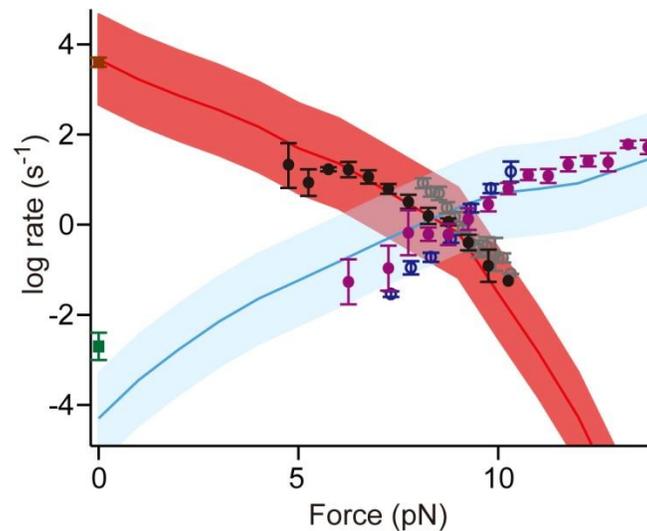
effective diffusion constant  $D_{\text{eff}}$ , which includes the diffusion of the protein plus the handles together. Under this framework, our result agrees well with previous finding that the effective diffusion of the construct is tether dependent (Berkovich et al., 2012).



**Figure 8.14 Direct estimate of an upper bound for the transition time.** Segments of the unfiltered constant-force extension trajectories, centered around the folding transitions, were extracted from the full trajectories, aligned on the center of the transition (red: 100 representative transitions), and averaged to remove Brownian fluctuations (black: average of over 3,300 transitions). An upper bound for the transition time was estimated at  $\sim 40 \mu\text{s}$  from the average time required to move between the inflection points of the folded and unfolded distributions (dashed lines).

Using the correctly deconvolved energy landscape profile which contains the information of the protein only, we were able to test the consistency of the deconvolved profile with the results found previously from FECs and constant force measurements. At 9.1 pN tension, the barrier is located  $\sim 8.5 \text{ nm}$  from the folded state and  $\sim 11.5 \text{ nm}$  from the unfolded state (with an error of one bin width, or 1.5 nm), in good agreement with the  $\Delta x^\ddagger$  values from constant-force trajectories. Tilting the profile back to zero force (Fig. 8.10b, red), we compared it to the results of the FEC kinetic analysis. The barriers

previously found at zero force for leaving the folded state (Fig. 8.10b, black) and the unfolded state (Fig. 8.10b, blue) agree within error with the barrier in the reconstruction, within error (the uncertainty in the barrier height after deconvolution is estimated at  $\pm 5$  kJ/mol). The free energy of the native state estimated from Crooks theorem (Fig. 8.10b, black circle) also agrees within error with the reconstructed landscape, validating the reconstruction and deconvolution. In addition, the shape of the landscape supports the earlier assumption of a linear-cubic profile for determining  $D$  and  $\tau_{ip}$ .



**Figure 8.15 Comparison between experimental and predicted rates.** The rates for folding and unfolding found from FEC analysis (black: folding; purple: unfolding) and constant-force measurements (grey: folding; blue: unfolding) agree within error with the rates calculated directly from the deconvolved landscape profile by Kramers theory, when tilted at various forces (red: folding; cyan: unfolding). Shaded areas represent error in calculated rates. Zero-force folding (brown) and unfolding (green) rates are from chemical denaturation measurements (Wildegger et al., 1999).

Next, we tilted the deconvolved profile over a wide range of forces and calculated the folding and unfolding rates predicted by Kramers theory (Eq. 2.1). Since the rates at all forces depend on the same diffusion constant  $D$ , this effectively amounts to fitting globally all the force-dependent rates in Fig. 8.11c, Fig. 8.11e, and Fig. 8.13c. The

global fit yielded  $D = 8 \times 10^{-11 \pm 1} \text{ m}^2/\text{s}$ , consistent within error with the result from the FEC kinetic analysis. The predicted rates (Fig. 8.15, red: folding, cyan: unfolding) recapitulate the force-dependent rates found from both the FEC kinetic analysis (Fig. 8.15, black: folding, purple: unfolding) and the constant-force measurements (Fig. 8.15, grey: folding, blue: unfolding), over four orders of magnitude. They also agree well, within error, with the rates found at zero force (Fig. 8.15, brown: folding, green: unfolding) from chemical denaturation studies (Wildegger et al., 1999).

Finally, we used the fitted value of  $D$  to recalculate  $\tau_{\text{p}}$  from the deconvolved profile. The result,  $\tau_{\text{p}} = 2 \times 10^{-2 \pm 1} \mu\text{s}$ , is somewhat faster than the previous result from the kinetic analysis, but consistent given the relatively large uncertainty.

## 8.4 Discussion

The results presented above show that the protein folding energy landscape profile reconstructed by measuring single molecules under tension can be used to predict key properties, ranging from commonly-studied observables such as the folding rates to rarely-characterized measures such as the transition path time, as well as fundamental quantities such as the diffusion constant. To assess the validity of these results, we compared them to independent measurements of PrP and other proteins using different techniques. Considering first the rates, the folding rate predicted at zero force from Fig. 8.10b,  $5 \times 10^{3 \pm 1} \text{ s}^{-1}$ , agrees well with the rate found for PrP from ensemble chemical denaturation measurements,  $\sim 4 \times 10^3 \text{ s}^{-1}$  (Wildegger et al., 1999). This agreement between ensemble and single-molecule results supports the notion that the same barrier is being probed in the two different measurements, despite the different modes of denaturation

(Dudko et al., 2011). The Kramers prefactor  $k_0$  calculated at zero force,  $10^{7\pm 1} \text{ s}^{-1}$  from Fig. 8.10 or  $10^{5\pm 0.5} \text{ s}^{-1}$  from the FEC kinetic analysis, is also consistent with the range found experimentally in previous work:  $10^6\text{--}10^8 \text{ s}^{-1}$  for contact formation in unfolded polypeptides and proteins (Krieger et al., 2003; Nettels et al., 2007), though it is lower than a previous estimate of  $10^9 \text{ s}^{-1}$  from unfolding ubiquitin by atomic force microscopy (Popa et al., 2011).

The force-dependence of the rates depends primarily on the distance to the barrier,  $\Delta x^\ddagger$ : very small  $\Delta x^\ddagger$  leads to “brittle” behavior where the rate changes little with force, whereas larger  $\Delta x^\ddagger$  results in “compliant” behavior with a strong force-dependence. PrP folding is clearly compliant, similar to proteins like helical coiled-coils (Gebhardt et al., 2010) or calmodulin (Junker et al., 2009), but contrasting with the folding of many other proteins like titin (Rief et al., 1997), ubiquitin (Popa et al., 2011), and GFP (Dietz and Rief, 2004). This suggests that the key interactions holding the structure together are located far from the points at which force is applied, allowing large extension fluctuations to occur before the protein reaches the transition state. Given that each unfolded amino acid extends  $\sim 0.22 \text{ nm}$  at  $F_{1/2}$ ,  $\Delta x^\ddagger$  corresponds to  $\sim 40$  amino acids unfolded. Such a non-compact transition state was recently suggested by phi-analysis of PrP folding (Hart et al., 2009), which showed a structural nucleus developing between the helices 2 and 3 of the native fold (Fig. 6.1). One possible transition state consistent with both the single-molecule and phi-analysis results would retain the core of the molecule (most of helix 2 and 3, the adjacent loop and  $\beta$ -strand 2) while unfolding  $\beta$ -strand 1 and helix 1. This picture is also supported by NMR studies showing that  $\beta$ -strand 1 is relatively unstable (Julien et al., 2009), but additional experiments, possibly combining phi-analysis with

single-molecule measurements, will be required to explore the nature of the transition state in more detail.\*

The most interesting aspect of the landscape analysis is the ability to quantify properties such as the diffusion constant over the barrier and the transition path time, which are extremely difficult to measure by any method. For the diffusion constant, an average of the results from the two methods used above (kinetic analysis and landscape deconvolution) yields the final result  $D = 1 \times 10^{-12 \pm 0.4} \text{ m}^2/\text{s}$ . This value is similar to the range reported from fluorescence studies of intrachain diffusion in unfolded proteins: typically  $D \sim 10^{-10} - 10^{-11} \text{ m}^2/\text{s}$  (Chen et al., 2009; Hagen et al., 1996; Nettels et al., 2007), although in one case as low as  $10^{-13} \text{ m}^2/\text{s}$  (Waldauer et al., 2010). Crucially, however, our result characterizes diffusion across the barrier itself, the critical parameter in Kramers theory. In contrast, previous work described diffusion within the unfolded state only, which may be different (Best and Hummer, 2010). The value we find is at the low end of the range, as would be expected since any structure formed near the transition state would presumably slow the configurational diffusion relative to the unstructured chain. We are not aware of any other published measurements of  $D$  for barrier crossing in protein

---

\* It was recently suggested from a study of apomyoglobin that compliant folding may be indicative of a molten globule rather than the native structure (Elms et al., 2012). We do not believe that we are observing a molten globule state for PrP, for several reasons. First, PrP was held at low force ( $< 2 \text{ pN}$ ) for 1–2 s between unfolding events, more than sufficient time to form the native structure given the very fast folding rate (Wildegger et al., 1999). Similar conditions in the apomyoglobin study were indeed claimed to lead to the native structure; the molten globule was only observed when insufficient refolding time was allowed at low force (*e.g.* in constant-force measurements). In contrast to this previous study, however, we did not see any change in  $\Delta x^\ddagger$  for unfolding when comparing FECs to constant force measurements. Second, there is independent support for an extended transition state from phi-analysis (Hart et al., 2009). Third, the agreement between the folding rate predicted from the landscape at zero force and the measured rate strongly suggests that the same barrier is being probed in the pulling measurements and ensemble denaturation studies.

folding, for more direct comparison. However, our results are consistent with a previous analysis of force spectroscopy measurements of titin unfolding (Carrion-Vazquez et al., 1999; Hummer and Szabo, 2003), which implied a value  $D = 1 \times 10^{-14 \pm 2} \text{ m}^2/\text{s}$  (although  $D$  was not explicitly calculated).

Turning to the transition path time, an average of the results from the kinetic analysis and landscape deconvolution yields  $\tau_{\text{tp}} = 2 \times 10^{0 \pm 0.4} \mu\text{s}$ . This is much faster than the characteristic timescale for folding at zero force,  $\sim 100 \mu\text{s}$ , but in excellent agreement with the only direct measurements of  $\tau_{\text{tp}}$  for folding proteins, which found  $\tau_{\text{tp}} \sim 2 \mu\text{s}$  for the WW domain of the formin binding protein and  $\tau_{\text{tp}} < 10 \mu\text{s}$  for the protein GB1 (Chung et al., 2012). Our result also agrees very well with an atomistic equilibrium molecular dynamics simulation in explicit solvent, which found  $\tau_{\text{tp}} = 0.5 \pm 0.1 \mu\text{s}$  for the FiP WW domain (Shaw et al., 2010), corresponding to  $\tau_{\text{tp}} \sim 1.5 \mu\text{s}$  after viscosity correction (Chung et al., 2012). Interestingly, PrP is twice as large as GB1 and three times as large as the WW domains, but still has a similar transition path time, suggesting that  $\tau_{\text{tp}}$  is relatively insensitive to protein size. The topology of PrP is also quite distinct from that of the other proteins, suggesting that  $\tau_{\text{tp}}$  is also insensitive to topology. In contrast, both protein size (Ivankov et al., 2003; Li et al., 2004) and the topology of the native fold (Baker, 2000) are known to play important roles in determining the folding rate.

In conclusion, these results demonstrate the power of single-molecule methods for quantifying the fundamental properties driving folding reactions. They also represent a stringent test of the underlying landscape theories, showing that landscape theory can be used to understand force spectroscopy measurements in a comprehensive, unified way. Not only can the location and height of free energy barriers be determined, and indeed the full profile of the landscape, but these can even be used to predict folding and unfolding

rates over many orders of magnitude, as well as the values of elusive quantities such as the diffusion constant and transition path time. A similar analysis could be applied to force spectroscopy measurements for many other proteins, pointing the way towards increasingly quantitative experimental applications of folding landscape theories.

## **9. Structural conversion of prion protein dimers via a multi-step pathway observed at the single-molecule level\***

Having described the folding pathways accessible to monomeric PrP, we now go one step further to study the interactions that drive misfolding of PrP in the context of aggregates by characterizing the folding pathways of dimers formed by joining two PrP molecules at their termini. Single dimer molecules held under tension in optical tweezers were unfolded and refolded by ramping the force up and down. Remarkably, although isolated PrP molecules folded rapidly into the native structure in a two-state process, neither domain of the dimer ever formed the native fold. Instead, the dimer invariably formed a stable, misfolded structure via multiple partially-folded intermediates. Only a single misfolding pathway was observed. We mapped the energy landscape underlying the misfolding and identified a key intermediate that leads to misfolding by blocking the formation of native structure. These results provide mechanistic insight into the formation of non-native structures by the prion protein (Yu et al., under preparation).

### **9.1 SMFS of PrP dimers and the observation of stable misfolded structures**

In order to observe directly the formation of stable misfolded structures, we used force spectroscopy to investigate the folding of individual dimers of PrP, as the smallest form of oligomer. PrP dimers have been suggested to play a role in conversion from PrP<sup>C</sup>

---

\* A version of this chapter has been submitted for publication. Yu, H., Dee, D. R., Liu, X., Brigley, A. M., Sosova, I. and Woodside, M. T. Structural conversion of prion protein dimers via a multi-step pathway observed at the single-molecule level. In this work, H. Y. helped build the instrument, performed experiments with X.L., and analysed data with D.R.D. Proteins were provided by A.M.B. and I.S..

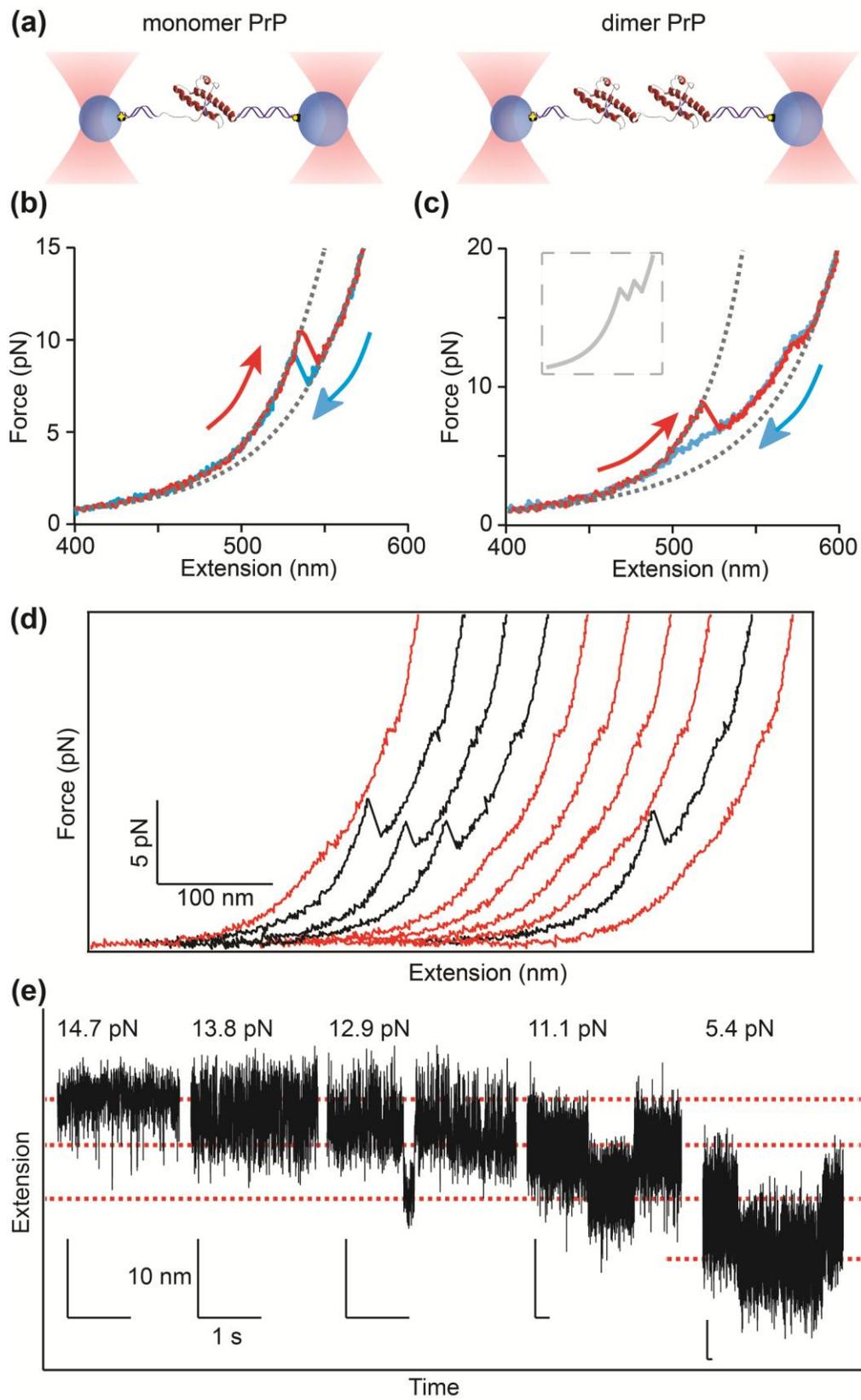
to PrP<sup>Sc</sup> (Tompa et al., 2002), although larger oligomers seem to be more infectious (Silveira et al., 2005). Recombinant PrP forms dimers both at low pH (Gerber et al., 2008; O'Sullivan et al., 2007) and upon dilution from 0.2 to 0.05% SDS (Jansen et al., 2001; Kaimann et al., 2008), while brain-derived PrP dimerizes *in vitro* (Meyer et al., 2000). A model for conversion of PrP<sup>C</sup> to amyloid based on dimeric domain swapping has been proposed (Lee and Eisenberg, 2003), inspired by the domain-swapped dimer found in the crystal structure of human PrP (Knaus et al., 2001). Most interestingly, synthetic PrP dimers have been shown to be toxic to neurons both in culture (Roostae et al., 2009; Simoneau et al., 2007) and in mouse models of prion disease (Simoneau et al., 2007). By covalently connecting two monomers of hamster PrP(90-231) end-to-end similarly to these synthetic dimers, we were able to compare the structures formed by two PrP molecules held in close contact to those formed by isolated monomers.

Dimers of SHaPrP were made in two ways: first, from monomers of PrP(90-231), the protease-resistant fragment of PrP<sup>Sc</sup>, linked via disulfide bonds between cysteine residues engineered at the termini (Zheng et al., 2011); and second, expressed as a single tandem-repeat protein from a synthetic gene. The DNA for the genetic dimer construct with an N-terminal His tag was synthesized (DNA 2.0) and inserted in the pJexpress406 plasmid. The 35-kDa protein was expressed in *E. coli* C41 (DE3) (Lucigen) and purified. Briefly, cell pellets were resuspended in lysis buffer (6 M GdnHCl, 50 mM phosphate, 500 mM NaCl, 0.5 mM PMSF, 20 mM imidazole, pH 7.4) and sonicated for three 10-s intervals.  $\beta$ -mercaptoethanol and Tween 20 were added to a final concentration of 10 mM and 0.5%, respectively, before the lysate was centrifuged, filtered, and purified by FPLC (GE Healthcare) using a Ni-NTA column (Qiagen) equilibrated with the lysis buffer. The sample was then washed with lysis buffer and eluted with elution buffer (6 M GdnHCl,

50 mM phosphate buffer, 250 mM imidazole, pH 7.4). Monomeric PrP(90-231) was expressed, purified, and refolded as described previously in chapter 7. Before attaching DNA handles, protein was reduced with TCEP in a 100:1 molar ratio with PrP for 30 min, desalted by spin column (Zeba, Thermo Scientific) to remove excess TCEP, and activated with 2,2'-dithiodipyridine (Sigma-Aldrich). Incompletely-activated PrP molecules formed dimers upon re-oxidation. Monomeric PrP and recombinant PrP dimers were expressed and purified by Angela Brigley and Iveta Sosova.

PrP dimers were attached to DNA handles linked specifically to labeled beads held in dual-beam optical tweezers (Fig. 9.1a), as previously described in chapter 7. The extension of the molecules was measured as the traps were moved apart at a constant rate to ramp up the force, and then again as the traps were brought back together, thereby generating FECs of the unfolding and refolding, respectively. FECs of dimeric PrP were compared to FECs of monomeric PrP measured in the same way.

In the case of the monomer, as described previously, the force increased monotonically as the DNA handles were stretched out, until the protein unfolded in a single step around 10 pN (Fig. 9.1b, red), creating a sawtooth “rip” pattern (due to the abrupt extension increase and concomitant force drop as the unfolded protein was stretched out by the tension) that is characteristic of cooperative unfolding (Borgia et al., 2008). Refolding FECs retraced the same two-state, cooperative behavior (Fig. 9.1b, cyan), with a slight hysteresis due to the high pulling speed. The change in contour length during the structural transitions,  $\Delta L_c$ , was found by fitting the FECs to extensible WLC models (Eq. 4.3) in series for the protein and handles.  $\Delta L_c$  agreed exactly with the result expected:  $34.3 \pm 0.4$  nm as compared to 34.3 nm from the NMR structure of natively-folded SHaPrP (James et al., 1997).



**Figure 9.1 SMFS measurements of PrP dimers show the presence of stable misfolded states.** (a) Individual PrP molecules (left: monomers, right: dimers) were linked via disulfide bonds to DNA handles bound to beads held in optical traps. (b) Unfolding (red) and refolding (cyan) FECs of PrP monomers show apparently two-state folding. Dashed lines show WLC fits for the native and unfolded states. (c) Unfolding (red) and refolding (cyan) FECs of PrP dimers show the formation of stable non-native structures, in contrast to the sequential unfolding of identical domains that would be expected for independently-folded PrP<sup>C</sup> domains (inset). (d) Two types of unfolding FECs were observed in successive unfolding/folding cycles: type 1 (black) and type 2 (red). (e) Representative constant force trajectories at different forces showed the complex misfolding pathway of PrP dimers. Red dashed lines label different states along the misfolding pathway.

Considering next the dimers, if the two monomeric domains were independently folded into their native structure, then one would expect to observe two consecutive sawtooth rips in the FECs, with unfolding forces and contour length changes matching the values seen for monomeric PrP (Fig. 9.1c, inset). Such behavior has been observed very commonly in force spectroscopy measurements of many different proteins where monomers were connected together to form tandem repeats (Cao and Li, 2007; Carrion-Vazquez et al., 2003; Dietz et al., 2006; Jollymore et al., 2009; Lee et al., 2006; Rief et al., 1997; Rief et al., 1999). The actual behavior observed in measurements of the PrP dimer, however, was strikingly different, indicating that neither domain was natively folded. FECs of dimer unfolding (Fig. 9.1c, red) often displayed one large rip at a force lower than native unfolding ( $\sim 8$  pN), followed by a shoulder-like region and then a distinctly smaller rip at a force higher than native unfolding ( $\sim 15$  pN). Refolding FECs retraced the small rip at high force, but not the large rip at lower force, displaying instead a broader shoulder region at  $\sim 5$ -10 pN (Fig. 9.1c, cyan). Repeated unfolding-refolding cycles showed that some of the unfolding FECs did not contain the large rip at low force, matching instead the shape of the refolding FECs; these curves, too, however, showed none of the characteristics of the unfolding of native PrP<sup>C</sup> (Fig. 9.1d). For convenience, the two classes of unfolding were denoted type 1 (Fig. 9.1d, black) and type 2 (Fig. 9.1d,

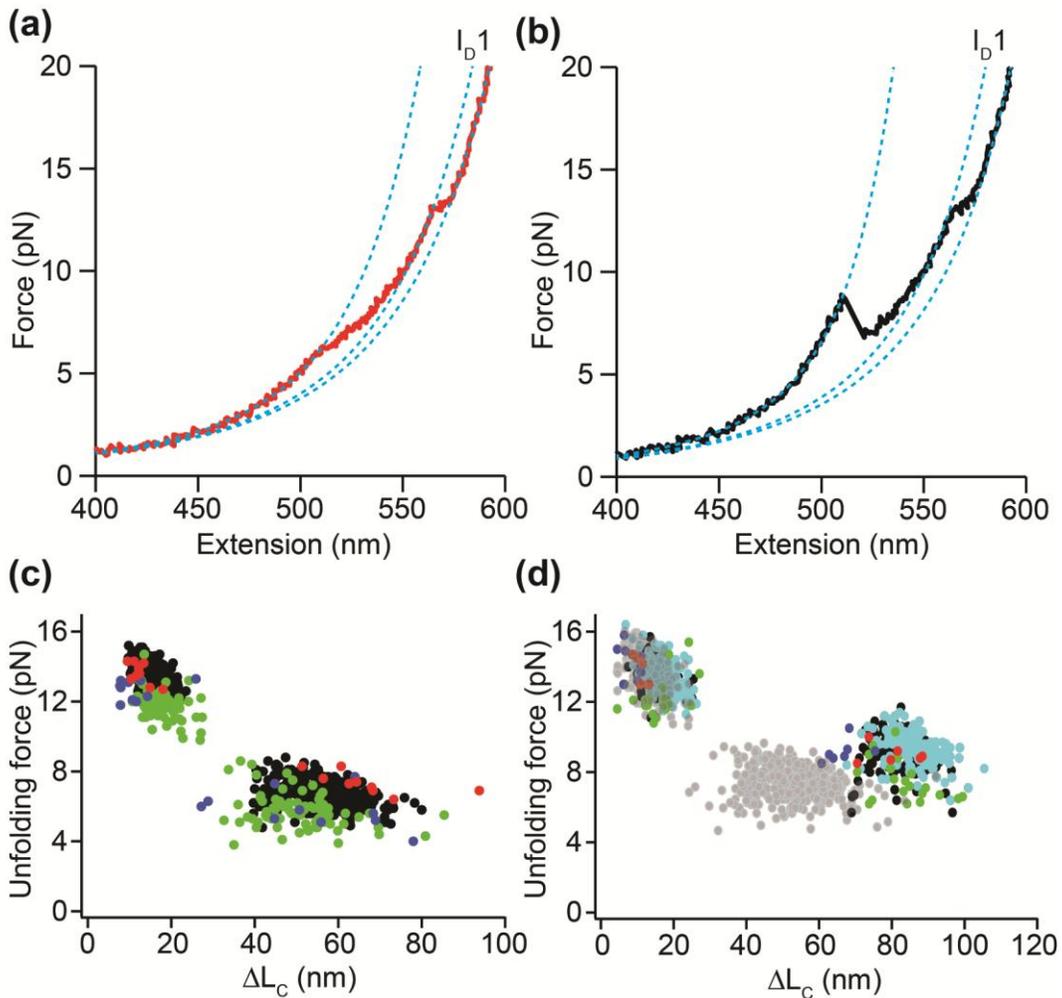
red). Constant force measurements of the dimer further confirmed the formation of non-native structures via a complex folding pathway (Fig. 9.1e).

To characterize the structures formed by the dimers more quantitatively, we measured the total change in contour length upon complete unfolding by fitting the FECs to WLC models as for the monomer (Fig. 9.2, dashed lines). In the case of the type 2 unfolding, the total  $\Delta L_c$  from 768 FECs was found to be  $56 \pm 3$  nm (Fig. 9.2a). This is significantly less than twice the  $\Delta L_c$  for unfolding PrP<sup>C</sup> (68.6 nm), indicating that some of the residues involved in the structured domain of PrP<sup>C</sup> were unfolded. In contrast, for type 1 unfolding the total  $\Delta L_c$  was  $81 \pm 1$  nm from 644 FECs (Fig. 9.2b), significantly more than twice the  $\Delta L_c$  for unfolding PrP<sup>C</sup>. The latter result indicates that some of the residues that are unstructured in the natively-folded monomer (namely residues 90–124) were in fact sequestered within the misfolded dimer structure. We estimated the number of amino acids,  $n_{aa}$ , in the misfolded dimer using the  $\Delta L_c$ - $n_{aa}$  relation:  $n_{aa} = (\Delta L_c + d_T)/L_c^{aa}$ . In the case of monomeric PrP<sup>C</sup>,  $d_T = 3.1$  nm is known from the NMR structure (James et al., 1997), but no such structural information is available for the misfolded dimer. Assuming slightly larger values of  $d_T \sim 4$  and 5 nm for type 2 and type 1 unfolding, respectively, the dimer structures involved approximately 170 amino acids for type 2 unfolding and 240 for type 1.

## 9.2 Metastable sequential intermediates on the misfolding pathway

The presence of multiple rips in the FECs indicates the existence of metastable intermediates in the misfolding pathway, in contrast to the two-state folding of monomeric PrP<sup>C</sup>. To investigate the misfolding pathways in more detail and determine

the number of intermediates, we fit the intermediate regions of the FECs to WLC models as done above for the high- and low-force regions. Since all unfolding and refolding FECs showed a discrete rip around 15 pN, we first fit this high-force transition (Fig. 9.2a). The contour length change was found to be the same for all unfolding FECs (whether type 1 or 2) as well as refolding FECs,  $15 \pm 1$  nm (from 2882 curves). We denoted this state  $I_{D1}$  (first intermediate in the dimer). The structure and stability of the states can be characterized in terms of  $\Delta L_c$ , and the unfolding force ( $F_{unf}$ ). Therefore the two dimensional  $F$ - $\Delta L_c$  plot of the different states provides the unique fingerprint of the specific pathway the molecule unfolds through. A total of 700 individual FECs collected from five dimer molecules were collected to examine the heterogeneity of the dimer folding pathways. The  $F$ - $\Delta L_c$  plot (Fig. 9.2 c and d) displayed identical clusters of the initial states and  $I_{D1}$  for both types of FEC behaviors, indicating all dimer molecules followed the same unfolding pathway.



**Figure 9.2 Intermediates on the misfolding pathway of PrP dimers.** (a) WLC fits (dashed lines) to different parts of a representative type 2 FEC (red) reveal the contour length changes ( $\Delta L_c$ ) in the dimer. The total  $\Delta L_c$  is 56 nm, more than the 34-nm change for a single PrP<sup>C</sup> domain but less than two. An intermediate, labeled  $I_{D1}$ , unfolds at ~15 pN, with  $\Delta L_c$  of 15 nm. A "shoulder" with non-WLC behavior at 5–10 pN indicates additional intermediate states. (b) WLC fits (dashed lines) to a representative type 1 FEC (black). The total  $\Delta L_c$  is 81 nm, more than twice the change for a single PrP<sup>C</sup> domain.  $I_{D1}$  is again observed unfolding at ~15 pN. (c)  $F-\Delta L_c$  plot of four different molecules (different colors) from type 2 FECs, showing the same unfolding force and contour length change. (d)  $F-\Delta L_c$  plot of five different molecules (different colors) from type 1 FECs. Compared with the fingerprint from type 2 FECs (grey), the intermediate state  $I_{D1}$  was on pathway to both types of dimer molecules.

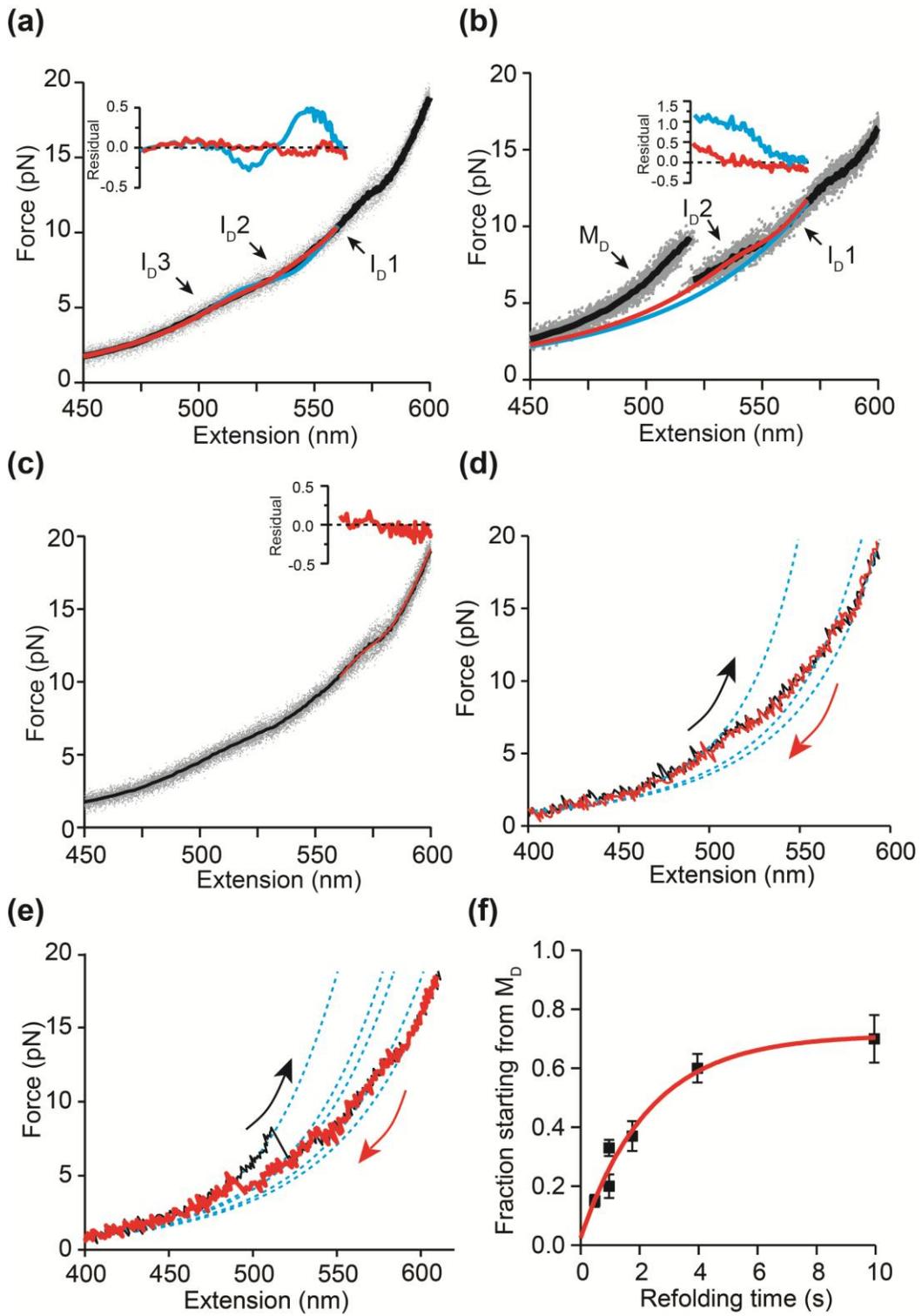
As noted above, in addition to the discrete rip around 15 pN the FECs also contain a shoulder-like feature around 5–10 pN, where the curves deviate from the expected WLC behavior. Such features can arise from transitions that are fast compared to the rate at which the force and extension are measured. When the folding/unfolding rate is slow, then only a single transition (as for native folding in Fig.9.1b) or possibly a small number (as for I<sub>D</sub>1 in Fig. 9.2a) are observed, each appearing as an abrupt rip in the FECs. When the transition rates are fast, however, the average effect of many transitions is measured at each point in the FEC, resulting in a smooth increase in extension as the time spent in the unfolded state increases with force. We modeled this shoulder-like feature by assuming a rapid quasi-equilibrium for each transition with a force-dependent unfolding probability, since the presence of a series of intermediates which unfold/refold at equilibrium confounded the identification of distinct species for fitting with the WLC model, which is classically used to fit non-equilibrium transitions. In this quasi-equilibrium WLC model, each transition was assumed to be two-state and each state was characterized by its contour length change ( $\Delta L_c$ ) and an equilibrium unfolding force ( $F_{1/2}$ ). The extension of the full construct was described as the sum of the extensions from the handle, the amount of unfolded polypeptide prior to the transition, and the average extension expected from the transition itself:

$$x(F) = x_H(F) + x_U(F) + \sum_{i=1}^n P_u^i(F) \Delta x_i(F), \quad (9.1)$$

where  $x_H(F)$  is the extension of the handles,  $x_U(F)$  is the extension of the unstructured portion of the protein,  $n$  is the number of structures being unfolded and  $\Delta x_i(F)$  is the extension of the  $i$ th structure of the protein. The force-dependent extension,  $\Delta x(F)$ , can be related to  $\Delta L_c$  by the WLC model (Eq. 4.3). The probability of any state being unfolded at

a given force value obeys the Boltzmann relation:  $P_u^i(F) = \{1 + \exp[(F_{1/2}^i - F) \cdot \Delta x_i(F)]\}^{-1}$  (Eq. 3.6).

Considering first the type 2 unfolding curves and the refolding curves, which both exhibit the same behavior, we averaged each set of curves measured on the same molecule and fit the average to Eq. 9.1 (Fig. 9.3a). We found that the shoulder feature in these curves could not be well fit by a single unfolding transition into  $I_{D1}$  (Fig. 9.3a, cyan), but it was well fit by two sequential transitions (Fig. 9.3a, red): we denoted the two additional states  $I_{D2}$  and  $I_{D3}$ , as shown. In type 2 curves, the dimer thus unfolded in the sequence  $I_{D3} \rightarrow I_{D2} \rightarrow I_{D1} \rightarrow U$  (where  $U$  denotes the unfolded state). Refolding curves, which simply retraced the type 2 unfolding curves (Fig. 9.3d), therefore followed the reverse sequence of states. From the fits, the average  $\Delta L_c$  values were  $23 \pm 1$  nm for the transition from  $I_{D3}$  to  $I_{D2}$  and  $17 \pm 2$  nm from  $I_{D2}$  to  $I_{D1}$ ; the average  $F_{1/2}$  values were respectively  $6.5 \pm 0.4$  pN and  $8.5 \pm 0.4$  pN. The same kind of fitting was applied to type 1 unfolding curves, but only to the portion of the curves after the large rip at low force. In this case, the average curves were well fit by a single transition with  $\Delta L_c = 17 \pm 2$  nm and  $F_{1/2} = 9.0 \pm 0.3$  pN (Fig. 9.3b, red). Since these values are the same as for  $I_{D2}$  to  $I_{D1}$ , we conclude that unfolding in type 1 curves followed the sequence  $M_D \rightarrow I_{D2} \rightarrow I_{D1} \rightarrow U$  (where  $M_D$  denotes the fully-misfolded state of the dimer).



**Figure 9.3 Quasi-equilibrium transitions between on-pathway intermediates in the shoulder region of FECs.** (a) An average (black) of multiple type 2 FECs (grey) was well fit in the shoulder region (5–10 pN) by a model that assumed two sequential, rapid transitions (red) rather than a single transition (cyan), as shown by the significant reduction of the fit residual (inset). The dimer unfolded sequentially through three different structures in type 2 curves:  $I_{D3}$ ,  $I_{D2}$ , and  $I_{D1}$ . (b) The shoulder-like feature in type 1 FECs was well fit assuming the presence of an additional, rapid transition (red), as shown by the significant reduction of the fit residual (inset), with the same characteristics as the  $I_{D2}$ – $I_{D1}$  transition in type 2 FECs. (c) Unfolding FECs (grey), here type 2, were averaged (black) and fit over the range 10–20 pN to a model assuming a single unfolding transition, corresponding to  $I_{D1}$  (red). (d) Refolding FECs (red) had the identical behavior as type 2 unfolding FECs (black). (e) A refolding curve (red) shows complete refolding from U to  $M_D$ . All intermediate states ( $I_{D1}$ ,  $I_{D2}$  and  $I_{D3}$ ) are on the pathway to  $M_D$ . A type 1 unfolding curve is shown for comparison (black). (f) The fraction of FECs exhibiting type 1 behavior rose with the waiting time at low force between successive pulls. A single-exponential fit (red) yielded a rate of  $0.5 \pm 0.2 \text{ s}^{-1}$  for the transition from  $I_{D3}$  to  $M_D$ .

Refolding FECs almost always ended up in the state  $I_{D3}$ , without any observable transition into  $M_D$ . Very rarely, however, complete sequential refolding from U to  $M_D$  passing through each of the three intermediates was in fact observed (Fig. 9.3e), indicating that all five states are indeed on a single folding pathway. The fact that  $M_D$  was almost never observed directly during refolding suggests that  $M_D$  must form slowly at low force. Consistent with this picture, the fraction of curves exhibiting type 1 behavior could be increased by increasing the waiting time at low force between successive unfolding curves (Fig. 9.3f). A single-exponential fit yielded a rate of  $0.5 \pm 0.2 \text{ s}^{-1}$  for forming  $M_D$ . We thus concluded that all FECs in fact represent the same underlying pathway consisting of 5 sequential states:  $M_D \leftrightarrow I_{D3} \leftrightarrow I_{D2} \leftrightarrow I_{D1} \leftrightarrow U$  (Fig. 9.4a). Type 1 unfolding curves don't show evidence for  $I_{D3}$  because it has a lower average unfolding force than  $M_D$  ( $7.0 \pm 0.5 \text{ pN}$  compared to  $9.1 \pm 0.5 \text{ pN}$ ) and hence the two states always unfold concurrently. Interestingly, we note that the first 3 transitions in the folding pathway (Fig. 9.4a) are fast compared to the native folding rate, since they all are near or at equilibrium (*i.e.*, the unfolding and refolding curves are coincident), whereas for the same pulling rates native folding is out of equilibrium at the same pulling rates (Fig. 9.1b).

In contrast, the final step in misfolding (formation of  $M_D$ ) is very much slower than the native folding rate,  $0.5 \pm 0.2 \text{ s}^{-1}$  compared to  $\sim 10^4 \text{ s}^{-1}$  (Hart et al., 2009; Wildegger et al., 1999).

### 9.3 PrP dimer energy landscape

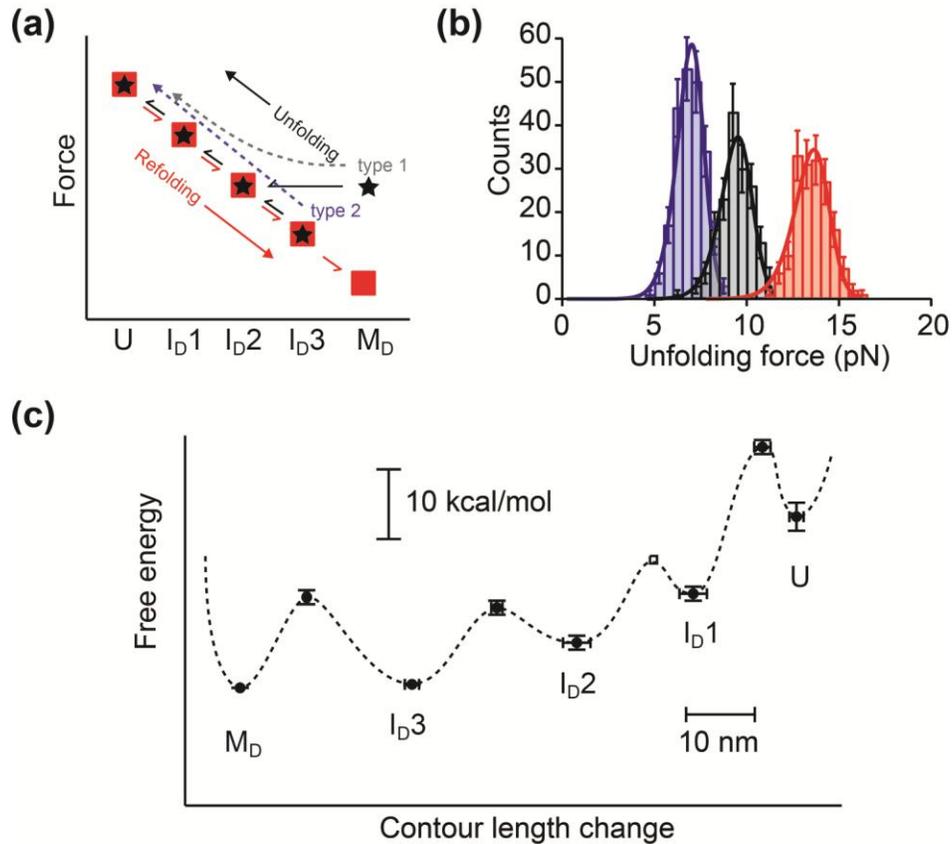
Having elucidated the folding pathway of the dimer from the sequence of transitions in the FECs, we analyzed the unfolding forces to deduce the shape of the fundamental energy landscape underlying the observed behavior. First, we found the free-energy difference between each of the 5 states in the sequential folding pathway, treating the pathway as a series of two-state transitions. In the case of the transitions between U,  $I_D1$ ,  $I_D2$ , and  $I_D3$ , all of which were fast and hence in quasi-equilibrium in the FECs, the free-energy difference in each transition was found from fitting the average FECs as in Fig. 9.3:  $\Delta G = F_{1/2} \cdot \Delta x(F_{1/2}) - \Delta G_{stretch}(F_{1/2})$  (Eq. 4.14). In the case of the transition between  $I_D3$  and  $M_D$ , which was not in equilibrium, we used the fit to the transition kinetics (Fig. 9.3f) to determine the equilibrium free energy from the asymptotic ratio of the state occupancies at long delay times:  $\Delta G = k_B T \ln(M_D / I_D3)$ . The results are listed in Table 9.1. As a consistency check, we also found the total free energy change for complete unfolding by integrating the FECs to obtain the distribution of non-equilibrium work done on the molecule and then applying the Jarzynski equality (Jarzynski, 1997) to determine the free energy from the distribution of non-equilibrium works. The result,  $\Delta G_{mf} = (29 \pm 2) - (6 \pm 2) = 23 \pm 3 \text{ kcal/mol}$ , agreed within error with the sum of the free energy changes for the individual transitions,  $24 \pm 2 \text{ kcal/mol}$ .

**Table 9.1 Structure, kinetic and stability parameters of the misfolded dimer and intermediate states.**

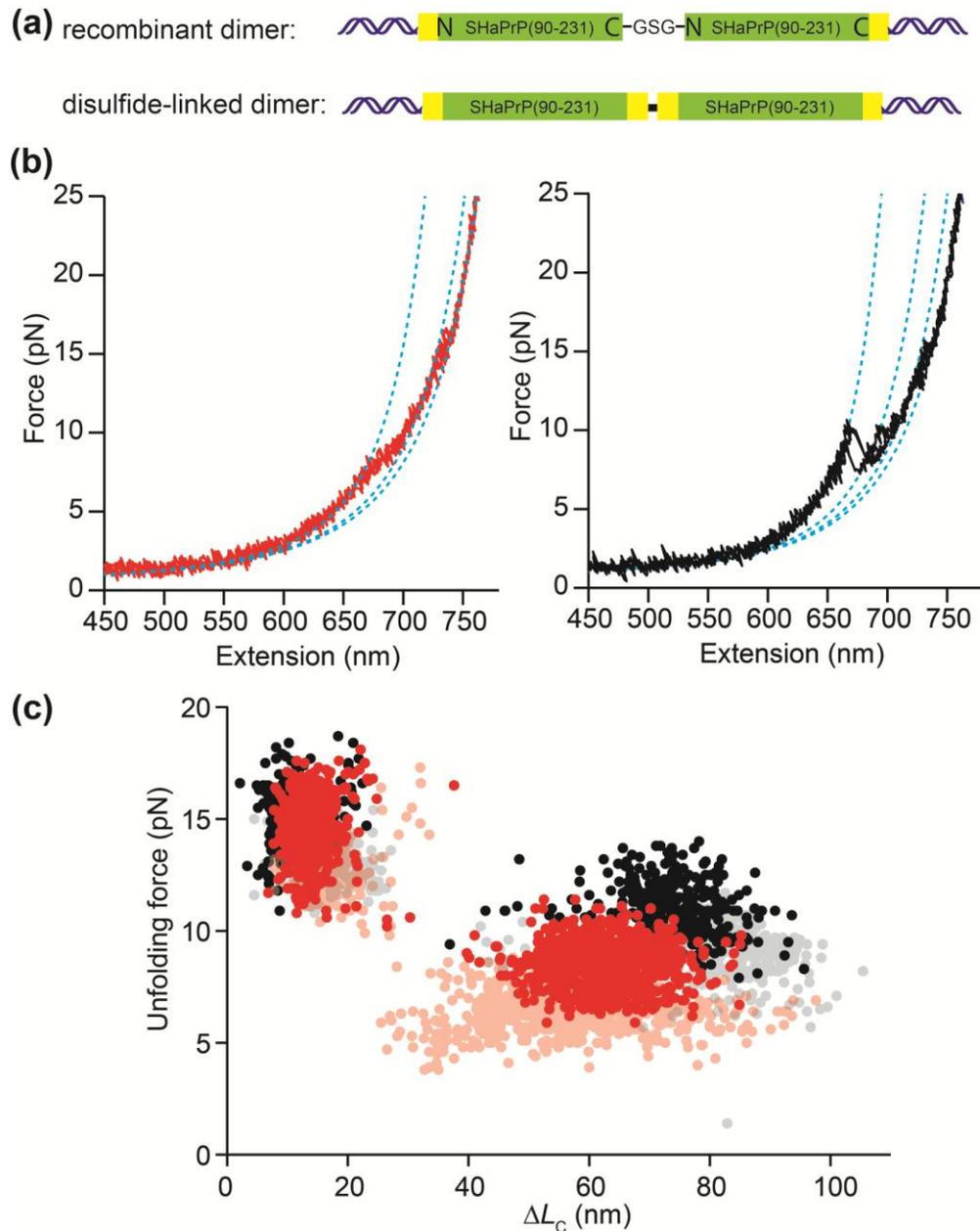
	<b>M<sub>D</sub></b>	<b>I<sub>D</sub>3</b>	<b>I<sub>D</sub>2</b>	<b>I<sub>D</sub>1</b>
<b><math>\Delta L_C</math> to U from WLC fits (nm)</b>	81 ± 1	56 ± 3	--	15 ± 1
<b><math>\Delta L_C</math> to next state from equilibrium fits (nm)</b>	--	23 ± 1	17 ± 2	14.4 ± 0.4
<b><math>F_{1/2}</math> from equilibrium fits (pN)</b>	--	6.5 ± 0.4	8.5 ± 0.4	13.1 ± 0.5
<b><math>\Delta G</math> to next state (kcal mol<sup>-1</sup>)</b>	0.5 ± 0.1	6 ± 1	7 ± 1	11 ± 1
<b>Log<math>k_{\text{unfold}}</math> (s<sup>-1</sup>)</b>	- 6.8 ± 0.3	- 5.0 ± 0.4	--	- 11.6 ± 0.9
<b><math>\Delta X^{\ddagger,0}</math> (nm)</b>	12.0 ± 0.3	13.1 ± 0.8	--	12.8 ± 1.2
<b><math>\Delta G^{\ddagger,0}</math> (kcal mol<sup>-1</sup>)</b>	13 ± 1	11 ± 1	--	21 ± 1

Finally, we determined the locations and heights of the barriers between each state, key properties of the misfolding energy landscape, from the distribution of unfolding forces (Fig. 9.4b). For a given transition, the unfolding force distribution can be related to the unfolding rate at zero force,  $k_{\text{unfold}}$ , the distance to the barrier along the reaction coordinate,  $\Delta x^{\ddagger}$ , and the height of the barrier,  $\Delta G^{\ddagger}$ , using the Dudko model (Dudko et al., 2006). The distributions for unfolding M<sub>D</sub>, I<sub>D</sub>3 and I<sub>D</sub>1 were each fit to Eq. 9.1, with the results listed in Table 9.1. A complementary analysis based on the force-dependent rates was also used (Dudko et al., 2008). Because discrete unfolding events could not be observed for the rapid transition between I<sub>D</sub>2 and I<sub>D</sub>1, the properties of the barrier between these states could not be quantified. The barrier positions and heights for each transition were then combined piecewise with the free-energy differences and

contour length changes between the 5 different states to reconstruct the energy landscape profile for the dimer misfolding (Fig. 9.4c).



**Figure 9.4 Pathway and energy landscape for PrP dimer misfolding.** (a) The refolding pathway for the dimer (red) leads from the unfolded state to  $M_D$  sequentially through each intermediate, at lower force. The unfolding pathway starts from  $M_D$  (type 1) or  $I_{D3}$  (type 2) sequentially through each intermediate to the unfolded state, skipping over  $I_{D3}$  in type 1 unfolding because of its low unfolding force. (b) Unfolding force distributions for  $M_D$  (black),  $I_{D3}$  (blue) and  $I_{D1}$  (red) are well fit by Equation 3, yielding parameters describing the energy landscape for misfolding. (c) PrP dimer misfolding energy landscape at zero force, reconstructed from FECs. Energy and contour length changes are plotted with respect to  $M_D$  (Error bars: s.e.m).



**Figure 9.5 Topology of the PrP dimers.** (a) PrP dimers were made in two ways: either expressed as a single protein with the two domains (NC-NC topology) connected by a GSG linker (top), or by covalent linkage of terminal Cys residues (yellow) by a disulfide bond (bottom). In each case, DNA handles (blue) were attached to terminal Cys residues. (b) Representative FECs from the recombinant dimer with WLC fits (cyan) of both type 1 (black) and type 2 (red). (c)  $F-\Delta L_c$  plot of recombinant dimer molecules from both type 1 (black) and type 2 (red) FECs, showing the same unfolding force and contour length change with the disulfide-linked dimers (type 1: grey; type 2: pink), suggesting all PrP dimers we measured have the same NC-NC topology.

## 9.4 Topology of the PrP dimer

The dimers formed from cysteine-labelled monomers could, in principle, contain monomers connected in three different orientations: NC-NC, NC-CN and CN-NC (where N and C represent the N- and C-termini). These three topologies should give very distinct folding/unfolding behavior, as the chain topology strongly affects folding (Shank et al., 2010). In practice, all the dimer molecules displayed identical folding behavior (Fig. 9.2c and d), indicating that these molecules shared the same topology. The topology of the disulfide-linked dimers was identified by comparison to the folding of a PrP dimer expressed as a single protein (thus ensuring the NC-NC topology, Fig. 9.5a). The FECs (Fig. 9.5b) and  $\Delta L_c - F_{unf}$  plots (Fig. 9.5c) from the recombinant dimer were qualitatively and quantitatively the same as those from the disulfide-linked dimers, indicating that they all shared the NC-NC topology. It is possible that dimers connected with other orientations do not form as readily (e.g. through a higher tendency to form aggregates), therefore only NC-NC topology was selectively observed in our experiments.

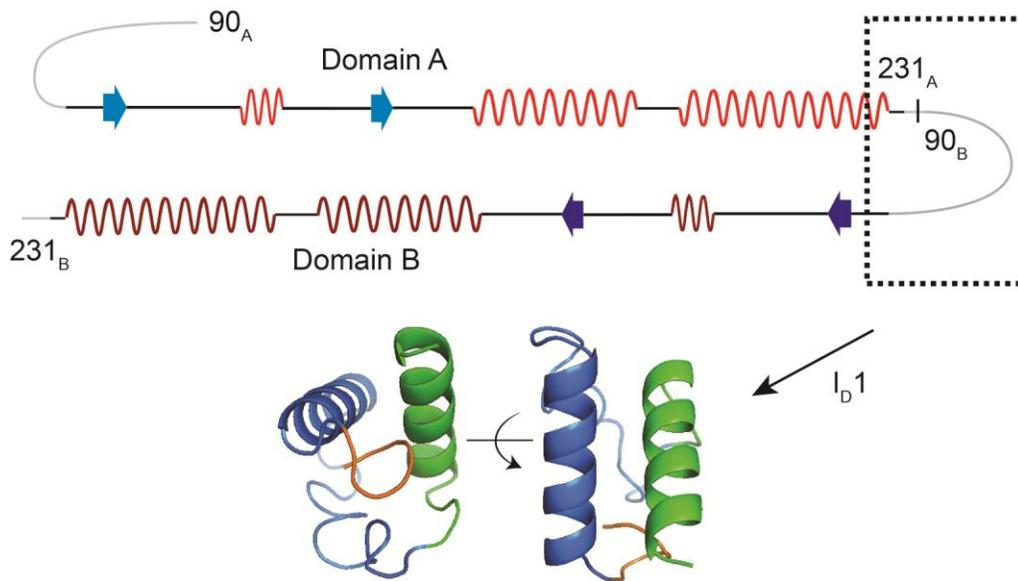
## 9.5 PrP dimer invariably misfolds

To our knowledge, the energy landscape for a misfolding protein has never before been reconstructed quantitatively. One of the most striking aspects of these results is how frequently and easily the PrP dimer misfolds: it invariably follows a single pathway to the same misfolded state,  $M_D$ . Moreover, whereas monomeric PrP folds as a simple two-state system, the dimer forms multiple metastable intermediates in sequence along the misfolding pathway. Simply joining two PrP domains together in tandem thus completely changes the folding landscape of each domain, leading to a different structural outcome. Interestingly, the uniformity of the PrP dimer misfolding contrasts sharply with

the heterogeneous behavior observed for other aggregation-prone proteins like  $\alpha$ -synuclein and A $\beta$ , dimers of which were found by force spectroscopy to form a variety of structures (Kim et al., 2011; Krasnoslobodtsev et al., 2012; Neupane et al.). Instead, the dimer misfolding behaves remarkably like the native folding of a “regular” protein that has evolved to have a minimally-frustrated landscape (Oliveberg and Wolynes, 2005), with none of the complexity that should arise from the rugged landscape expected for aggregation (Eichner and Radford, 2011). Even the change from two-state to multi-state folding as the structure grows from 104 amino acids in PrP<sup>C</sup> to ~ 240 in the misfolded dimer is consistent with the behavior of well-behaved natively-folded proteins, since proteins larger than 100-150 amino acids almost invariably fold via intermediates (Brockwell and Radford, 2007; Jackson, 1998).

The fact that the dimer never forms native structure stands in stark contrast to the behavior observed for the many other proteins that have also been studied at the single-molecule level as tandem-repeat oligomers. Such oligomers have been widely studied, especially using AFM force spectroscopy (Cao and Li, 2007; Carrion-Vazquez et al., 2003; Dietz et al., 2006; Jollymore et al., 2009; Lee et al., 2006; Rief et al., 1997; Rief et al., 1999), and in almost all cases native folding was predominant, with little or no misfolding reported. In work where the extent of misfolding was quantified, typically the level was low. For example, AFM measurements of repeats of the titin I27 domain found that neighboring domains misfolded into a larger structure approximately 2% of the time (Oberhauser et al., 1999), whereas misfolding occurred 4% of the time for repeats of the tenascin FN III domain (Oberhauser et al., 1999) and 3-8% of the time for spectrin repeats (Randles et al., 2007). Misfolding of I27 tandem repeats was also measured using single-molecule fluorescence and found to occur up to 5% of the time (Borgia et al.,

2011). Optical tweezers measurements of calmodulin, which consists of two independently-folding domains connected by a helical linker, also revealed rare misfolding events (less than 1% of the time) (Stigler et al., 2011). Notably, however, none of these misfolded states were thermodynamically stable, since they converted to the native structure over timescales of seconds (Stigler et al., 2011) to days (Borgia et al., 2011).



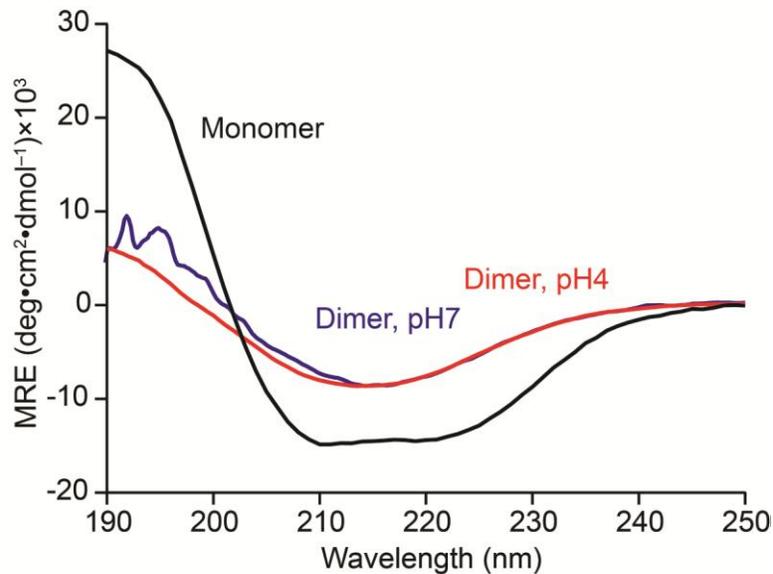
**Figure 9.6 Intermediate state I<sub>D</sub>1.** The secondary structure of PrP<sup>C</sup> is mapped onto the dimer as it would occur if each domain folded natively. The intermediate I<sub>D</sub>1 guides the dimer down the misfolding pathway by helping prevent native structure formation. Since it does not form in monomeric PrP, it likely spans the region between the two natively-structured domains. A possible 52-amino acid structure for I<sub>D</sub>1 obtained from *ab initio* modeling is illustrated.

In contrast, the misfolded conformation of the PrP dimer is indeed thermodynamically stable: the stability of M<sub>D</sub>,  $\Delta G_{unf} = 24 \pm 2$  kcal/mol, is slightly higher than the sum of the stabilities of two natively folded domains ( $\Delta G_{unf} = 11 \pm 1$  kcal/mol per monomer). The stability of the misfolded dimer also contrasts with the misfolded states of the monomer, all of which were considerably less stable than the native structure.

Nevertheless, the fact that  $M_D$  is the thermodynamic minimum does not account for the observation exclusively of misfolding: given that the energy difference between  $M_D$  and two  $PrP^C$  domains is only  $\sim 2$  kcal/mol, equilibrium considerations would suggest that the native structure should be observed  $\sim 4\%$  of the time. One possible reason for the lack of native folding in the dimer is that the intermediate  $I_D1$ , which forms at a force substantially higher than the force at which  $PrP^C$  folds, prevents the formation of  $PrP^C$  in either domain, thereby selecting preferentially for the misfolding pathway on kinetic as well as thermodynamic grounds. This intermediate thus likely plays a key role in the misfolding. Once the dimer has formed the misfolded state  $M_D$ , the barrier to return to the native structure (via the unfolded state) would be very high,  $\sim 36$  kcal/mol, since  $M_D$  must be unfolded completely. Interestingly, this barrier height is similar in magnitude to the result estimated from an ensemble kinetic study of misfolding (Baskakov et al., 2001).

Since  $I_D1$  was not observed in monomeric PrP, it must involve interactions between residues in both domains of the dimer. From the 15-nm contour length change upon unfolding, we estimate that  $I_D1$  consists of 45-55 amino acids. It thus almost certainly encompasses the region spanning the link between the two domains, *i.e.* the C-terminal residues of the first domain and the N-terminal residues of the second domain (Fig. 9.6). Residues 109-122 of PrP are predicted to have a strong propensity to form  $\alpha$ -helical structure (Zhang et al., 1995), and a structurally-ambivalent “chameleon sequence” capable of forming different secondary structures depending on the context provided by neighboring sequences (Kuznetsov and Rackovsky, 2003) was identified in residues 114-125 (Kuznetsov and Rackovsky, 2004), suggesting that this region could indeed form the nucleus for a stable structure in the context of interactions with the C-terminal residues of a neighboring domain, despite the fact that it is unstructured in monomeric PrP.

Secondary structure prediction and *ab initio* modeling (Xu and Zhang, 2012) applied to 50-amino acid windows spanning the segment from the C terminus of the first domain to the chameleon sequence of the second domain yielded possible structures for I<sub>D</sub>1, including a helix-turn-helix motif (Fig. 9.6). Previous force spectroscopy measurements of similar structures in leucine zippers (Gebhardt et al., 2010; Xi et al., 2012) yielded similar mechanical properties, suggesting that such a model is plausible, if speculative.



**Figure 9.7 CD spectrum of PrP dimers.** CD spectrum of recombinant dimers shows an increased  $\beta$ -sheet content compared with the spectrum of monomeric PrP<sup>C</sup> (black). The dimer CD spectrum was similar at pH4 (red) and pH7 (blue), yielding the same secondary structure content within error. A long path length (2 mm) was used for the spectrum at pH7 owing to the need to reduce the protein concentration (1  $\mu$ M) to prevent aggregation.

## 9.6 Structure of the misfolded dimer and relevance of the misfolded states

Force spectroscopy yields a wealth of information about the pathway for the misfolding, but it does not provide a high-resolution structure. Nevertheless, the results provide constraints for building or testing structural models. For example, in the only high-resolution crystal structure of PrP (Knaus et al., 2001), made from human PrP(90-231), PrP forms a domain-swapped dimer in which each monomeric domain is structured as in monomeric PrP<sup>C</sup> (Zahn et al., 2000) but the largest helix (helix 3) swaps positions in the two domains. This domain-swapped structure could, in principle, form in the tandem dimer, but in fact it is incompatible with the force spectroscopy results, since the stable intermediate I<sub>D</sub>1 is formed in part from the residues in the N-terminus that remain disordered in the crystal structure. Other structural models for PrP dimers were generated from MD simulations using constraints from chemical cross-linking and mass-spectrometry analysis (Kaimann et al., 2008). A structure with the N termini from each monomer domain in close proximity but distal to the C termini would not be possible in our system owing to the constraints imposed by the domain connection topology. However, evidence was also found for an alternate structure in which G90 cross-linked with E221, which would be consistent with the repeat topology; unfortunately no model for this structure was published. A dimer of the structural model proposed for PrP<sup>Sc</sup> based on MD simulations (DeMarco and Daggett, 2004) would yield  $\Delta L_c$  upon unfolding of ~ 90 nm, somewhat larger than the observed value but in the right general size range. A parallel  $\beta$ -sheet model of PrP amyloid (Cobb et al., 2007) would imply  $\Delta L_c \sim 70$  nm for the case of a dimer, again in the right range although somewhat smaller than observed. In contrast, the observed  $\Delta L_c$  is much lower than what would be predicted by models in

which each monomeric domain is structured from residues 90 to 230,  $\Delta L_c \sim 100$  nm (Smirnovas et al., 2011; Yang et al., 2005). The high stability of our dimers also contrasts with the suggestion of a partially-denatured dimeric amyloid precursor existing in an equilibrium with monomeric PrP (Stohr et al., 2008).

A key question is whether the misfolded dimer remains mostly helical, like PrP<sup>C</sup>, or has been converted to a more  $\beta$ -rich form, like PrP<sup>Sc</sup>. To test this question, we measured the CD spectrum of the tandem dimer. Whereas monomeric PrP exhibits the classic spectrum of a helical protein (Fig. 9.7, black), the spectrum of the dimer is qualitatively different, exhibiting the shape characteristic of a structure rich in  $\beta$ -sheets (Fig. 9.7, red). There is thus a substantial conversion of structure from helices to sheets in the misfolded dimer. Such a conversion to a soluble,  $\beta$ -rich form has been studied previously as a potential intermediate step in PrP<sup>Sc</sup> formation, being induced by low pH (Bjorndahl et al., 2011) or a combination of moderately low pH and partially-denaturing conditions (Baskakov, 2002; Gerber et al., 2008; O'Sullivan et al., 2007). Here, however, low pH is not required:  $\beta$ -rich structures are seen both at pH 4 and neutral pH (Fig. 9.7, blue). Quantifying the secondary structure content from the CD spectra, we found that the structure converted from 43% helical and 13% sheets in the monomer (characteristic of PrP<sup>C</sup>) to 11% helical and 35% sheets in the dimer. Suggestively, the low helical content is consistent with recent work showing that the helical C terminus of PrP<sup>C</sup> is likely fully converted to  $\beta$ -strands in PrP<sup>Sc</sup> (Smirnovas et al., 2011), in contrast to earlier models positing the retention of significant C-terminal helix content (DeMarco and Daggett, 2004; Govaerts et al., 2004). However, relating the bulk CD spectroscopy to the structures observed with force spectroscopy is complicated by the fact that the dimers oligomerize in bulk, as observed previously (Roostae et al., 2009; Simoneau et al., 2007).

We thus cannot rule out the possibility that individual misfolded dimers undergo additional restructuring upon oligomerization, before the CD spectra are measured. Nevertheless, these results do suggest that the misfolded dimer,  $M_D$ , is likely rich in  $\beta$ -strands. This interpretation is also consistent with the results of previous SM fluorescence studies of PrP aggregation, which found evidence for the rapid formation of  $\beta$ -rich dimers as the first step in aggregation (Post et al., 1998).

Dimerization has long been suspected to play an important role in pathogenic conversion of PrP<sup>C</sup> (Tompa et al., 2002). Despite the previous evidence for the relevance of dimeric states, however, including the toxicity of “artificial” dimers formed from linked molecules (Simoneau et al., 2007), it is very difficult to ascertain whether any of the species we have observed at the single-molecule level are in fact related to pathogenesis *in vivo*. Nevertheless, although any connection to disease is only speculative, our results show that PrP is uniquely pre-disposed to conversion into misfolded structures through intermolecular interactions. Indeed, even the smallest oligomer, a dimer, can rapidly and reliably convert to an apparently  $\beta$ -rich form that is more stable than PrP<sup>C</sup>. This misfolded structure likely acts as a first step along the aggregation pathway, as reflected in the fact that the dimer aggregates much more rapidly than monomeric PrP<sup>C</sup>. By mapping out the pathway for structural conversion of a disease-related protein, our work represents an exciting new approach for understanding the microscopic mechanisms of protein misfolding.

## 10. Conclusions and future directions

### 10.1 Summary and implications of the work

In this thesis, we demonstrated a high-resolution SMFS method to study folding and misfolding of single RNA and protein molecules using optical tweezers. The folding and misfolding pathways of *add* adenine riboswitch (Neupane et al., 2011), PrP monomers (Yu et al., 2012b) and dimers (Yu et al., under preparation) have been characterized, and the corresponding energy landscapes have been analyzed to recover important parameters that govern molecular folding, such as the diffusion constant, transition path time and rates (Yu et al., 2012a). The intermediate and misfolded states observed in our measurements were correlated to key structural components and interactions within the folding pathways. We also suggested how the single-molecule results could relate to the actual function or dysfunction of the molecules. Taken together, our results should help to build a deeper understanding of the protein folding problem, especially as it applies to prion proteins and misfolding. They have also helped establish a general platform for single-molecule studies of disease-related proteins.

#### 1. Biophysics of folding

Molecular folding is governed by energy landscapes. A better characterization of the energy profile, which is difficult to measure experimentally, will provide a more quantitative understanding of the physical principles underlying the folding process. We have pushed the state-of-the-art in experimental measurements of energy landscapes by validating the Hummer-Szabo formalism for energy landscape reconstruction from non-equilibrium measurements (Gupta et al., 2011). Using this method, in combination with a deconvolution procedure (Gebhardt et al., 2010; Woodside et al., 2006a), we recovered

the free energy profile for the native folding pathway of PrP. This is only the second time the landscape profile has been reconstructed for any protein and the first time it has been done using non-equilibrium measurements.

Although in principle both methods for landscape reconstruction (equilibrium and non-equilibrium) should result in the same information, they each have advantages and drawbacks. Landscape reconstructions based on equilibrium measurements typically provide higher spatial resolution. Given that the molecule explores all possible conformations at equilibrium, however, problems can arise if there are multiple folding/misfolding pathways (as for PrP), since all the pathways are projected onto the same molecular extension axis, complicating the interpretation. In contrast, non-equilibrium measurements offer the potential to select for specific pathways kinetically, as we did for PrP, because the fast force ramps reflected primarily the properties of the most prominent (here, native) folding pathway.

Our reconstruction also allowed the best experimental test of the 70-year-old Kramers theory, as it applies to protein folding. Specifically, we were able to fit the rates predicted by Kramers theory for the measured landscape profile to the rates measured both under tension and at zero force in ensemble measurements. This was done over approximately 6 orders of magnitude, with only a single unknown parameter (the diffusion constant). In addition to extracting the diffusion constant, we were also able to use landscape analysis to determine the transition path time for PrP folding. The transition time had been previously found for only one other molecule, using a different approach. These results demonstrate how advances in single-molecule theory and experiment can be used to harness the power of landscape formalisms to describe the mechanics of folding quantitatively.

## 2. Prion science

Our results provided the first detailed, quantitative look at the folding landscape of PrP, revealing a two state native folding pathway along with several misfolding pathways. From our results, at neutral pH and without chemical denaturant there was no intermediate detectable on the native folding pathway with lifetime longer than 50-100  $\mu$ s, the limit of our resolution. This observation does not completely rule out the possibility of a partially-native intermediate, since an intermediate involving restructuring that leaves the end-to-end extension unchanged, such as rearrangement of helix 1 (De Simone et al., 2007), would not be detected in our measurement (although it could be detected by changing the experimental geometry to pull on helix 1 instead of the terminus). It does suggest, however, that the intermediates observed previously may be a function of the particular measurement conditions, such as the presence of chemical denaturants, rather than a general feature of PrP folding.

The most distinctive finding from the force spectroscopy of monomeric PrP was the discovery of at least three different types of partially-folded intermediates off the native folding pathway, under physiological conditions. The fact that this misfolding did not require low pH urges us to reconsider the notion that PrP<sup>Sc</sup> develops in a low pH environment such as endosomes (Caughey et al., 2009), as evidenced by the fact that PrP<sup>Sc</sup> localizes there (Veith et al., 2009) and low pH encourages misfolding (Bjorndahl et al., 2011; Hornemann and Glockshuber, 1998). It is not evident *a priori* that the environments in which most PrP<sup>Sc</sup> is found in the cell—likely those leading to the fastest growth of PrP<sup>Sc</sup>—are necessarily the same as those in which the first PrP<sup>Sc</sup> seeds form. Given that PrP explores non-native states at neutral pH, it is possible that seed formation could take place in non-acidic environments such as the endoplasmic reticulum (*e.g.*,

after being unfolded during membrane translocation), the cytosol (*e.g.*, after retro-translocation (Ma et al., 2002)), or the cell surface (Caughey and Raymond, 1991).

The misfolding of monomeric PrP originated solely from the unfolded state, suggesting that the role of the unfolded state may be under-appreciated (Gerum et al., 2009). Although the misfolded states were quite unstable in wild-type-PrP, they might be stabilized by aggregation-enhancing mutations if they are involved in aggregation. This speculation was supported by measurements of the C179A/C214A mutant, which lacks an internal disulfide bond and is highly aggregation-prone in bulk (Maiti and Surewicz, 2001). The most common misfolded states were significantly enhanced in the mutant, suggesting a role as intermediates on the aggregation pathway. Interestingly, the mutant's aggregation propensity is thus likely driven by the differences in the refolding from the unfolded state rather than the lack of the disulfide bridge, given the absence of disulfide bond in all proteins we used in our experiments.

Equally interesting is what was not observed. First, the amyloidogenic part of the unstructured region (Kuwata et al., 2003) did not form even transient structure on its own, indicating that any structure in this region must be induced by conditions not present in the measurements, such as interactions with other molecules. Second, we found no evidence for a stable misfolded form of monomeric PrP, as proposed recently (Zhou et al., 2012). Since measurements in constant force are in equilibrium, all possible structures are explored during the trajectories, but PrP<sup>C</sup> was the only stable, long-lived structure observed. Hence any alternate structure must either require the truncated residues 23-90 or else specific experimental conditions not found in the single-molecule measurements.

The most striking result was the conversion of PrP into a stable misfolded structure upon forcing two monomers to interact closely as linked dimers. The dimer exclusively misfolded via multiple partially-folded states; kinetic trapping of a metastable non-native intermediate between two adjacent PrP domains apparently prevents the native folding pathway, serving as the molecular basis for the thermodynamically stable protein misfolding. Putting our observations on PrP dimers together with our previous results for monomers begins to suggest a possible answer to the critical question of how misfolding and aggregation start. It is reasonable to speculate that at the monomeric stages, the unfolded PrP molecules, which could for example result from the retro-translocation of PrP to the cytosol, have the ability to sample different misfolding pathways frequently, however all misfolded states are partially-structured and unstable. Upon interaction with another PrP molecule, a misfolded conformation may be stabilized (namely the misfolded dimer state we observed). The requirement for the molecules to be unfolded during this conversion is not implausible, given that neurotoxicity has been found when retro-translocated PrP accumulates in the cytosol (Ma et al., 2002). Most techniques do not have the ability of accessing such an early stage of the prion aggregation process, with only a few molecules participating, but this is exactly the regime that takes best advantage of the capabilities of single-molecule approaches.

### 3. General platform for protein misfolding studies

We made what are, to our knowledge, the first direct measurements of non-native folding pathways for any single-domain protein (the simplest and most commonly-studied type), as well as the first single-molecule observation of a thermodynamically stable misfolded state. In the work presented here, methods for studying rare or transient intermediate/misfolded states based on SMFS were developed, which should be useful

for further investigation of misfolding and aggregation behaviours of disease-related proteins. The threshold for detection of rare and transient states has been pushed to unprecedented levels: occupancies as low as 0.001% and lifetimes as short as 50  $\mu$ s, representing significant improvements over previously published results. Given the widespread existence of prion-like mechanisms in neurodegenerative diseases, these advances enable a general platform for studying the misfolding and transmission of other disease-related proteins at the single-molecule level. This offers the potential of wide applicability for understanding disease pathogenesis and the development of new disease-modifying therapies.

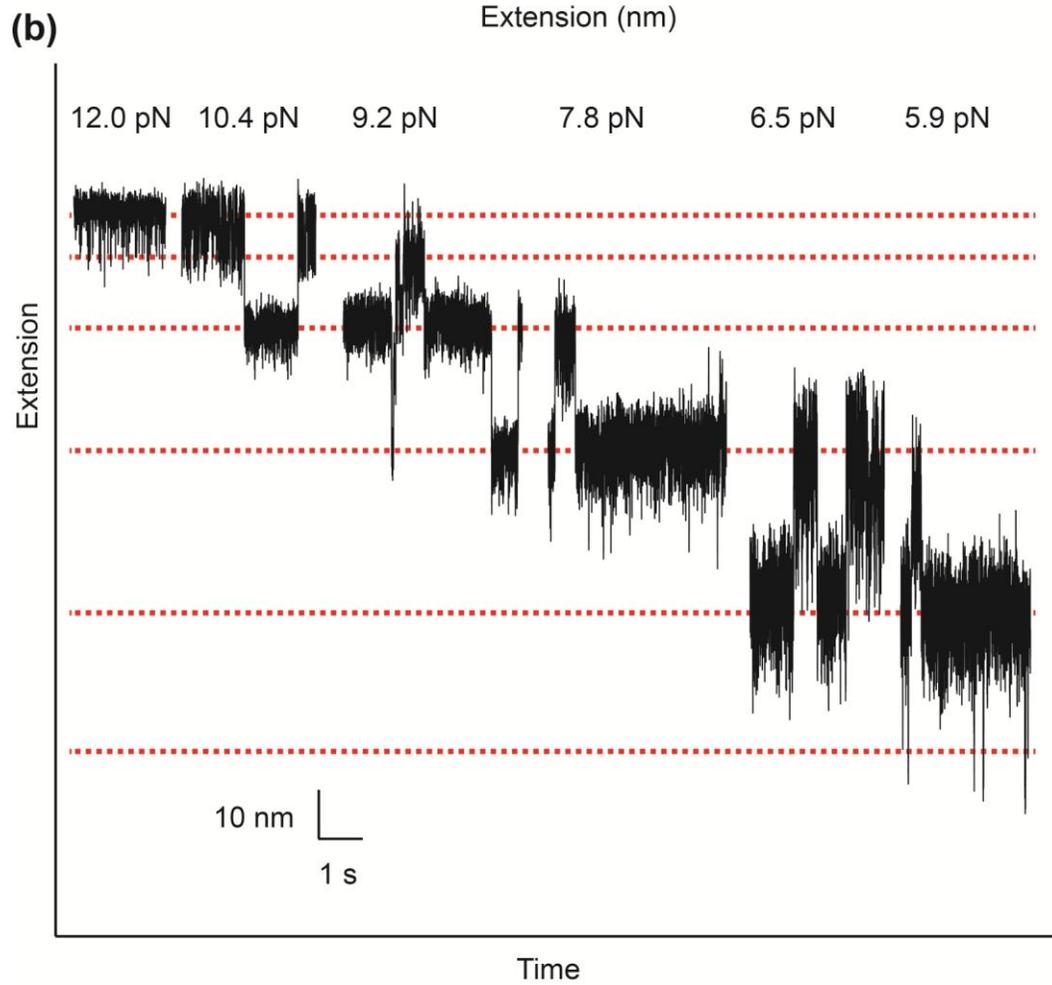
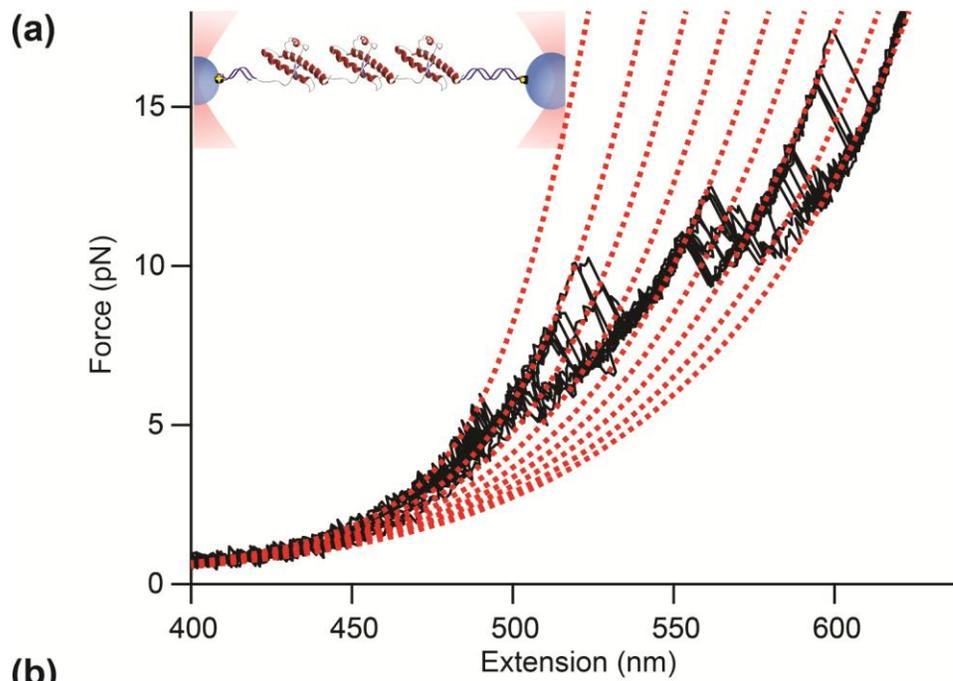
## **10.2 Future directions**

Although many significant questions have been answered and new windows have been opened by the research presented in this thesis, there is still a long way to go before the general mechanism of protein misfolding in disease propagation can be fully understood. Continued studies of small oligomers, larger aggregates and mutational studies at the single-molecule level will lead a better understanding of this question. However, no single technique can fully answer the question by itself. Therefore it is necessary to bring multiple methods to bear on the problem. Computational simulations, in-vivo biochemistry assays, and further development of optical tweezers with FRET abilities promise a more complete picture of prionopathies.

### **10.2.1 SMFS of oligomers and aggregates**

We started our work with the aim of understanding the simplest aspect of the prion problem (within the larger problem of protein folding and aggregation, Fig. 1.1),

namely the folding behavior of isolated PrP monomers. Then we took the first step towards studying aggregation by linking two monomers together to investigate the interactions within a dimer. Although our results suggested that the minimum size of oligomers required for stable misfolding of PrP is the dimer, it is not at all clear what the actual relation is with regard to the misfolded dimer observed in our single-molecule assay and what are the subsequent pathways leading to amyloid fibrils. Since we propose dimerization initiates the aggregation process, how the conversion progresses with increasing oligomer size, and whether these structures have any toxic or pathogenic attributes, would be the next questions to address.

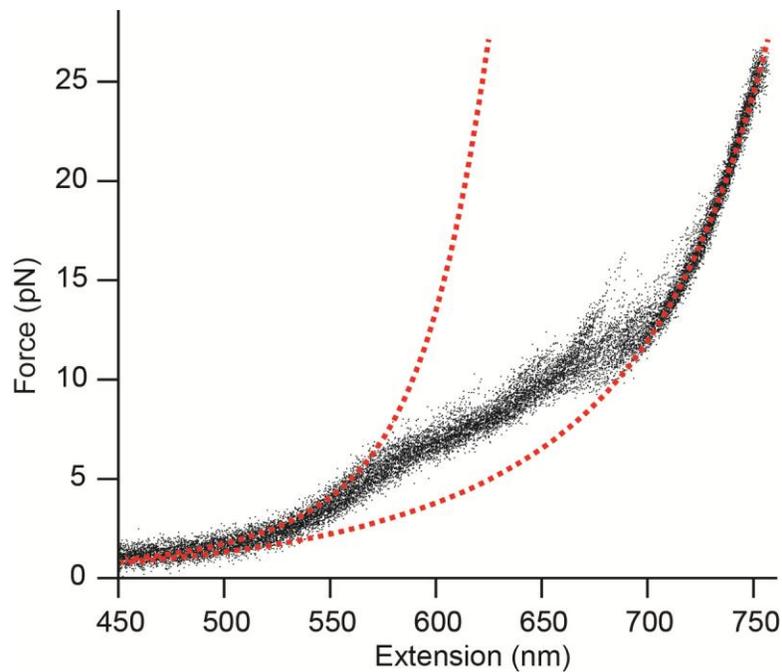


**Figure 10.1 SMFS of PrP trimers.** (a) Representative FECs of PrP trimers show a complex folding pathway of a non-native trimer structure. WLC fits of different states along the folding pathway are shown in dashed lines. The average contour length changes for each state relative to the unfolded state are: 15 nm, 33 nm, 51 nm, 68 nm, 88 nm, 108 nm, 128 nm and 155 nm. (b) Representative constant force trajectories of PrP trimers at different forces are shown in small segments, suggesting the complexity of the PrP trimer folding pathway. The forces are decreasing from the left trace to the right. The dashed lines are indicating the propagation of states as force drops.

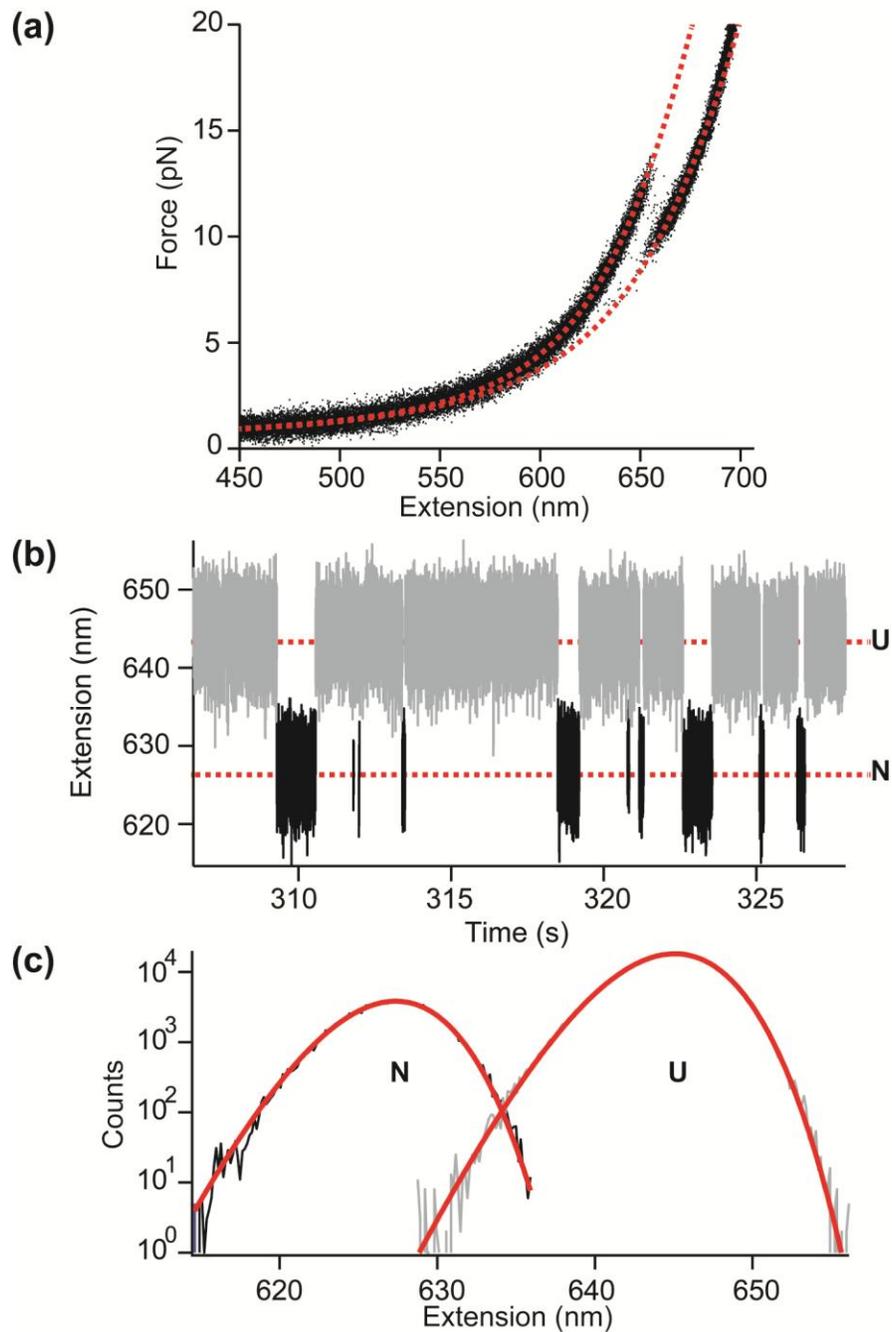
To this end, we will need to investigate higher-order oligomers and aggregates.

For example, PrP trimers have been proposed to serve as the building-blocks for both the spiral (DeMarco and Daggett, 2004) and  $\beta$ -helix (Govaerts et al., 2004) models of PrP<sup>Sc</sup>. Preliminary SMFS results of PrP trimers made from disulfide-linked monomers showed the formation of non-native structures via multiple intermediates (Fig. 10.1). Compared with the monomers and the dimers, the PrP trimers reflected a much more complex folding pathway, which was also implied from the constant force measurement (Fig. 10.1b). The total contour length change for unfolding the entire trimer structure was estimated to be 155 nm, corresponding to the unfolding of a stable structure containing ~430 structured amino acids, equal to the sum of all the amino acids contained within each of the monomers. Since the structured part of PrP<sup>C</sup> contains only 104 amino acids (James et al., 1997), this result implies that monomeric domains inside a trimer interact to form new structures incorporating the formerly unstructured part in PrP<sup>C</sup>. Such a structural outcome is predicted in several models of PrP<sup>Sc</sup> (Govaerts et al., 2004; Smirnovas et al., 2011), but how the structures formed by trimers in force spectroscopy measurements may be connected to pathogenesis, and in what way they are related to the structures formed in misfolded dimers, remains to be determined. Nevertheless, single-molecule studies of small oligomers evidently provide rich opportunities for probing structural conversion of PrP. Such methods can be further extended to study the structural properties of PrP amyloid, as demonstrated by recent work applying force spectroscopy in combination

with fluorescent imaging to map out inter- and intra-molecular interactions in yeast prion amyloid (Dong et al., 2010). However, the folding pathways of higher-order aggregates would be expected to become increasingly complex, as demonstrated in a set of FEC measurements of a possible PrP aggregate of larger size (Fig. 10.2). The broad shoulder feature of the FECs indicates the presence of a large number of intermediates under equilibrium during the unfolding of the oligomer, which are difficult to resolve from the measurement.



**Figure 10.2 FECs of a PrP aggregate showing the complex aggregation pathway.** A set of 44 FECs of aggregated H187R PrP overlaid (black) with WLC fits of the initial and the final state (dashed lines). The contour length change is 190 nm, corresponding to a size of a PrP tetramer or a larger oligomer. The broad shoulder-like feature at 5–20 pN indicates the interconversion of a large number of intermediates under equilibrium during unfolding.



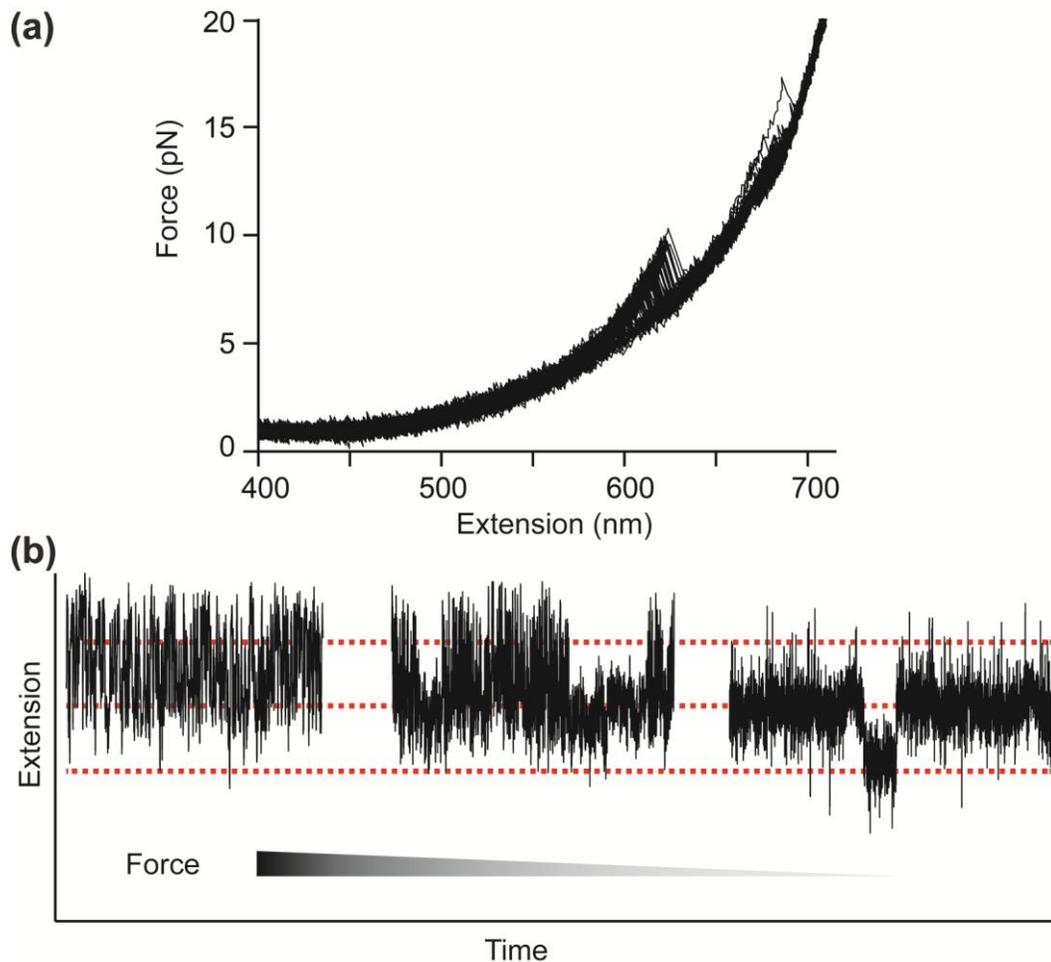
**Figure 10.3 SMFS of monomeric H187R mutant molecules.** (a) A set of 100 FECs of a H187R mutant (black) with WLC fits of the native and the unfolded state (dashed lines) shows a two-state folding behavior with a contour length change  $\sim 33$  nm, similar as the wild-type PrP. (b) Representative constant force trajectories of a H187R mutant confirm a two-state folding of the molecule. (c) No evidence of misfolded states which are present in the wild-type PrP has been found from the histogram analysis of the extension trajectories.

### 10.2.2 Disease-related mutations

More than 30 point mutations are related to prion diseases, and many of them have significant effects on the stability of PrP as well (van der Kamp and Daggett, 2009). Are the mutations going to cause intermediates to be populated, as deduced from previous experiments (Apetri et al., 2004)? How do the folding landscape and transition state change by introducing mutations into the protein? Can we relate these changes to the PrP<sup>Sc</sup> conversion process? Understanding how mutations alter the folding pathways and the misfolding behavior of PrP will help us to understand the misfolding-related disease mechanism. Although this thesis concentrated on wild-type PrP, we have begun to explore mutations by measuring the C179/214A variant which is aggregation-prone but not disease-related. As shown in chapter 7, the comparison of the mutant with the wild-type PrP provided important implication about how protein misfolding may be promoted by point mutations.

Studies of disease-related mutations of PrP have also been started, however, the increased propensity of the disease-related mutants to aggregate created many challenges for sample preparation. Two different mutants, F198S and H187R have been expressed and measured with optical tweezers. Protocols for attaching handles were modified to proceed in pH 4 sodium acetate buffer in order to minimize the aggregation of the proteins, however, the efficiency of the attachment reaction was greatly reduced under this condition. We were unsuccessful in measuring the monomeric F198S mutant, but three molecules of monomeric H187R mutant were measured, and found to fold into a structure with same contour length change as the native structure of PrP<sup>C</sup> (Fig. 10.3a). One H187R molecule was measured at constant force, showing no significant population of any of the three misfolded states observed in wild-type monomers (Fig. 10.3b and c),

but additional data would be required to make reliable conclusions about the effects of the mutation on the folding.



**Figure 10.4 SMFS of F198S dimers showing similar behavior to the wild-type dimers.** (a) Representative FECs of F198S dimers (black) show a similar behavior as the wild-type PrP dimers. (b) Representative constant force trajectories of F198S dimers.

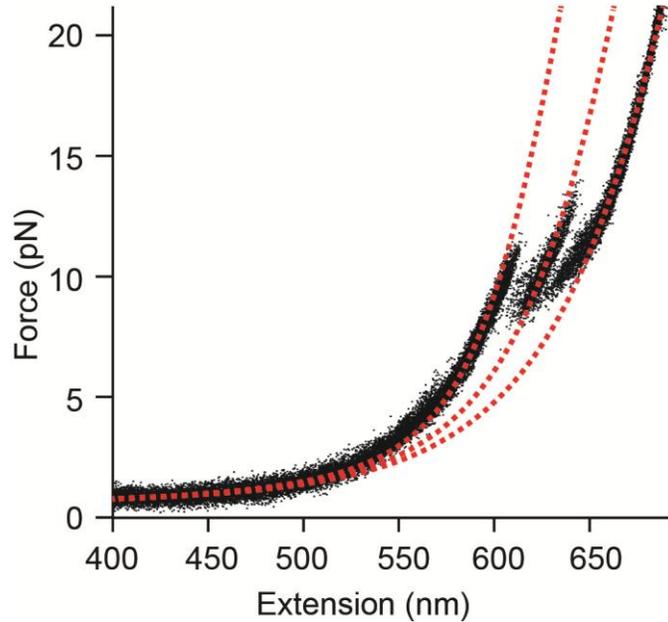
Given our observation of a stable misfolded dimer structure and the potential (although speculative) that it might play a role in pathogenic conversion, it would be interesting to probe the effects of disease-related mutations that significantly change the propensity of PrP<sup>Sc</sup> conversion on dimers, since presumably the mutations would affect the thermodynamics and/or kinetics of the dimer misfolding. Changes in the dimerization

process caused by pathogenic mutations would help provide evidence for the role of dimerization as an important initial step during the conversion. We made initial measurements on PrP dimers made from C179/214A, F198S and H187R monomers linked with disulfide bonds. The preliminary results showed largely the same behavior as for wild-type dimers: out of 3 F198S dimers and 3 H187R dimers (as well as 5 C179/214A dimers, although this is not a pathogenic mutation), all showed the same pattern of sequential unfolding through multiple intermediates as wild-type dimers, with similar unfolding forces (Fig. 10.4). One H187R dimer molecule unfolded differently (Fig. 10.5), however, indicating the possible presence of different structures of PrP dimers. Considerably more measurements on these mutations will be needed, as well as investigation of more types of mutations, before firm conclusions can be drawn. However, the strong aggregation propensity resulting from mutations will likely necessitate the development of a more efficient sample preparation method.

### **10.2.3 Structural prediction in combination with computer simulations**

Although single-molecule approaches have opened a new window on the misfolding of PrP, by themselves they are likely not sufficient to address the main issues that interest prion researchers. For example, even though our measurements show in detail that dimers form misfolded structures, we can't determine what this structure is from our results. However, measurements like these deliver a richly-detailed set of constraints that can be used to build or improve structural models of misfolded PrP as well as to guide models or simulations of the conversion process. In particular they provide distance and energy constraints for the possible structures (including intermediates), the pathways available (native and non-native), transition rates, even

information about the barriers. Integration of single-molecule experiments with computational modeling approaches should therefore prove a very fruitful source of future insight.



**Figure 10.5 FECs of H187R dimers showing a different behavior from the wild-type dimers.** A set of FECs of one H187R dimer molecule (black) shows two-step unfolding behavior. Dashed lines are WLC fits of the three states along the unfolding pathway. The contour length changes from the unfolded state are 36 nm and 75 nm, respectively.

To help test models, additional information could be obtained from single-molecule assays by introducing different reaction coordinates into the measurements. This can be done by pulling not just on the termini of the molecules, but also on different parts of the sequence, as shown in previous work with other proteins where handles were attached at various places in the structure (Bertz et al., 2008; Brockwell et al., 2003; Carrion-Vazquez et al., 2003; Dietz et al., 2006; Junker and Rief, 2010; Yu et al., 2012c). By using multiple different pulling axes, extra sets of constraints can be obtained to help build structural models (Dietz et al., 2006). However, under this approach each

measurement is still limited to a single reaction coordinate. To monitor more than one reaction coordinate at once, we could modify our optical tweezers system to add the ability to do FRET measurements as well. Combination of SMFS with FRET allows the folding reaction to be measured in two dimensions simultaneously: one reaction coordinate is the end-to-end extension of the molecule and the other is the distance between the two fluorophores. Such experiments would give flexibility in the choice of reaction coordinates and provide more robust constraints for the structure prediction procedures.

#### **10.2.4 Biological relevance of single-molecule results**

One of the challenges for single-molecule researchers is to connect their results *in vitro* to the full biological system and show they are relevant. In future work, we would like to explore the toxicity and infectivity of the misfolded conformations we have observed. To test the toxicity, we could use neuronal cell lines (Bedecs, 2008) or mouse models (Groschup and Buschmann, 2008) and assay for cell death, as done in previous work (Simoneau et al., 2007). To test infectivity, we might try several approaches. First, we could try applying the PMCA assay (Castilla et al., 2005; Saborio et al., 2001) to the misfolded oligomers, to see if they can act as templates for amplification. PMCA is an ultrasensitive method, with the ability in principle to detect as little as possibly a single PrP<sup>Sc</sup> molecule (Saa et al., 2006). An amplification of misfolded oligomers would provide support for their ability to convert native PrP<sup>C</sup>, so that the misfolded structure maintains a biological-relevant infectivity. The protease digestion pattern of aggregates made by PMCA as compared to the result for PrP<sup>Sc</sup> could also offer evidence for the oligomers

sharing similar structural characteristics with PrP<sup>Sc</sup> (Atarashi et al., 2007; Atarashi et al., 2008; Smirnovas et al., 2009).

Another approach to study infectivity would be to use a variant of our force spectroscopy assay. In order to probe the mechanism of the templated conformation change of PrP, which is believed as the basic mechanism underlying the propagation of prion disease, an experimental system could be developed by linking a dimer to a monomer using a much longer linker. By subjecting to force in optical tweezers, the dimer is kept physically separate from the monomer at a few pN tension, allowing the monomer folding into the native conformation while the dimer taking the misfolding pathway respectively, given our current result that PrP<sup>C</sup> refolds at ~ 10 pN and PrP dimer started to misfold at ~ 15 pN. As we bring the force down to zero, the misfolded dimer is brought into contact with the natively folded monomer to investigate any possible structural conversion of PrP<sup>C</sup> resulting from its interaction with the misfolded dimer. The hypothesis would be that the misfolded dimer is able to recruit and trigger the misfolding of PrP<sup>C</sup>, thereby strongly supporting the infectious nature of the misfolded dimer and providing a direct relationship with PrP<sup>Sc</sup>.

Finally, it would be interesting to explore the effects of anti-prion agents that help to clear up the disease. Several chemicals with anti-prion activity have been identified, such as tetrapyrroles (Caughey et al., 1998; Dee et al., 2012; Nicoll et al., 2010) and curcumin (Caughey et al., 1998), but their mechanism of action is not yet known. SMFS could be used to observe the effects of these compounds on the folding of single PrP monomers and oligomers, to provide insight into how they work, potentially leading to more effective anti-prion therapies. Single-molecule approaches have much to offer

protein folding/misfolding studies, and we foresee their increasing deployment to help solve the central scientific questions posed by prions.

## Bibliography

- Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. and Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460-465 (2005).
- Agarwal, P. K. Enzymes: An integrated view of structure, dynamics and function. *Microb Cell Fact* **5**, 2 (2006).
- Aguzzi, A., Heikenwalder, M. and Polymenidou, M. Insights into prion strains and neurotoxicity. *Nat Rev Mol Cell Biol* **8**, 552-561 (2007).
- Ainavarapu, S. R., Brujic, J., Huang, H. H., Wiita, A. P., Lu, H., Li, L., Walther, K. A., Carrion-Vazquez, M., Li, H. and Fernandez, J. M. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys J* **92**, 225-233 (2007).
- Al-Hashimi, H. M. and Walter, N. G. RNA dynamics: it is about time. *Curr Opin Struct Biol* **18**, 321-329 (2008).
- Anderson, M., Bocharova, O. V., Makarava, N., Breydo, L., Salnikov, V. V. and Baskakov, I. V. Polymorphism and ultrastructural organization of prion protein amyloid fibrils: an insight from high resolution atomic force microscopy. *J Mol Biol* **358**, 580-596 (2006).
- Andronesco, M., Aguirre-Hernandez, R., Condon, A. and Hoos, H. H. RNAsoft: a suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Res* **31**, 3416-3422 (2003).
- Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).
- Antonyuk, S. V., Trevitt, C. R., Strange, R. W., Jackson, G. S., Sangar, D., Batchelor, M., Cooper, S., Fraser, C., Jones, S., Georgiou, T., Khalili-Shirazi, A., Clarke, A. R., Hasnain, S. S. and Collinge, J. Crystal structure of human prion protein bound to a therapeutic antibody. *Proc Natl Acad Sci U S A* **106**, 2554-2558 (2009).
- Apetri, A. C., Maki, K., Roder, H. and Surewicz, W. K. Early intermediate in human prion protein folding as evidenced by ultrarapid mixing experiments. *J Am Chem Soc* **128**, 11673-11678 (2006).
- Apetri, A. C., Surewicz, K. and Surewicz, W. K. The effect of disease-associated mutations on the folding pathway of human prion protein. *J Biol Chem* **279**, 18008-18014 (2004).
- Apetri, A. C. and Surewicz, W. K. Kinetic intermediate in the folding of human prion protein. *J Biol Chem* **277**, 44589-44592 (2002).

- Araki, K. and Nagata, K. Protein folding and quality control in the ER. *Cold Spring Harb Perspect Biol* **4**, a015438 (2012).
- Ashkin, A. Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophys J* **61**, 569-582 (1992).
- Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E. and Chu, S. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt Lett* **11**, 288 (1986).
- Atarashi, R., Moore, R. A., Sim, V. L., Hughson, A. G., Dorward, D. W., Onwubiko, H. A., Priola, S. A. and Caughey, B. Ultrasensitive detection of scrapie prion protein using seeded conversion of recombinant prion protein. *Nat Methods* **4**, 645-650 (2007).
- Atarashi, R., Wilham, J. M., Christensen, L., Hughson, A. G., Moore, R. A., Johnson, L. M., Onwubiko, H. A., Priola, S. A. and Caughey, B. Simplified ultrasensitive prion detection by recombinant PrP conversion with shaking. *Nat Methods* **5**, 211-212 (2008).
- Baird, N. J. and Ferré-D'Amaré, A. R. Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis *RNA* **16**, 598-609 (2010).
- Baker, D. A surprising simplicity to protein folding. *Nature* **405**, 39-42 (2000).
- Baral, P. K., Wieland, B., Swayampakula, M., Polymenidou, M., Rahman, M. H., Kav, N. N., Aguzzi, A. and James, M. N. Structural studies on the folded domain of the human prion protein bound to the Fab fragment of the antibody POM1. *Acta Crystallogr D Biol Crystallogr* **68**, 1501-1512 (2012).
- Baskakov, I. V. Pathway complexity of prion protein assembly into amyloid. *J Biol Chem* **277**, 21140-21148 (2002).
- Baskakov, I. V., Legname, G., Prusiner, S. B. and Cohen, F. E. Folding of prion protein to its native alpha-helical conformation is under kinetic control. *J Biol Chem* **276**, 19687-19690 (2001).
- Batey, R. T., Gilbert, S. D. and Montange, R. K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411-415 (2004).
- Bechtluft, P., van Leeuwen, R. G., Tyreman, M., Tomkiewicz, D., Nouwen, N., Tepper, H. L., Driessen, A. J. and Tans, S. J. Direct observation of chaperone-induced changes in a protein folding pathway. *Science* **318**, 1458-1461 (2007).
- Bedecs, K. Cell culture models to unravel prion protein function and aberrancies in prion diseases. *Methods Mol Biol* **459**, 1-20 (2008).
- Bell, G. I. Models for the specific adhesion of cells to cells. *Science* **200**, 618-627 (1978).

- Benkovic, S. J. and Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301**, 1196-1202 (2003).
- Berkovich, R., Hermans, R. I., Popa, I., Stirnemann, G., Garcia-Manyes, S., Berne, B. J. and Fernandez, J. M. Rate limit of protein elastic response is tether dependent. *Proc Natl Acad Sci U S A* **109**, 14416-14421 (2012).
- Bertz, M., Kunfermann, A. and Rief, M. Navigating the folding energy landscape of green fluorescent protein. *Angew Chem Int Ed Engl* **47**, 8192-8195 (2008).
- Best, R. B. Atomistic molecular simulations of protein folding. *Curr Opin Struct Biol* **22**, 52-61 (2012).
- Best, R. B. and Hummer, G. Coordinate-dependent diffusion in protein folding. *Proc Natl Acad Sci U S A* **107**, 1088-1093 (2010).
- Best, R. B., Paci, E., Hummer, G. and Dudko, O. K. Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J Phys Chem B* **112**, 5968-5976 (2008).
- Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. and Kiefhaber, T. The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc Natl Acad Sci U S A* **96**, 9597-9601 (1999).
- Bieschke, J., Giese, A., Schulz-Schaeffer, W., Zerr, I., Poser, S., Eigen, M. and Kretzschmar, H. Ultrasensitive detection of pathological prion protein aggregates by dual-color scanning for intensely fluorescent targets. *Proc Natl Acad Sci U S A* **97**, 5468-5473 (2000).
- Birkmann, E., Schafer, O., Weinmann, N., Dumpitak, C., Beekes, M., Jackman, R., Thorne, L. and Riesner, D. Detection of prion particles in samples of BSE and scrapie by fluorescence correlation spectroscopy without proteinase K digestion. *Biol Chem* **387**, 95-102 (2006).
- Bjorndahl, T. C., Zhou, G. P., Liu, X., Perez-Pineiro, R., Semenchenko, V., Saleem, F., Acharya, S., Bujold, A., Sobsey, C. A. and Wishart, D. S. Detailed biophysical characterization of the acid-induced PrP(c) to PrP(beta) conversion process. *Biochemistry* **50**, 1162-1173 (2011).
- Borgia, A., Wensley, B. G., Soranno, A., Nettels, D., Borgia, M. B., Hoffmann, A., Pfeil, S. H., Lipman, E. A., Clarke, J. and Schuler, B. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat Commun* **3**, 1195 (2012).
- Borgia, A., Williams, P. M. and Clarke, J. Single-molecule studies of protein folding. *Annu Rev Biochem* **77**, 101-125 (2008).

- Borgia, M. B., Borgia, A., Best, R. B., Steward, A., Nettels, D., Wunderlich, B., Schuler, B. and Clarke, J. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* **474**, 662-665 (2011).
- Brockwell, D. J., Paci, E., Zinober, R. C., Beddard, G. S., Olmsted, P. D., Smith, D. A., Perham, R. N. and Radford, S. E. Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat Struct Biol* **10**, 731-737 (2003).
- Brockwell, D. J. and Radford, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* **17**, 30-37 (2007).
- Brundin, P., Melki, R. and Kopito, R. Prion-like transmission of protein aggregates in neurodegenerative diseases. *Nat Rev Mol Cell Biol* **11**, 301-307 (2010).
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. and Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195 (1995).
- Bryngelson, J. D. and Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* **84**, 7524-7528 (1987).
- Buchner, J. and Kiefhaber, T. *Protein Folding Handbook*. (Wiley-VCH, 2005).
- Buck, J., Fürtig, B., Noeske, J., Wöhnert, J. and Schwalbe, H. Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proc Natl Acad Sci U S A* **104**, 15699-16704 (2007).
- Buck, J., Noeske, J., Wöhnert, J. and Schwalbe, H. Dissecting the influence of Mg<sup>2+</sup> on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res* **38**, 4143-4153 (2010).
- Bustamante, C., Chemla, Y. R., Forde, N. R. and Izhaky, D. Mechanical processes in biochemistry. *Annu Rev Biochem* **73**, 705-748 (2004).
- Calzolari, L., Lysek, D. A., Perez, D. R., Guntert, P. and Wuthrich, K. Prion protein NMR structures of chickens, turtles, and frogs. *Proc Natl Acad Sci U S A* **102**, 651-655 (2005).
- Cao, S. and Chen, S. J. Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B* **115**, 4216-4226 (2011).
- Cao, Y. and Li, H. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nat Mater* **6**, 109-114 (2007).
- Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F. and Fernandez, J. M. The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* **10**, 738-743 (2003).

- Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J. and Fernandez, J. M. Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci U S A* **96**, 3694-3699 (1999).
- Castel, S. E. and Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* **14**, 100-112 (2013).
- Castilla, J., Saa, P., Hetz, C. and Soto, C. In vitro generation of infectious scrapie prions. *Cell* **121**, 195-206 (2005).
- Caughey, B., Baron, G. S., Chesebro, B. and Jeffrey, M. Getting a grip on prions: oligomers, amyloids, and pathological membrane interactions. *Annu Rev Biochem* **78**, 177-204 (2009).
- Caughey, B. and Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* **26**, 267-298 (2003).
- Caughey, B. and Raymond, G. J. The scrapie-associated form of PrP is made from a cell surface precursor that is both protease- and phospholipase-sensitive. *J Biol Chem* **266**, 18217-18223 (1991).
- Caughey, W. S., Raymond, L. D., Horiuchi, M. and Caughey, B. Inhibition of protease-resistant prion protein formation by porphyrins and phthalocyanines. *Proc Natl Acad Sci U S A* **95**, 12117-12122 (1998).
- Cecconi, C., Shank, E. A., Bustamante, C. and Marqusee, S. Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057-2060 (2005).
- Cecconi, C., Shank, E. A., Dahlquist, F. W., Marqusee, S. and Bustamante, C. Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J* **37**, 729-738 (2008).
- Chakroun, N., Prigent, S., Dreiss, C. A., Noinville, S., Chapuis, C., Fraternali, F. and Rezaei, H. The oligomerization properties of prion protein are restricted to the H2H3 domain. *FASEB J* **24**, 3222-3231 (2010).
- Chaudhury, S. and Makarov, D. E. A harmonic transition state approximation for the duration of reactive events in complex molecular rearrangements. *J Chem Phys* **133**, 034118 (2010).
- Chen, Y., Parrini, C., Taddei, N. and Lapidus, L. J. Conformational properties of unfolded HypF-N. *J Phys Chem B* **113**, 16209-16213 (2009).
- Chiti, F. and Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-366 (2006).

- Chiti, F. and Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* **5**, 15-22 (2009).
- Christen, B., Hornemann, S., Damberger, F. F. and Wuthrich, K. Prion protein NMR structure from tammar wallaby (*Macropus eugenii*) shows that the beta2-alpha2 loop is modulated by long-range sequence effects. *J Mol Biol* **389**, 833-845 (2009).
- Chung, H. S., Louis, J. M. and Eaton, W. A. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc Natl Acad Sci U S A* **106**, 11837-11844 (2009).
- Chung, H. S., McHale, K., Louis, J. M. and Eaton, W. A. Single-molecule fluorescence experiments determine protein folding transition path times. *Science* **335**, 981-984 (2012).
- Cobb, N. J., Sonnichsen, F. D., McHaourab, H. and Surewicz, W. K. Molecular architecture of human prion protein amyloid: a parallel, in-register beta-structure. *Proc Natl Acad Sci U S A* **104**, 18946-18951 (2007).
- Cohen, F. E., Pan, K. M., Huang, Z., Baldwin, M., Fletterick, R. J. and Prusiner, S. B. Structural clues to prion replication. *Science* **264**, 530-531 (1994).
- Cohen, F. E. and Prusiner, S. B. Pathologic conformations of prion proteins. *Annu Rev Biochem* **67**, 793-819 (1998).
- Colby, D. W. and Prusiner, S. B. Prions. *Cold Spring Harb Perspect Biol* **3**, a006833 (2011).
- Collin, D., Ritort, F., Jarzynski, C., Smith, B. C., Tinoco, I. and Bustamante, C. Verification of the crooks fluctuations theorem and recovery of RNA folding free energies. *Nature* **437**, 231-234 (2005).
- Come, J. H., Fraser, P. E. and Lansbury, P. T., Jr. A kinetic model for amyloid formation in the prion diseases: importance of seeding. *Proc Natl Acad Sci U S A* **90**, 5959-5963 (1993).
- Coppins, R. L., Hall, K. B. and Groisman, E. A. The intricate world of riboswitches. *Curr Opin Microbiol* **10**, 176-181 (2007).
- Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **60**, 2721-2726 (1999).
- Davies, P. and Brown, D. R. The chemistry of copper binding to PrP: is there sufficient evidence to elucidate a role for copper in protein function? *Biochem J* **410**, 237-244 (2008).

- De Simone, A., Zagari, A. and Derreumaux, P. Structural and hydration properties of the partially unfolded states of the prion protein. *Biophys J* **93**, 1284-1292 (2007).
- Dee, D. R., Gupta, A. N., Anikovskiy, M., Sosova, I., Grandi, E., Rivera, L., Vincent, A., Brigley, A. M., Petersen, N. O. and Woodside, M. T. Phthalocyanine tetrasulfonates bind to multiple sites on natively-folded prion protein. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1824**, 826-832 (2012).
- DeMarco, M. L. and Daggett, V. From conversion to aggregation: protofibril formation of the prion protein. *Proc Natl Acad Sci U S A* **101**, 2293-2298 (2004).
- Dietz, H., Berkemeier, F., Bertz, M. and Rief, M. Anisotropic deformation response of single protein molecules. *Proc Natl Acad Sci U S A* **103**, 12724-12728 (2006).
- Dietz, H. and Rief, M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc Natl Acad Sci U S A* **101**, 16192-16197 (2004).
- Dill, K. A. and Chan, H. S. From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-19 (1997).
- Dong, J., Castro, C. E., Boyce, M. C., Lang, M. J. and Lindquist, S. Optical trapping with high forces reveals unexpected behaviors of prion fibrils. *Nat Struct Mol Biol* **17**, 1422-1430 (2010).
- Dudko, O. K., Graham, T. G. W. and Best, R. B. Locating the Barrier for Folding of Single Molecules under an External Force. *Phys Rev Lett* **107**, 208301 (2011).
- Dudko, O. K., Hummer, G. and Szabo, A. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys Rev Lett* **96**, 108101 (2006).
- Dudko, O. K., Hummer, G. and Szabo, A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc Natl Acad Sci U S A* **105**, 15755-15760 (2008).
- Efron, B. Bootstrap methods: another look at the jackknife. *Ann Statist* **7**, 1-26 (1979).
- Eghiaian, F., Grosclaude, J., Lesceu, S., Debey, P., Doublet, B., Treguer, E., Rezaei, H. and Knossow, M. Insight into the PrPC-->PrPSc conversion from the structures of antibody-bound ovine prion scrapie-susceptibility variants. *Proc Natl Acad Sci U S A* **101**, 10254-10259 (2004).
- Eichner, T. and Radford, S. E. A diversity of assembly mechanisms of a generic amyloid fold. *Mol Cell* **43**, 8-18 (2011).
- Elms, P. J., Chodera, J. D., Bustamante, C. and Marqusee, S. The molten globule state is unusually deformable under mechanical force. *Proc Natl Acad Sci U S A* **109**, 3796-3801 (2012).

- Evans, E. and Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys J* **72**, 1541-1555 (1997).
- Fabian, H. and Naumann, D. Methods to study protein folding by stopped-flow FT-IR. *Methods* **34**, 28-40 (2004).
- Faver, J. C., Benson, M. L., He, X., Roberts, B. P., Wang, B., Marshall, M. S., Sherrill, C. D. and Merz, K. M., Jr. The energy computation paradox and ab initio protein folding. *PLoS One* **6**, e18868 (2011).
- Fischer, M., Rulicke, T., Raeber, A., Sailer, A., Moser, M., Oesch, B., Brandner, S., Aguzzi, A. and Weissmann, C. Prion protein (PrP) with amino-proximal deletions restoring susceptibility of PrP knockout mice to scrapie. *EMBO J* **15**, 1255-1264 (1996).
- Fitzpatrick, J. A. and Lillemeier, B. F. Fluorescence correlation spectroscopy: linking molecular dynamics to biological function in vitro and in situ. *Curr Opin Struct Biol* **21**, 650-660 (2011).
- Foster, D. A. High resolution optical tweezers for single molecule studies of hierarchical folding in the *pbuE* riboswitch aptamer. Master of Science thesis, University of Alberta, Edmonton (2010).
- Frankenfield, K. N., Powers, E. T. and Kelly, J. W. Influence of the N-terminal domain on the aggregation properties of the prion protein. *Protein Sci* **14**, 2154-2166 (2005).
- Frauenfelder, H., Fenimore, P. W. and Young, R. D. Protein dynamics and function: Insights from the energy landscape and solvent slaving. *IUBMB Life* **59**, 506-512 (2007).
- Frost, B. and Diamond, M. I. Prion-like mechanisms in neurodegenerative diseases. *Nat Rev Neurosci* **11**, 155-159 (2010).
- Fujii, F., Horiuchi, M., Ueno, M., Sakata, H., Nagao, I., Tamura, M. and Kinjo, M. Detection of prion protein immune complex for bovine spongiform encephalopathy diagnosis using fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy. *Anal Biochem* **370**, 131-141 (2007).
- Fung, S. Y., Keyes, C., Duhamel, J. and Chen, P. Concentration effect on the aggregation of a self-assembling oligopeptide. *Biophys J* **85**, 537-548 (2003).
- Gallo, M., Paludi, D., Cicero, D. O., Chiovitti, K., Millo, E., Salis, A., Damonte, G., Corsaro, A., Thellung, S., Schettini, G., Melino, S., Florio, T., Paci, M. and Aceto, A. Identification of a conserved N-capping box important for the structural autonomy of the prion alpha 3-helix: the disease associated D202N mutation destabilizes the helical conformation. *Int J Immunopathol Pharmacol* **18**, 95-112 (2005).

- Ganchev, D. N., Cobb, N. J., Surewicz, K. and Surewicz, W. K. Nanomechanical properties of human prion protein amyloid as probed by force spectroscopy. *Biophys J* **95**, 2909-2915 (2008).
- Gebhardt, J. C., Bornschlogl, T. and Rief, M. Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci U S A* **107**, 2013-2018 (2010).
- Gerber, R., Tahiri-Alaoui, A., Hore, P. J. and James, W. Oligomerization of the human prion protein proceeds via a molten globule intermediate. *J Biol Chem* **282**, 6300-6307 (2007).
- Gerber, R., Tahiri-Alaoui, A., Hore, P. J. and James, W. Conformational pH dependence of intermediate states during oligomerization of the human prion protein. *Protein Sci* **17**, 537-544 (2008).
- Gerum, C., Silvers, R., Wirmer-Bartoschek, J. and Schwalbe, H. Unfolded-state structure and dynamics influence the fibril formation of human prion protein. *Angew Chem Int Ed Engl* **121**, 9616-9620 (2009).
- Gilbert, S. D., Stoddard, C. D., Wise, S. J. and Batey, R. T. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol* **359**, 754-768 (2006).
- Goedert, M., Clavaguera, F. and Tolnay, M. The propagation of prion-like protein inclusions in neurodegenerative diseases. *Trends Neurosci* **33**, 317-325 (2010).
- Gong, B., Ramos, A., Vazquez-Fernandez, E., Silva, C. J., Alonso, J., Liu, Z. and Requena, J. R. Probing structural differences between PrP(C) and PrP(Sc) by surface nitration and acetylation: evidence of conformational change in the C-terminus. *Biochemistry* **50**, 4963-4972 (2011).
- Gore, J., Ritort, F. and Bustamante, C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc Natl Acad Sci U S A* **100**, 12564-12569 (2003).
- Gossert, A. D., Bonjour, S., Lysek, D. A., Fiorito, F. and Wuthrich, K. Prion protein NMR structures of elk and of mouse/elk hybrids. *Proc Natl Acad Sci U S A* **102**, 646-650 (2005).
- Govaerts, C., Wille, H., Prusiner, S. B. and Cohen, F. E. Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci U S A* **101**, 8342-8347 (2004).
- Grabenaus, M., Wytttenbach, T., Sanghera, N., Slade, S. E., Pinheiro, T. J., Scrivens, J. H. and Bowers, M. T. Conformational stability of Syrian hamster prion protein PrP(90-231). *J Am Chem Soc* **132**, 8816-8818 (2010).

- Grad, L. I., Guest, W. C., Yanai, A., Pokrishevsky, E., O'Neill, M. A., Gibbs, E., Semenchenko, V., Yousefi, M., Wishart, D. S., Plotkin, S. S. and Cashman, N. R. Intermolecular transmission of superoxide dismutase 1 misfolding in living cells. *Proc Natl Acad Sci U S A* **108**, 16398-16403 (2011).
- Greenleaf, W. J., Frieda, K. L., Foster, D. A., Woodside, M. T. and Block, S. M. Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **319**, 630-633 (2008).
- Greenleaf, W. J., Woodside, M. T., Abbondanzieri, E. A. and Block, S. M. Passive all-optical force clamp for high-resolution laser trapping. *Phys Rev Lett* **95**, 208102 (2005).
- Greenleaf, W. J., Woodside, M. T. and Block, S. M. High-resolution, single-molecule measurements of biomolecular motion. *Annu Rev Biophys Biomol Struct* **36**, 171-190 (2007).
- Griffith, J. S. Self-replication and scrapie. *Nature* **215**, 1043-1044 (1967).
- Groschup, M. H. and Buschmann, A. Rodent models for prion diseases. *Vet Res* **39**, 32 (2008).
- Gross, P., Laurens, N., Oddershede, L. B., Bockelmann, U., Peterman, E. J. G. and Wuite, G. J. L. Quantifying how DNA stretches, melts and changes twist under tension. *Nat Phys* **7**, 731-736 (2011).
- Gupta, A. N., Vincent, A., Neupane, K., Yu, H., Wang, F. and Woodside, M. T. Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat Phys* **7**, 631-634 (2011).
- Hagen, S. J., Hofrichter, J., Szabo, A. and Eaton, W. A. Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc Natl Acad Sci U S A* **93**, 11615-11617 (1996).
- Hall, K. B. RNA in motion. *Curr Opin Chem Biol* **12**, 612-618 (2008).
- Hanggi, P., Talkner, P. and Borkovec, M. Reaction-rate theory - 50 years after Kramers. *Rev Mod Phys* **62**, 251-341 (1990).
- Harada, Y. and Asakura, T. Radiation forces on a dielectric sphere in the Rayleigh scattering regime. *Opt Commun* **124**, 529-541 (1996).
- Harris, N. C., Song, Y. and Kiang, C.-H. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality. *Phys Rev Lett* **99**, 068101 (2007).

- Hart, T., Hosszu, L. L., Trevitt, C. R., Jackson, G. S., Waltho, J. P., Collinge, J. and Clarke, A. R. Folding kinetics of the human prion protein probed by temperature jump. *Proc Natl Acad Sci U S A* **106**, 5651-5656 (2009).
- Hartl, F. U. and Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16**, 574-581 (2009).
- Henzler-Wildman, K. and Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972 (2007).
- Hillger, F., Nettels, D., Dorsch, S. and Schuler, B. Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *J Fluoresc* **17**, 759-765 (2007).
- Hinczewski, M., von Hansen, Y. and Netz, R. R. Deconvolution of dynamic mechanical networks. *Proc Natl Acad Sci U S A* **107**, 21493-21498 (2010).
- Hoffmann, A., Neupane, K. and Woodside, M. T. Single-molecule assays for investigating protein misfolding and aggregation. *Phys Chem Chem Phys* (2013).
- Hofrichter, J. Laser temperature-jump methods for studying folding dynamics. *Methods Mol Biol* **168**, 159-191 (2001).
- Hohng, S., Zhou, R., Nahas, M. K., Yu, J., Schulten, K., Lilley, D. M. and Ha, T. Fluorescence-force spectroscopy maps two-dimensional reaction landscape of the holliday junction. *Science* **318**, 279-283 (2007).
- Hornemann, S. and Glockshuber, R. A scrapie-like unfolding intermediate of the prion protein domain PrP(121-231) induced by acidic pH. *Proc Natl Acad Sci U S A* **95**, 6010-6014 (1998).
- Hosszu, L. L., Baxter, N. J., Jackson, G. S., Power, A., Clarke, A. R., Waltho, J. P., Craven, C. J. and Collinge, J. Structural mobility of the human prion protein probed by backbone hydrogen exchange. *Nat Struct Biol* **6**, 740-743 (1999).
- Hosszu, L. L., Wells, M. A., Jackson, G. S., Jones, S., Batchelor, M., Clarke, A. R., Craven, C. J., Waltho, J. P. and Collinge, J. Definable equilibrium states in the folding of human prion protein. *Biochemistry* **44**, 16649-16657 (2005).
- Hugel, T., Michaelis, J., Hetherington, C. L., Jardine, P. J., Grimes, S., Walter, J. M., Falk, W., Anderson, D. L. and Bustamante, C. Experimental test of connector rotation during DNA packaging into bacteriophage phi29 capsids. *PLoS Biol* **5**, e59 (2007).
- Hummer, G. From transition paths to transition states and rate coefficients. *J Chem Phys* **120**, 516-523 (2004).
- Hummer, G. and Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc Natl Acad Sci U S A* **98**, 3658-3661 (2001).

- Hummer, G. and Szabo, A. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys J* **85**, 5-15 (2003).
- Hüttenhofer, A. and Noller, H. F. Footprinting mRNA-ribosome complexes with chemical probes. *EMBO J* **13**, 3892-3901 (1994).
- Ivankov, D. N., Garbuzynskiy, S. O., Alm, E., Plaxco, K. W., Baker, D. and Finkelstein, A. V. Contact order revisited: influence of protein size on the folding rate. *Protein Sci* **12**, 2057-2062 (2003).
- Jackson, S. E. How do small single-domain proteins fold? *Fold Des* **3**, R81-91 (1998).
- Jain, S. and Udgaonkar, J. B. Evidence for stepwise formation of amyloid fibrils by the mouse prion protein. *J Mol Biol* **382**, 1228-1241 (2008).
- James, T. L., Liu, H., Ulyanov, N. B., Farr-Jones, S., Zhang, H., Donne, D. G., Kaneko, K., Groth, D., Mehlhorn, I., Prusiner, S. B. and Cohen, F. E. Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform. *Proc Natl Acad Sci U S A* **94**, 10086-10091 (1997).
- Jansen, K., Schäfer, O., Birkmann, E., Post, K., Serban, H., Prusiner, S. B. and Riesner, D. Structural intermediates in the putative pathway from the cellular prion protein to the pathogenic form. *Biol Chem* **382**, 683-691 (2001).
- Jansson, P. A. *Deconvolution of images and spectra*. 2 edn, (Academic Press, New York, 1997).
- Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys Rev Lett* **78**, 2690-2693 (1997).
- Jenkins, D. C., Pearson, D. S., Harvey, A., Sylvester, I. D., Geeves, M. A. and Pinheiro, T. J. Rapid folding of the prion protein captured by pressure-jump. *Eur Biophys J* **38**, 625-635 (2009).
- Jenkins, D. C., Sylvester, I. D. and Pinheiro, T. J. T. The elusive intermediate on the folding pathway of the prion protein. *FEBS Journal* **275**, 1323-1335 (2008).
- Jetha, N. N., Semenchenko, V., Wishart, D. S., Cashman, N. R. and Marziali, A. Nanopore analysis of wild-type and mutant prion protein (PrP(C)): single molecule discrimination and PrP(C) kinetics. *PLoS One* **8**, e54982 (2013).
- Jollymore, A., Lethias, C., Peng, Q., Cao, Y. and Li, H. Nanomechanical properties of tenascin-X revealed by single-molecule force spectroscopy. *J Mol Biol* **385**, 1277-1286 (2009).
- Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. and Ha, T. Advances in single-molecule fluorescence methods for molecular biology. *Annu Rev Biochem* **77**, 51-76 (2008).

- Julien, O., Chatterjee, S., Thiessen, A., Graether, S. P. and Sykes, B. D. Differential stability of the bovine prion protein upon urea unfolding. *Protein Sci* **18**, 2172-2182 (2009).
- Junker, J. P. and Rief, M. Evidence for a broad transition-state ensemble in calmodulin folding from single-molecule force spectroscopy. *Angew Chem Int Ed Engl* **49**, 3306-3309 (2010).
- Junker, J. P., Ziegler, F. and Rief, M. Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science* **323**, 633-637 (2009).
- Kaimann, T., Metzger, S., Kuhlmann, K., Brandt, B., Birkmann, E., Holtje, H. D. and Riesner, D. Molecular model of an alpha-helical prion protein dimer and its monomeric subunits as derived from chemical cross-linking and molecular modeling calculations. *J Mol Biol* **376**, 582-596 (2008).
- Khan, M. Q., Sweeting, B., Mulligan, V. K., Arslan, P. E., Cashman, N. R., Pai, E. F. and Chakrabarty, A. Prion disease susceptibility is affected by  $\beta$ -structure folding propensity and local side-chain interactions in PrP. *Proc Natl Acad Sci U S A* **107**, 19808-19813 (2010).
- Kim, B. H., Palermo, N. Y., Lovas, S., Zaikova, T., Keana, J. F. and Lyubchenko, Y. L. Single-molecule atomic force microscopy force spectroscopy study of A $\beta$ -40 interactions. *Biochemistry* **50**, 5154-5162 (2011).
- Kim, J. N. and Breaker, R. R. Purine sensing by riboswitches. *Biol Cell* **100**, 1-11 (2008).
- Kim, P. S. and Baldwin, R. L. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* **59**, 631-660 (1990).
- Knaus, K. J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W. K. and Yee, V. C. Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat Struct Biol* **8**, 770-774 (2001).
- Konermann, L., Pan, J. and Liu, Y. H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev* **40**, 1224-1234 (2011).
- Koonin, E. V., Wolf, Y. I. and Karev, G. P. The structure of the protein universe and genome evolution. *Nature* **420**, 218-223 (2002).
- Kosuri, P., Alegre-Cebollada, J., Feng, J., Kaplan, A., Ingles-Prieto, A., Badilla, C. L., Stockwell, B. R., Sanchez-Ruiz, J. M., Holmgren, A. and Fernandez, J. M. Protein folding drives disulfide formation. *Cell* **151**, 794-806 (2012).
- Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284-304 (1940).

- Krasnoslobodtsev, A. V., Peng, J., Asiago, J. M., Hindupur, J., Rochet, J. C. and Lyubchenko, Y. L. Effect of spermidine on misfolding and interactions of alpha-synuclein. *PLoS One* **7**, e38099 (2012).
- Krieger, F., Fierz, B., Bieri, O., Drewello, M. and Kiefhaber, T. Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J Mol Biol* **332**, 265-274 (2003).
- Kubelka, J., Hofrichter, J. and Eaton, W. A. The protein folding 'speed limit'. *Curr Opin Struct Biol* **14**, 76-88 (2004).
- Kubota, H. Quality control against misfolded proteins in the cytosol: a network for cell survival. *J Biochem* **146**, 609-616 (2009).
- Kuwata, K., Li, H., Yamada, H., Legname, G., Prusiner, S. B., Akasaka, K. and James, T. L. Locally disordered conformer of the hamster prion protein: a crucial intermediate to PrPSc? *Biochemistry* **41**, 12277-12283 (2002).
- Kuwata, K., Matumoto, T., Cheng, H., Nagayama, K., James, T. L. and Roder, H. NMR-detected hydrogen exchange and molecular dynamics simulations provide structural insight into fibril formation of prion protein fragment 106-126. *Proc Natl Acad Sci U S A* **100**, 14790-14795 (2003).
- Kuznetsov, I. B. and Rackovsky, S. On the properties and sequence context of structurally ambivalent fragments in proteins. *Protein Sci* **12**, 2420-2433 (2003).
- Kuznetsov, I. B. and Rackovsky, S. Comparative computational analysis of prion proteins reveals two fragments with unusual structural properties and a pattern of increase in hydrophobicity associated with disease-promoting mutations. *Protein Sci* **13**, 3230-3244 (2004).
- Lansbury, P. T., Jr. and Caughey, B. The chemistry of scrapie infection: implications of the 'ice 9' metaphor. *Chem Biol* **2**, 1-5 (1995).
- Lapidus, L. J., Eaton, W. A. and Hofrichter, J. Measuring the rate of intramolecular contact formation in polypeptides. *Proc Natl Acad Sci U S A* **97**, 7220-7225 (2000).
- Lawson, V. A., Collins, S. J., Masters, C. L. and Hill, A. F. Prion protein glycosylation. *J Neurochem* **93**, 793-801 (2005).
- Lee, G., Abdi, K., Jiang, Y., Michaely, P., Bennett, V. and Marszalek, P. E. Nanospring behaviour of ankyrin repeats. *Nature* **440**, 246-249 (2006).
- Lee, S. and Eisenberg, D. Seeded conversion of recombinant prion protein to a disulfide-bonded oligomer by a reduction-oxidation process. *Nat Struct Biol* **10**, 725-730 (2003).

- Lee, T. H., Lapidus, L. J., Zhao, W., Travers, K. J., Herschlag, D. and Chu, S. Measuring the folding transition time of single RNA molecules. *Biophys J* **92**, 3275-3283 (2007).
- Lemay, J. F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Massé, E. and Lafontaine, D. A. Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet* **7**, e1001278 (2011).
- Lemay, J. F., Penedo, J. C., Tremblay, R., Lilley, D. M. and Lafontaine, D. A. Folding of the adenine riboswitch. *Chem Biol* **13**, 857-868 (2006).
- Levinthal, C. Are there pathways for protein folding. *J Chim Phys Phys-Chim Biol* **65**, 44-& (1968).
- Levinthal, C. *Mossbauer spectroscopy in biological systems: proceedings of a meeting held at Allerton House, Monticello, Illinois*. 22-24 (University of Illinois Press, 1969).
- Li, M. S., Klimov, D. K. and Thirumalai, D. Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer* **45**, 573-579 (2004).
- Li, P. T., Viereggs, J. and Tinoco, I., Jr. How RNA unfolds and refolds. *Annu Rev Biochem* **77**, 77-100 (2008).
- Li, P. T. X., Bustamante, C. and Tinoco Jr, I. Real-time control of the energy landscape by force directs the folding of RNA molecules. *Proc Natl Acad Sci U S A* **104**, 7039-7044 (2007).
- Liemann, S. and Glockshuber, R. Influence of amino acid substitutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein. *Biochemistry* **38**, 3258-3267 (1999).
- Lin, J.-C. and Thirumalai, D. Relative stability of helices determines the folding landscape of adenine riboswitch aptamers. *J Am Chem Soc* **130**, 14080-14081 (2008).
- Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I. J. and Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* **296**, 1832-1835 (2002).
- Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. J. and Bustamante, C. Reversible unfolding of single RNA molecules by mechanical force. *Science* **292**, 733-737 (2001).
- Lopez Garcia, F., Zahn, R., Riek, R. and Wuthrich, K. NMR structure of the bovine prion protein. *Proc Natl Acad Sci U S A* **97**, 8334-8339 (2000).

- Lorimer, G. H. A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. *FASEB J* **10**, 5-9 (1996).
- Lysek, D. A., Schorn, C., Nivon, L. G., Esteve-Moya, V., Christen, B., Calzolari, L., von Schroetter, C., Fiorito, F., Herrmann, T., Guntert, P. and Wuthrich, K. Prion protein NMR structures of cats, dogs, pigs, and sheep. *Proc Natl Acad Sci U S A* **102**, 640-645 (2005).
- Ma, J., Wollmann, R. and Lindquist, S. Neurotoxicity and neurodegeneration when PrP accumulates in the cytosol. *Science* **298**, 1781-1785 (2002).
- Maiti, N. R. and Surewicz, W. K. The role of disulfide bridge in the folding and stability of the recombinant human prion protein. *J Biol Chem* **276**, 2427-2431 (2001).
- Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. and Breaker, R. R. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**, 577-586 (2003).
- Mandal, M. and Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**, 29-35 (2004).
- Marko, J. F. and Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759-8770 (1995).
- Martins, S. M., Chapeaurouge, A. and Ferreira, S. T. Folding intermediates of the prion protein stabilized by hydrostatic pressure and low temperature. *J Biol Chem* **278**, 50449-50455 (2003).
- Matouschek, A. Protein unfolding--an important process in vivo? *Curr Opin Struct Biol* **13**, 98-109 (2003).
- Matouschek, A. and Bustamante, C. Finding a protein's Achilles heel. *Nat Struct Biol* **10**, 674-676 (2003).
- McKinley, M. P., Meyer, R. K., Kenaga, L., Rahbar, F., Cotter, R., Serban, A. and Prusiner, S. B. Scrapie prion rod formation in vitro requires both detergent extraction and limited proteolysis. *J Virol* **65**, 1340-1351 (1991).
- Mead, S. Prion disease genetics. *Eur J Hum Genet* **14**, 273-281 (2006).
- Meyer, R. K., Lustig, A., Oesch, B., Fatzer, R., Zurbriggen, A. and Vandevelde, M. A monomer-dimer equilibrium of a cellular prion protein (PrPC) not observed with recombinant PrP. *J Biol Chem* **275**, 38081-38087 (2000).
- Minh, D. D. Free-energy reconstruction from experiments performed under different biasing programs. *Phys Rev E Stat Nonlin Soft Matter Phys* **74**, 061120 (2006).
- Minor, D. L., Jr. and Kim, P. S. Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**, 730-734 (1996).

- Moffitt, J. R., Chemla, Y. R., Smith, S. B. and Bustamante, C. Recent advances in optical tweezers. *Annu Rev Biochem* **77**, 205-228 (2008).
- Mossa, A., de Lorenzo, S., Huguet, J. M. and Ritort, F. Measurement of work in single-molecule pulling experiments. *J Chem Phys* **130**, 234116 (2009).
- Moult, J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol* **15**, 285-289 (2005).
- Mukhopadhyay, S., Krishnan, R., Lemke, E. A., Lindquist, S. and Deniz, A. A. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc Natl Acad Sci U S A* **104**, 2649-2654 (2007).
- Munch, C., O'Brien, J. and Bertolotti, A. Prion-like propagation of mutant superoxide dismutase-1 misfolding in neuronal cells. *Proc Natl Acad Sci U S A* **108**, 3548-3553 (2011).
- Nath, S., Meuvlis, J., Hendrix, J., Carl, S. A. and Engelborghs, Y. Early aggregation steps in alpha-synuclein as measured by FCS and FRET: evidence for a contagious conformational change. *Biophys J* **98**, 1302-1311 (2010).
- Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riek, C., Grothe, R. and Eisenberg, D. Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435**, 773-778 (2005).
- Nettels, D., Gopich, I. V., Hoffmann, A. and Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci U S A* **104**, 2655-2660 (2007).
- Neuman, K. C. and Block, S. M. Optical trapping. *Rev Sci Instrum* **75**, 2787-2809 (2004).
- Neuman, K. C. and Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods* **5**, 491-505 (2008).
- Neupane, K., Ritchie, D. B., Yu, H., Foster, D. A., Wang, F. and Woodside, M. T. Transition path times for nucleic acid folding determined from energy-landscape analysis of single-molecule trajectories. *Phys Rev Lett* **109**, 068102 (2012).
- Neupane, K., Solanki, A., Sosova, I., Belov, M. and Woodside, M. T. Simultaneous characterization of structure, stability, and rates in single oligomeric aggregates of  $\alpha$ -synuclein using force spectroscopy (submitted).
- Neupane, K., Yu, H., Foster, D. A., Wang, F. and Woodside, M. T. Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res* **39**, 7677-7687 (2011).
- Nguyen, J. T., Inouye, H., Baldwin, M. A., Fletterick, R. J., Cohen, F. E., Prusiner, S. B. and Kirschner, D. A. X-ray diffraction of scrapie prion rods and PrP peptides. *J Mol Biol* **252**, 412-422 (1995).

- Nicoll, A. J., Trevitt, C. R., Tattum, M. H., Risse, E., Quarterman, E., Ibarra, A. A., Wright, C., Jackson, G. S., Sessions, R. B., Farrow, M., Waltho, J. P., Clarke, A. R. and Collinge, J. Pharmacological chaperone for the structured domain of human prion protein. *Proc Natl Acad Sci U S A* **107**, 17610-17615 (2010).
- Noeske, J., Buck, J., Fürtig, B., Nasiri, H. R., Schwalbe, H. and Wöhnert, J. Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res* **35**, 572-583 (2007).
- Noeske, J., Richter, C., Grundl, M. A., Nasiri, H. R., Schwalbe, H. and Wöhnert, J. An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. *Proc Natl Acad Sci U S A* **102**, 1372-1377 (2005).
- Nojima, T., Lin, A. C., Fujii, T. and Endo, I. Determination of the termination efficiency of the transcription terminator using different fluorescent profiles in green fluorescent protein mutants. *Anal Sci* **21**, 1479-1481 (2005).
- O'Sullivan, D. B., Jones, C. E., Abdelraheim, S. R., Thompsett, A. R., Brazier, M. W., Toms, H., Brown, D. R. and Viles, J. H. NMR characterization of the pH 4 beta-intermediate of the prion protein: the N-terminal half of the protein remains unstructured and retains a high degree of flexibility. *Biochem J* **401**, 533-540 (2007).
- Oberhauser, A. F., Marszalek, P. E., Carrion-Vazquez, M. and Fernandez, J. M. Single protein misfolding events captured by atomic force microscopy. *Nat Struct Biol* **6**, 1025-1028 (1999).
- Oesch, B., Westaway, D., Walchli, M., McKinley, M. P., Kent, S. B., Aebersold, R., Barry, R. A., Tempst, P., Teplow, D. B., Hood, L. E. and et al. A cellular gene encodes scrapie PrP 27-30 protein. *Cell* **40**, 735-746 (1985).
- Oliveberg, M. and Wolynes, P. G. The experimental survey of protein-folding energy landscapes. *Q Rev Biophys* **38**, 245-288 (2005).
- Oluwole, B. O., Du, W., Mills, I. and Sumpio, B. E. Gene regulation by mechanical forces. *Endothelium* **5**, 85-93 (1997).
- Onuchic, J. N., Luthey-Schulten, Z. and Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **48**, 545-600 (1997).
- Onuchic, J. N. and Wolynes, P. G. Theory of protein folding. *Curr Opin Struct Biol* **14**, 70-75 (2004).
- Orte, A., Birkett, N. R., Clarke, R. W., Devlin, G. L., Dobson, C. M. and Klenerman, D. Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc Natl Acad Sci U S A* **105**, 14424-14429 (2008).

- Ottink, O. M., Rampersad, S. M., Tessari, M., Zaman, G. J. R., Heus, H. A. and Wijmenga, S. S. Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined–induced fit mechanism *RNA* **13**, 2202-2212 (2007).
- Paik, D. H. and Perkins, T. T. Overstretching DNA at 65 pN does not require peeling from free ends or nicks. *J Am Chem Soc* **133**, 3219-3221 (2011).
- Palassini, M. and Ritort, F. Improving free-energy estimates from unidirectional work measurements: theory and experiment. *Phys Rev Lett* **107**, 060601 (2011).
- Pan, K. M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R. J., Cohen, F. E. and et al. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci U S A* **90**, 10962-10966 (1993).
- Pang, Y. and Gordon, R. Optical trapping of a single protein. *Nano Lett* **12**, 402-406 (2012).
- Park, S. and Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J Chem Phys* **120**, 5946-5961 (2004).
- Pathria, R. K. *Statistical mechanics*. (Elsevier Science, 1996).
- Pauling, L. and Corey, R. B. The pleated sheet, a new layer configuration of polypeptide chains. *Proc Natl Acad Sci U S A* **37**, 251-256 (1951).
- Pechmann, S., Willmund, F. and Frydman, J. The ribosome as a hub for protein quality control. *Mol Cell* **49**, 411-421 (2013).
- Perez, D. R., Damberger, F. F. and Wuthrich, K. Horse prion protein NMR structure and comparisons with related variants of the mouse prion protein. *J Mol Biol* **400**, 121-128 (2010).
- Piana, S. and Laio, A. A bias-exchange approach to protein folding. *J Phys Chem B* **111**, 4553-4559 (2007).
- Piana, S., Lindorff-Larsen, K. and Shaw, D. E. Protein folding kinetics and thermodynamics from atomistic simulation. *Proc Natl Acad Sci U S A* **109**, 17845-17850 (2012).
- Piana, S., Lindorff-Larsen, K. and Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc Natl Acad Sci U S A* **110**, 5915-5920 (2013).
- Plotkin, S. S. and Onuchic, J. N. Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q Rev Biophys* **35**, 111-167 (2002).

- Popa, I., Fernandez, J. M. and Garcia-Manyes, S. Direct quantification of the attempt frequency determining the mechanical unfolding of ubiquitin protein. *J Biol Chem* **286**, 31072-31079 (2011).
- Post, K., Pitschke, M., Schafer, O., Wille, H., Appel, T. R., Kirsch, D., Mehlhorn, I., Serban, H., Prusiner, S. B. and Riesner, D. Rapid acquisition of beta-sheet structure in the prion protein prior to multimer formation. *Biol Chem* **379**, 1307-1317 (1998).
- Prakash, S. and Matouschek, A. Protein unfolding in the cell. *Trends Biochem Sci* **29**, 593-600 (2004).
- Priola, S. A. and Lawson, V. A. Glycosylation influences cross-species formation of protease-resistant prion protein. *EMBO J* **20**, 6692-6699 (2001).
- Proshkin, S., Rahmouni, A. R., Mironov, A. and Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504-508 (2010).
- Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136-144 (1982).
- Prusiner, S. B. Prions. *Proc Natl Acad Sci U S A* **95**, 13363-13383 (1998).
- Prusiner, S. B., McKinley, M. P., Bowman, K. A., Bolton, D. C., Bendheim, P. E., Groth, D. F. and Glenner, G. G. Scrapie prions aggregate to form amyloid-like birefringent rods. *Cell* **35**, 349-358 (1983).
- Randles, L. G., Rounsevell, R. W. and Clarke, J. Spectrin domains lose cooperativity in forced unfolding. *Biophys J* **92**, 571-577 (2007).
- Redecke, L., Bergen, M. v., Clos, J., Konarev, P. V., Svergun, D. I., Fittschen, U. E. A., Broekaert, J. A. C., Bruns, O., Georgieva, D., Mandelkow, E., Genov, N. and Betzel, C. Structural characterization of  $\beta$ -sheeted oligomers formed on the pathway of oxidative prion protein aggregation in vitro. *J Struct Biol* **157**, 308-320 (2007).
- Reynolds, R., Bermúdez-Cruz, R. M. and Chamberlin, M. J. Parameters affecting transcription termination by *Escherichia coli* RNA polymerase: I. Analysis of 13 rho-independent terminators *J Mol Biol* **224**, 31-51 (1992).
- Rezaei-Ghaleh, N., Blackledge, M. and Zweckstetter, M. Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *ChemBioChem* **13**, 930-950 (2012).
- Rieder, R., Lang, K., Graber, D. and Micura, R. Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *ChemBioChem* **8**, 896-902 (2007).

- Rieder, U., Kreutz, C. and Micura, R. Folding of a transcriptionally acting PreQ1 riboswitch. *Proc Natl Acad Sci U S A* **107**, 10804-10809 (2010).
- Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. and Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109-1112 (1997).
- Rief, M., Pascual, J., Saraste, M. and Gaub, H. E. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J Mol Biol* **286**, 553-561 (1999).
- Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R. and Wuthrich, K. NMR structure of the mouse prion protein domain PrP(121-231). *Nature* **382**, 180-182 (1996).
- Ritort, F. Nonequilibrium fluctuations in small systems: from physics to biology. in *Advances in Chemical Physics* 31-123 (John Wiley & Sons, Inc., 2008).
- Robinson, P. J. and Pinheiro, T. J. The unfolding of the prion protein sheds light on the mechanisms of prion susceptibility and species barrier. *Biochemistry* **48**, 8551-8558 (2009).
- Rohrbach, A. and Stelzer, E. H. Optical trapping of dielectric particles in arbitrary fields. *J Opt Soc Am A Opt Image Sci Vis* **18**, 839-853 (2001).
- Rohrbach, A. and Stelzer, E. H. Trapping forces, force constants, and potential depths for dielectric spheres in the presence of spherical aberrations. *Appl Opt* **41**, 2494-2507 (2002).
- Roostae, A., Cote, S. and Roucou, X. Aggregation and amyloid fibril formation induced by chemical dimerization of recombinant prion protein in physiological-like conditions. *J Biol Chem* **284**, 30907-30916 (2009).
- Roth, A. and Breaker, R. R. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* **78**, 305-334 (2009).
- Rudd, P. M., Merry, A. H., Wormald, M. R. and Dwek, R. A. Glycosylation and prion protein. *Curr Opin Struct Biol* **12**, 578-586 (2002).
- Russell, R. RNA misfolding and the action of chaperones. *Front Biosci* **13**, 1-20 (2008).
- Russell, R., Zhuang, X., Babcock, H. P., Millett, I. S., Doniach, S., Chu, S. and Herschlag, D. Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci U S A* **99**, 155-160 (2002).
- Saa, P., Castilla, J. and Soto, C. Ultra-efficient replication of infectious prions by automated protein misfolding cyclic amplification. *J Biol Chem* **281**, 35245-35252 (2006).

- Sabelko, J., Ervin, J. and Gruebele, M. Observation of strange kinetics in protein folding. *Proc Natl Acad Sci U S A* **96**, 6031-6036 (1999).
- Saborio, G. P., Permanne, B. and Soto, C. Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* **411**, 810-813 (2001).
- Saenger, W. *Principles of nucleic acid structure*. (Springer, 1984).
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* **324**, 163-166 (1986).
- Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, J. J., McFarlane, H. T., Madsen, A. O., Riek, C. and Eisenberg, D. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453-457 (2007).
- Schechter, A. N. Measurement of fast biochemical reactions. *Science* **170**, 273-280 (1970).
- Schellenberg, M. J., Ritchie, D. B., Wu, T., Markin, C. J., Spyropoulos, L. and MacMillan, A. M. Context-dependent remodeling of structure in two large protein fragments. *J Mol Biol* **402**, 720-730 (2010).
- Sela, M., White, F. H., Jr. and Anfinsen, C. B. Reductive cleavage of disulfide bridges in ribonuclease. *Science* **125**, 691-692 (1957).
- Seol, Y., Skinner, G. M. and Visscher, K. Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys Rev Lett* **93**, 118102-118105 (2004).
- Serganov, A. and Nudler, E. A decade of riboswitches. *Cell* **152**, 17-24 (2013).
- Serganov, A., Yuan, Y. R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R. and Patel, D. J. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* **11**, 1729-1741 (2004).
- Shank, E. A., Cecconi, C., Dill, J. W., Marqusee, S. and Bustamante, C. The folding cooperativity of a protein is controlled by its chain topology. *Nature* **465**, 637-640 (2010).
- Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y. and Wriggers, W. Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341-346 (2010).

- Silveira, J. R., Raymond, G. J., Hughson, A. G., Race, R. E., Sim, V. L., Hayes, S. F. and Caughey, B. The most infectious prion protein particles. *Nature* **437**, 257-261 (2005).
- Simoneau, S., Rezaei, H., Sales, N., Kaiser-Schulz, G., Lefebvre-Roque, M., Vidal, C., Fournier, J. G., Comte, J., Wopfner, F., Grosclaude, J., Schatzl, H. and Lasmezas, C. I. In vitro and in vivo neurotoxicity of prion protein oligomers. *PLoS Pathog* **3**, e125 (2007).
- Slabinski, L., Jaroszewski, L., Rodrigues, A. P., Rychlewski, L., Wilson, I. A., Lesley, S. A. and Godzik, A. The challenge of protein structure determination--lessons from structural genomics. *Protein Sci* **16**, 2472-2482 (2007).
- Smirnovas, V., Baron, G. S., Offerdahl, D. K., Raymond, G. J., Caughey, B. and Surewicz, W. K. Structural organization of brain-derived mammalian prions examined by hydrogen-deuterium exchange. *Nat Struct Mol Biol* **18**, 504-506 (2011).
- Smirnovas, V., Kim, J. I., Lu, X., Atarashi, R., Caughey, B. and Surewicz, W. K. Distinct structures of scrapie prion protein (PrP<sup>Sc</sup>)-seeded versus spontaneous recombinant prion protein fibrils revealed by hydrogen/deuterium exchange. *J Biol Chem* **284**, 24233-24241 (2009).
- Smith, S. B., Cui, Y. and Bustamante, C. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795-799 (1996).
- Smock, R. G. and Gierasch, L. M. Sending signals dynamically. *Science* **324**, 198-203 (2009).
- Sokolowski, F., Modler, A. J., Masuch, R., Zirwer, D., Baier, M., Lutsch, G., Moss, D. A., Gast, K. and Naumann, D. Formation of critical oligomers is a key event during conformational transition of recombinant Syrian hamster prion protein. *J Biol Chem* **278**, 40481-40492 (2003).
- Soranno, A., Buchli, B., Nettels, D., Cheng, R. R., Muller-Spath, S., Pfeil, S. H., Hoffmann, A., Lipman, E. A., Makarov, D. E. and Schuler, B. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc Natl Acad Sci U S A* **109**, 17800-17806 (2012).
- Soto, C. *Prions: The new biology of proteins*. (CRC Press, 2005).
- Spiro, T. G. and Gaber, B. P. Laser Raman scattering as a probe of protein structure. *Annu Rev Biochem* **46**, 553-572 (1977).
- Stahl, N., Borchelt, D. R., Hsiao, K. and Prusiner, S. B. Scrapie prion protein contains a phosphatidylinositol glycolipid. *Cell* **51**, 229-240 (1987).

- Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C. and Rief, M. The complex folding network of single calmodulin molecules. *Science* **334**, 512-516 (2011).
- Stoddard, C. D., Gilbert, S. D. and Batey, R. T. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* **14**, 675-684 (2008).
- Stohr, J., Watts, J. C., Mensinger, Z. L., Oehler, A., Grillo, S. K., DeArmond, S. J., Prusiner, S. B. and Giles, K. Purified and synthetic Alzheimer's amyloid beta (A $\beta$ ) prions. *Proc Natl Acad Sci U S A* **109**, 11025-11030 (2012).
- Stohr, J., Weinmann, N., Wille, H., Kaimann, T., Nagel-Steger, L., Birkmann, E., Panza, G., Prusiner, S. B., Eigen, M. and Riesner, D. Mechanisms of prion protein assembly into amyloid. *Proc Natl Acad Sci U S A* **105**, 2409-2414 (2008).
- Straub, J. E. and Thirumalai, D. Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annu Rev Phys Chem* **62**, 437-463 (2011).
- Swietnicki, W., Petersen, R., Gambetti, P. and Surewicz, W. K. pH-dependent stability and conformation of the recombinant human prion protein PrP(90-231). *J Biol Chem* **272**, 27517-27520 (1997).
- Swietnicki, W., Petersen, R. B., Gambetti, P. and Surewicz, W. K. Familial mutations and the thermodynamic stability of the recombinant human prion protein. *J Biol Chem* **273**, 31048-31052 (1998).
- Thirumalai, D., O'Brien, E. P., Morrison, G. and Hyeon, C. Theoretical perspectives on protein folding. *Annu Rev Biophys* **39**, 159-183 (2010).
- Tinoco, I., Chen, G. and Qu, X. RNA reactions one molecule at a time. *Cold Spring Harb Perspect Biol* **2**, a003624 (2010).
- Tompa, P., Tusnady, G. E., Friedrich, P. and Simon, I. The role of dimerization in prion replication. *Biophys J* **82**, 1711-1718 (2002).
- Truhlar, D. G., Garrett, B. C. and Klippenstein, S. J. Current status of transition-state theory. *J Phys Chem* **100**, 12771-12800 (1996).
- Tyedmers, J., Mogk, A. and Bukau, B. Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* **11**, 777-788 (2010).
- Valentine, M. T., Guydosh, N. R., Gutierrez-Medina, B., Fehr, A. N., Andreasson, J. O. and Block, S. M. Precision steering of an optical trap by electro-optic deflection. *Opt Lett* **33**, 599-601 (2008).
- van der Kamp, M. W. and Daggett, V. The consequences of pathogenic mutations to the human prion protein. *Protein Eng Des Sel* **22**, 461-468 (2009).

- Vanik, D. L. and Surewicz, W. K. Disease-associated F198S mutation increases the propensity of the recombinant prion protein for conformational conversion to scrapie-like form. *J Biol Chem* **277**, 49065-49070 (2002).
- Veith, N. M., Plattner, H., Stuermer, C. A. O., Schulz-Schaeffer, W. J. and Bürkle, A. Immunolocalisation of PrPSc in scrapie-infected N2a mouse neuroblastoma cells by light and electron microscopy. *Eur J Cell Biol* **88**, 45-63 (2009).
- Vogel, U. and Jensen, K. F. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J Bacteriol* **176**, 2807-2813 (1994).
- Waldauer, S. A., Bakajin, O. and Lapidus, L. J. Extremely slow intramolecular diffusion in unfolded protein L. *Proc Natl Acad Sci U S A* **107**, 13713-13717 (2010).
- Wang, F., Wang, X., Yuan, C. G. and Ma, J. Generating a prion with bacterially expressed recombinant prion protein. *Science* **327**, 1132-1135 (2010).
- Wang, M. D., Yin, H., Landick, R., Gelles, J. and Block, S. M. Stretching DNA with optical tweezers. *Biophys J* **72**, 1335-1346 (1997).
- Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A. B., Riek, R. and Meier, B. H. Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319**, 1523-1526 (2008).
- Watanabe, Y., Hiraoka, W., Shimoyama, Y., Horiuchi, M., Kuwabara, M. and Inanami, O. Instability of familial spongiform encephalopathy-related prion mutants. *Biochem Biophys Res Commun* **366**, 244-249 (2008).
- Wen, J. D., Lancaster, L., Hodges, C., Zeri, A. C., Yoshimura, S. H., Noller, H. F., Bustamante, C. and Tinoco, I. Following translation by single ribosomes one codon at a time. *Nature* **452**, 598-603 (2008).
- Wen, Y., Li, J., Yao, W., Xiong, M., Hong, J., Peng, Y., Xiao, G. and Lin, D. Unique structural characteristics of the rabbit prion protein. *J Biol Chem* **285**, 31682-31693 (2010).
- Westergaard, L., Christensen, H. M. and Harris, D. A. The cellular prion protein (PrP(C)): its physiological function and role in disease. *Biochim Biophys Acta* **1772**, 629-644 (2007).
- Wickiser, J. K., Cheah, M. T., Breaker, R. B. and Crothers, D. M. The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44**, 13404-13414 (2005a).
- Wickiser, J. K., Winkler, W. C., Breaker, R. R. and Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* **18**, 49-60 (2005b).

- Wildegger, G., Liemann, S. and Glockshuber, R. Extremely rapid folding of the C-terminal domain of the prion protein without kinetic intermediates. *Nat Struct Biol* **6**, 550-553 (1999).
- Wille, H., Bian, W., McDonald, M., Kendall, A., Colby, D. W., Bloch, L., Ollesch, J., Borovinskiy, A. L., Cohen, F. E., Prusiner, S. B. and Stubbs, G. Natural and synthetic prion structure from X-ray fiber diffraction. *Proc Natl Acad Sci U S A* **106**, 16990-16995 (2009).
- Wille, H., Prusiner, S. B. and Cohen, F. E. Scrapie infectivity is independent of amyloid staining properties of the N-terminally truncated prion protein. *J Struct Biol* **130**, 323-338 (2000).
- Wolynes, P. G. The protein folding energy landscape: a primer. in *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches* 49-69 (The Royal Society of Chemistry, 2008).
- Woodside, M. T., Anthony, P. C., Behnke-Parks, W. M., Larizadeh, K., Herschlag, D. and Block, S. M. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* **314**, 1001-1004 (2006a).
- Woodside, M. T., Behnke-Parks, W. M., Larizadeh, K., Travers, K., Herschlag, D. and Block, S. M. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc Natl Acad Sci U S A* **103**, 6190-6195 (2006b).
- Woodside, M. T., Garcia-Garcia, C. and Block, S. M. Folding and unfolding single RNA molecules under tension. *Curr Opin Chem Biol* **12**, 640-646 (2008).
- Woodside, M. T. and Valentine, M. T. Single-molecule manipulation using optical traps. in *Handbook of Single-Molecule Biophysics* (eds Hinterdorfer, P. and Oijen, A.) Ch. 12, (Springer, 2009).
- Wuthrich, K. and Riek, R. Three-dimensional structures of prion proteins. *Adv Protein Chem* **57**, 55-82 (2001).
- Xi, Z., Gao, Y., Sirinakis, G., Guo, H. and Zhang, Y. Single-molecule observation of helix staggering, sliding, and coiled coil misfolding. *Proc Natl Acad Sci U S A* **109**, 5711-5716 (2012).
- Xu, D. and Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **80**, 1715-1735 (2012).
- Yang, S., Levine, H., Onuchic, J. N. and Cox, D. L. Structure of infectious prions: stabilization by domain swapping. *FASEB J* **19**, 1778-1782 (2005).

- Yu, H., Dee, D. R., Liu, X., Brigley, A. M., Sosova, I. and Woodside, M. T. Structural conversion of prion protein dimers via a multi-step pathway observed at the single-molecule level. (under preparation).
- Yu, H., Dee, D. R. and Woodside, M. T. Single-molecule approaches to prion protein misfolding. *Prion* **7**, 140-146 (2013).
- Yu, H., Gupta, A. N., Liu, X., Neupane, K., Brigley, A. M., Sosova, I. and Woodside, M. T. Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *Proc Natl Acad Sci U S A* **109**, 14452-14457 (2012a).
- Yu, H., Liu, X., Neupane, K., Gupta, A. N., Brigley, A. M., Solanki, A., Sosova, I. and Woodside, M. T. Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc Natl Acad Sci U S A* **109**, 5283-5288 (2012b).
- Yu, J., Malkova, S. and Lyubchenko, Y. L. alpha-Synuclein misfolding: single molecule AFM force spectroscopy study. *J Mol Biol* **384**, 992-1001 (2008).
- Yu, Z., Koirala, D., Cui, Y., Easterling, L. F., Zhao, Y. and Mao, H. Click chemistry assisted single-molecule fingerprinting reveals a 3D biomolecular folding funnel. *J Am Chem Soc* **134**, 12338-12341 (2012c).
- Yusupova, G. Z., Yusupov, M. M., Cate, J. H. D. and Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **106**, 233-241 (2001).
- Zahn, R., Liu, A., Luhrs, T., Riek, R., von Schroetter, C., Lopez Garcia, F., Billeter, M., Calzolari, L., Wider, G. and Wuthrich, K. NMR solution structure of the human prion protein. *Proc Natl Acad Sci U S A* **97**, 145-150 (2000).
- Zerr, I. Human prion diseases: progress in clinical trials. *Brain* **136**, 996-997 (2013).
- Zhang, H., Kaneko, K., Nguyen, J. T., Livshits, T. L., Baldwin, M. A., Cohen, F. E., James, T. L. and Prusiner, S. B. Conformational transitions in peptides containing two putative alpha-helices of the prion protein. *J Mol Biol* **250**, 514-526 (1995).
- Zhang, J., Lau, M. W. and Ferré-D'Amaré, A. R. Ribozymes and riboswitches: modulation of RNA function by small molecules. *Biochemistry* **49**, 9123-9131 (2010).
- Zhang, Y. Protein structure prediction: when is it useful? *Curr Opin Struct Biol* **19**, 145-155 (2009).
- Zhang, Y., Swietnicki, W., Zagorski, M. G., Surewicz, W. K. and Sonnichsen, F. D. Solution structure of the E200K variant of human prion protein. Implications for the mechanism of pathogenesis in familial prion diseases. *J Biol Chem* **275**, 33650-33654 (2000).

- Zheng, P., Cao, Y. and Li, H. Facile method of constructing polyproteins for single-molecule force spectroscopy studies. *Langmuir* **27**, 5713-5718 (2011).
- Zhou, M., Ottenberg, G., Sferrazza, G. F. and Lasmezas, C. I. Highly neurotoxic monomeric alpha-helical prion protein. *Proc Natl Acad Sci U S A* **109**, 3113-3118 (2012).
- Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).
- Zwanzig, R., Szabo, A. and Bagchi, B. Levinthal's paradox. *Proc Natl Acad Sci U S A* **89**, 20-22 (1992).