

Transition Paths In Folding Reactions

by

Noel Quenton Hoffer

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

Doctor of philosophy

Department of Physics
University of Alberta

© Noel Quenton Hoffer, 2021

Abstract

Folding is the process by which biomolecules spontaneously self-assemble into specific, complex, three-dimensional structures from simple one-dimensional polypeptide chains. Folding is a critical process in biology as there exists a tight link between the structure and function of biomolecules, and misfolding into the wrong structure often leads to disease. Traditional studies of folding have used ensemble measurements to monitor how the statistics of the folded and unfolded states, respond to perturbations to the stability of the states. Ensemble studies have yielded important results and indeed, are the primary basis for our current understanding of folding reactions. However, due to the asynchronous nature of folding reactions, such studies are incapable of probing transition paths.

Transition paths comprise those parts of a folding trajectory where the molecule passes through the high-energy transition states separating the folded and unfolded states. The transition states determine the folding kinetics and mechanism but are difficult to observe because of their brief duration. Single-molecule experiments have in recent years, begun to characterize transition paths in folding reactions, allowing the microscopic conformational dynamics that occur as a molecule traverses the energy barriers, to be probed directly.

In this thesis, I present the first direct measurements of transition-path trajectories. I then show how, using single-molecule force spectroscopy, I have been able to make the first-ever measurements of several transition-path properties including: the local velocity along the paths, the path shapes, and the transition state

dynamics inferred by them. I discuss how these measurements have been related to theories of folding as diffusion over an energy landscape, to deduce properties such as the diffusion coefficient, and how they further our understanding of folding. The richly detailed information available from transition path measurements holds great promise for an improved understanding of microscopic mechanisms in folding.

Preface

The original research presented in this thesis involved many members of the Woodside research group.

Chapter 6 was originally published as two separate manuscripts. Sections 6.1- 6.4 were published as K. Neupane, N.Q. Hoffer, M.T. Woodside “*Measuring the velocity along transition paths during the folding of single biological molecules.*” Phys. Rev. Lett., 121:018102 (2018). Sections 6.5- 6.7 were published as K. Neupane, N.Q. Hoffer, M.T. Woodside “*Testing Kinetic Identities Involving Transition-Path Properties using Single-Molecule Folding Trajectories.*” J. Phys. Chem. B, 122:11095-11099 (2018). For these papers, I participated in the design of the research, sample preparation, measurements, analysis, and the writing of the paper.

Chapter 7 was originally published as A.G.T. Pyo, N.Q. Hoffer, K. Neupane, M.T. Woodside, “*Transition-path properties for folding reactions in the limit of small barriers*” J. Chem. Phys. 149:115101 (2018). For this paper, I participated in the sample preparation, and measurements as well as aided in the development of the theory.

Chapter 8 was originally published in the manuscript N.Q. Hoffer, K. Neupane, A.G.T. Pyo and M.T. Woodside, “*Measuring the average shape of transition paths during the folding of a single biological molecule*” Proc. Natl. Acad. Sci. 116:8125-8130 (2019). For this paper, I participated in the design of the research, sample preparation, measurements, analysis, and the writing of the paper.

Dedication

To Raven,

ask simple questions, demand simple answers.

Acknowledgments

First, I would like to thank my professor, Michael Woodside, for all of his advice and guidance, both inside and outside of the lab. I had the great fortune of having my desk directly outside of Michael's office, providing me with near unrestricted access to ask questions, discuss results, and receive feedback. From Michael, I have learned more about science and experiment than I have from nine years of courses.

I thank Krishna Neupane for teaching me how to properly operate the optical traps. Krishna's wealth of practical knowledge about the traps and how to take measurements with them was an invaluable resource for me.

A very special thank you to my mother, who always encouraged me to make my own path. I will never forget a time, five years after completing high school when my mother told me that I could only row my proverbial boat around for so long, and eventually I would have to pick a direction and head to shore. Shortly after this discussion, I enrolled in university.

I thank my father, whose hard work and sacrifice have afforded me the opportunity to pursue higher education.

Last, but certainly not least, I thank my wife, Amy, for her patience and support, and for always trusting in my ability to make a career out of a passion. It can be hard to be married to an academic and Amy always took up the slack when I was busy with course work or working in the lab. Thank you.

Table of Contents

Abstract	ii
Preface	iv
Dedication	v
Acknowledgments	vi
Table of Contents	vii
List of Tables	ix
List of Figures	ix
List of Symbols and Abbreviations	xi
Chapter 1. Introduction	1
1.1 Biomolecular folding	1
1.2 The structure of DNA	4
1.3 Thesis outline	6
Chapter 2. The theory of folding	8
2.1 Energy landscapes	8
2.2 Kramers reaction rate theory	10
2.3 Transition paths	12
2.3.1 Langevin dynamics	14
2.3.2 Most probable paths	19
2.3.3 Folding committers	20
2.3.4 The Bryngelson and Wolynes landscape picture	22
Chapter 3. Optical tweezers	27
3.1 Theory of optical trapping	27
3.2 Optical tweezers design	31
3.2.1 Trapping laser	31
3.2.2 Optics	33
3.2.3 Position detection	37
3.3 Stiffness calibration of the optical trap	40
3.3.1 Power spectral density calibration	40
3.3.2 Variance calibration method	42
3.3.3 Stokes drag calibration method	43
3.4 Summary	43
Chapter 4. Optical tweezers measurements	45
4.1 Equilibrium measurements	46
4.1.1 Constants force measurements	46
4.1.2 Constants trap separation measurements	48
4.2 Non-equilibrium measurements	50
4.2.1 Force vs extension measurements	50
4.3 Measuring transition paths	53
4.3.1 Requirements for measuring transition paths	53
4.3.2 Instrumental artifacts in kinetic rates and transition path properties	57
4.3.3 Reaction coordinate quality	60
4.5.3 Specifics of transition-path measurements	63
4.6 Summary	64
Chapter 5. Transition-path studies	66

5.1 smFRET measurements of average transition-path times	66
5.2 Force spectroscopy measurements of transition-path times	70
5.3 Transition paths in protein-protein complexes	74
Chapter 6. Transition-path velocities	77
6.1 Transition-path velocity measurements	77
6.2 Diffusion coefficients from transition-path velocities	84
6.3 Deviations from 1D harmonic theories	85
6.4 Comparing transition-state theory to Kramers' theory	88
6.5 Transition-path velocity and position probability distribution	91
6.6 Calculating the committor from transition-path velocities	94
6.7 Discussion of transition path velocity results	96
Chapter 7. Transition path properties in the limit of small barriers	97
7.1 Large barrier approximations	97
7.2 The need for small barrier approximations	98
7.3 Derivation of the small barrier limit equations	99
7.4. Application to Brownian dynamics simulations	104
7.5 Application to anharmonic barriers	109
7.6 Application to experimental data	111
7.7 Discussion of results	113
Chapter 8. Transition-state dynamics	117
8.1 Pausing along transition paths	117
8.2 Pause durations	123
8.3 Diffusion coefficients from pausing statistics	125
8.4 Discussion of transition path pausing results	128
Chapter 9. The average shapes of transition paths	131
9.1 Motivation for studying transition-path shapes	131
9.2 Time domain averaged transition-path shapes	133
9.3 Extension domain averaged transition-path shapes	135
9.4 Exploring the discrepancy between the time and extension domain averages	138
9.5 Comparison of results to theory	138
9.6 Variability in transition-path shapes	140
9.7 Discussion of average transition-path shape results	142
Chapter 10. Future work and outlook	149
References	152
Appendices	164
Appendix A DNA hairpin sequences	164
Appendix B Sample preparation	165
Appendix C Smoothing of transition-path trajectories	165
Appendix D Boundary effects in transition-path measurements	168
Appendix E Detection system upgrades	170

List of Tables

Table 4.1. Kinetic parameters of DNA hairpins.	59
Table 6.1. Comparison of D from $\langle v(x^\ddagger) \rangle$, transition times, τ_{TP} , and rates....	85
Table 6.2. Comparison of κ from rates, velocities, and Kramers' theory.....	90
Table 8.1. Results from fitting pause-duration distributions... ..	126
Table 9.1. Diffusion coefficients from fitting dominant path shapes... ..	140
Table 9.2. Diffusion coefficients from fitting extension-domain averages	148
Table A.1. DNA hairpin sequences used in this thesis	164

List of Figures

Fig. 1.1. DNA structure.....	6
Fig. 2.1. Energy landscape representation of protein folding... ..	10
Fig. 2.2. Kramers' reaction rate theory.....	12
Fig. 2.3. Transition paths.. ..	14
Fig. 2.4. Lattice representation of funneled energy landscapes.. ..	26
Fig. 3.1 Ray optics view of optical trapping.....	31
Fig. 3.2 Absorption spectrum of water	33
Fig. 3.3 Optics layout.....	36
Fig. 3.4 Back focal plane geometry	36
Fig. 3.5 Power spectral density of a trapped bead	42
Fig. 4.1 Constant force measurements.....	48
Fig. 4.2 Active and passive force clamps	48
Fig. 4.3 Constant position measurements	50
Fig. 4.4 Force extension curves	53
Fig. 4.5 Instrument response time measurement	57
Fig. 4.6 Reaction coordinate quality tests for DNA hairpin 30R50/T4	63
Fig. 4.7 Transition-path measurements	64
Fig. 5.1 smFRET measurements	69
Fig. 5.2 smFRET measurements of transition-path times	70
Fig. 5.3 SMFS measurements of transition-path times	74
Fig. 5.4 Measuring transition paths in an encounter complex	76
Fig. 6.1 Measuring transition path velocity with optical tweezers	81
Fig. 6.2 Transition-path velocities	82
Fig. 6.3 Transition-path velocity distributions	83
Fig. 6.4 Simulated transition-path velocity distributions	87

Fig. 6.5 Effects of smoothing on transition-path velocity distributions	87
Fig. 6.6 Simulated velocity profiles	88
Fig. 6.7 Rate measurements	91
Fig. 6.8 Test of the relation between transition-path velocity and occupancy	94
Fig. 6.9 Comparison of empirical p_{fold} and p_{fold} calculated from transition-path velocity.....	95
Fig. 7.1 Average transition-path velocity from simulated transitions over harmonic barriers	107
Fig. 7.2 Average transition-path times from simulated transitions over harmonic barriers	108
Fig. 7.3 Estimates of the diffusion coefficient for simulations over harmonic barriers	109
Fig. 7.4 Estimates of the diffusion coefficient for anharmonic barriers	111
Fig. 8.1 Observing pausing along transition paths.....	120
Fig. 8.2 Pausing within the transition states	121
Fig. 8.3 Measurements of reference construct.....	121
Fig. 8.4 Distribution of 'pauses' in reference construct trajectories	122
Fig. 8.5 Microscopic motions during hairpin folding	122
Fig. 8.6 Pause durations	124
Fig. 8.7 Models of microwells in the barrier region	125
Fig. 8.8 Position dependence of D	127
Fig. 9.1 Time-domain average transition-path shape $\langle t(x) \rangle$ for hairpin 30R50/T4	135
Fig. 9.2 Time-domain average transition-path shapes $\langle t(x) \rangle$ for different hairpins.....	135
Fig. 9.3 Extension-domain average transition-path shapes $\langle x(t \tau) \rangle_\tau$ for hairpin 30R50/T4	136
Fig. 9.4 Extension-domain average transition-path shapes $\langle x(t \tau) \rangle_\tau$ for different hairpins.....	136
Fig. 9.5 Transition-path shapes from Brownian-dynamics simulations	140
Fig. 9.6 Comparison of path-shape variance for different hairpins	142
Fig. 9.7 Effects of noise on time-domain and extension-domain averages	147
Fig. 9.8 Extension-domain average path shape at τ_{TP}	147
Fig. 9.9 Diversity in transition-path ensembles.....	148
Fig. A.1 DNA hairpin legend.....	164
Fig. C.1 Effects of smoothing on velocity profiles	168

List of Symbols and Abbreviations

Symbol	Definition
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
dsDNA	double-stranded deoxyribonucleic acid
ssDNA	single-stranded deoxyribonucleic acid
A	adenine
G	guanine
C	cytosine
T	thymine
U	uracil
1D	one dimensional
$k_{\text{fold/unfold}}$	folding/unfolding rate
D	conformational diffusion coefficient
κ_w	well curvature
κ_b	barrier curvature
β	inverse thermal energy
ΔG^\ddagger	free energy barrier height
TP	transition path
OT	optical trap
F_{scatt}	scattering force
c	speed of light in vacuum
λ	wavelength
F_{grad}	gradient force
$\lambda/2$	half waveplate
Nd:YAG	neodymium-doped yttrium aluminum garnet
Nd:YVO ₄	neodymium-doped yttrium orthovanadate
EOD	electro-optic deflector
AOD	acousto-optic deflector
PSD	duolateral position sensitive diode
QPD	quadrant photodiode
NA	numerical aperture
SMFS	single-molecule force spectroscopy
FRET	Förster resonance energy transfer
smFRET	single-molecule FRET
TEM	transverse electromagnetic mode
HeNe	Helium Neon
MFD	mode field diameter
k_B	Boltzmann constant
η	fluid viscosity
FEC	force extension curve
WLC	worm-like chain
L_p	persistence length
L_c	contour length
bp	base pair

τ_{TP}	average transition-path time
x^{\ddagger}	barrier location
v_{th}	thermal velocity
TST	transition-state theory
p_{fold}	committor probability

Chapter 1

Introduction

1.1 Biomolecular folding

Folding is the process by which biological macromolecules such as nucleic acids and proteins, spontaneously self-assemble into specific, complex, three-dimensional structures, from simple one-dimensional polypeptide chains. These polypeptide chains consist of amino acids in the case of proteins and nucleotide bases in nucleic acids. Folding into the correct structure is crucial, as there exists a tight link between the structure and function of biomolecules. The complexity of these structures came as a surprise when the first atomic-resolution protein crystal structures were resolved in the early 1960s. When John Kendrew resolved the structure of myoglobin, an oxygen-binding protein found in muscle cells, he observed a globular compact structure consisting of 8 alpha helices, interconnected by loops and all tightly packed together in irregular ways (1). Although the presence of helices had been anticipated by Linus Pauling (2), the irregularity and overall lack of symmetry of the structure was unexpected. At roughly the same time, Anfinsen and colleagues observed that proteins can reversibly fold in solution, unaided by outside forces (3). This observation led to the thermodynamic hypothesis of folding. The hypothesis states that, in its standard physiological environment, the native structure is the one in which the Gibbs free energy for the system is lowest; that is, that the native structure is determined by the totality of the interatomic interactions and hence, by the protein's amino acid sequence (4).

Together, these two discoveries of protein structure and the folding process spawned what would be known as the folding ‘problem’.

The problem is really a set of three very closely related questions: (a) for a given amino acid sequence, how are the interatomic forces balanced in order to reliably produce the specific structure of the protein?; (b) for a given amino acid sequence, how can we predict a protein’s native folded structure?; and (c) what mechanisms allow proteins to fold quickly and reliably? (5).

Although postulated in the context of proteins, the folding problem applies equally to nucleic acids. Like proteins, nucleic acids fold into their functional conformations. The physical principles that govern the folding are the same for both proteins and nucleic acids, but the strength or importance, of any particular interaction or mechanism, depends on the physical makeup of the individual molecule. The major differences between the folding of proteins and nucleic acids arise in the early stages of folding (6). The initial events leading to chain compaction in nucleic acids are driven by counterion-mediated collapse, whereas lowering of the free energy of the polypeptide chain by burying the hydrophobic residues, is the main driving force in the collapse of proteins (6)

The vast majority of nucleic acid folding studies have focused on ribonucleic acid (RNA) molecules, owing to the highly diverse set of functions and associated structures of RNA in the cell (7). Despite this, arguably the most famous structure in all of biology is the deoxyribonucleic acid (DNA) double helix. The discovery of the double-helix structure of DNA by Rosalind Franklin, James Watson, and Francis Crick in the 1950s, is viewed as a major milestone in biology and modern

science as a whole (8). This discovery brought with it not only the structural underpinnings of nucleic acids, base-pairing, and base stacking interactions, but also the mechanistic understanding of genetics. This knowledge has allowed for the manipulation and design of specific DNA structures. Despite these revelations, the physical process by which nucleic acid molecules adopt their native structure remains unclear.

The classical physical picture of folding is that of a random walk through configurational phase space, where each possible combination of bond angles, comprises a single conformer and single point in the configurational phase space. This picture appears to be perfectly valid from a thermodynamic perspective but a simple calculation of the timescale for folding implied by such a picture immediately indicates that something is missing from the picture. A typical protein is a polypeptide chain of 300 amino acids connected by 299 peptide bonds, and each of these bonds has two dihedral angles, known as the phi and psi bond angles. If each of these bond angles can be in one of three stable conformations (typical of proteins), then there are 3^{598} possible conformations. Even if the molecule was sampling these conformations at a rate of inverse picoseconds, it would take vastly longer than the lifetime of the universe, for such a protein to find its unique folded state. This clearly does not match the timescale of living organisms or the observed timescale of protein folding (on the order of milliseconds). This calculation was first done by Levinthal in 1969 (9) and is now commonly referred to as Levinthal's paradox. The way out of the paradox comes in the form of a funneled energy landscape, where the search is energetically biased toward the lowest energy folded

state (10).

1.2 The structure of DNA

In this thesis I investigate the physical process of folding, examining physical quantities, properties, and mechanisms that are equally applicable to proteins and nucleic acids. For simplicity, I use the DNA hairpin as a model structure for studying folding. It is therefore important to introduce DNA and provide context on both its structure and its intimately related biological functions.

The primary roles of DNA are to store the genetic information (genes) used to code for the amino acid sequences of proteins and to replicate that code, allowing it to pass from mother to daughter cell during cell division. The structure of DNA makes it ideally suited to these tasks. The primary structure of DNA is a polymer chain composed of monomeric nucleotide units. The nucleotides consist of one of four nucleobases ([A] adenine, [T] thymine, [C] cytosine, or [G] guanine) forming a glycosidic bond to a deoxyribose sugar attached covalently to a phosphate group. The nucleotides are then tethered together by covalent phosphodiester bonds between the deoxyribose sugar of one nucleotide to the phosphate group of the next, forming a sugar-phosphate backbone (Fig 1.1).

The principle secondary structure of DNA is the famous right-handed double helix consisting of two complementary strands of DNA paired together with strict base-pairing rules (A pairs with T and C pairs with G). Each base pair is held together with hydrogen bonding and forms a purine-pyrimidine set with the same shape and size as any other purine-pyrimidine set (Fig 1.1). The homogeneity of the purine-pyrimidine sets allows the bases to neatly stack on top of one another

forming Van der Waals stacking interactions and further stabilizing the helical structure(11).

While the strictness of the base-pairing results in high fidelity of information transfer and the complementary nature of the double helix provides a simple mechanism for replication, the specifics of the structure are also highly important. The most common form of the DNA double helix, B-form DNA, has a diameter of 2.0 nm with a 3.4 angstrom vertical distance between base pairs and completes one turn of the helix for every 10.5 base pairs. The sugar-phosphate backbones of the complementary strands are not directly adjacent to each other resulting in a large (major) groove and a smaller (minor) groove running along the sides of the double helix and forming specific binding sites. For example, during transcription, proteins known as transcription factors bind specifically to the major groove of the DNA double helix and serve to regulate gene expression within the cell.

Many forms of the DNA double helix, A-form, B-form, and Z-form are known to occur in nature but there are also many other secondary DNA structures including the G-quadruplex, i-motif, and stem-loop. The DNA stem-loop or DNA hairpin is the structure focused on in this thesis. A DNA hairpin is formed by a single strand of DNA with regions of self-complementarity. The regions of complementary nucleotides base-pair to form a helical stem region that is connected by an unpaired loop. It is one of the simplest folded structures with sequences that can be easily manipulated in the lab, making it a model structure for studying the underlying mechanisms of folding reactions.

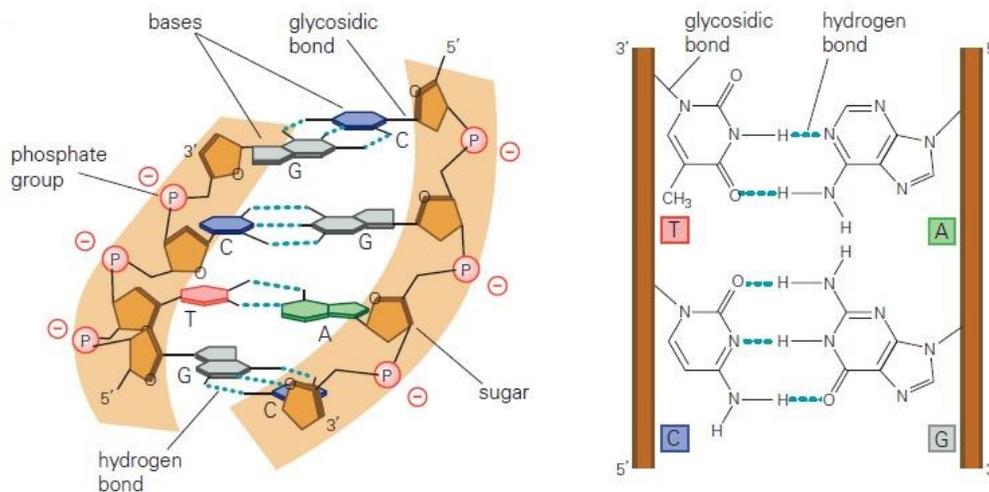


Fig. 1.1: DNA structure. The secondary structure (right) and tertiary structure (left) of B-form DNA shows the Watson-Crick base pairing as well as the phosphodiester backbone. The alignment of the planar structure of the bases, shown in the tertiary structure, leads to base stacking interactions. Adapted from Ref 11 with permission.

1.3 Thesis outline

This thesis has two primary goals: to demonstrate the ability to measure transition paths in folding reactions directly, using optical tweezers; and to show how said measurements can be used to elucidate the mechanisms underlying folding reactions.

The thesis begins with a breakdown of the key concepts behind the modern theory of folding, as well as the theoretical descriptions of transition paths. I then present and explain, the technique and instrument, that I have used to measure transition paths in the lab, explaining why these measurements are technically challenging to make, as well as how we have overcome these challenges. This is followed by an overview of the major advances that have been made with regard to experimental studies of transition paths, excluding my own work. I then present the

original research that I have been a part of during my graduate studies. I begin by examining the local velocity of transition paths, using the measured velocities to test established theories of folding. I then examine transition paths in the limit of small activation barriers. This is followed by an examination of the transient pauses present in the transition-path trajectories, describing how the statistics of these pauses can be used to probe transition-state dynamics. Next, I present measurements of the average and most probable transition-path shapes, demonstrating that these measurements contain information about the temporal sequence of events within transition paths. Lastly, I discuss the directions of future work and the outlook for experimental studies of transition paths.

Chapter 2

The theory of folding

This chapter provides an overview of the prevailing theoretical understanding of folding. We begin by discussing the concept of energy landscapes and how these landscapes provide a qualitative description of folding. We then show how the parameters of the energy landscape provide a quantitative description of folding rates through reaction rate theory. This is followed by an introduction to the main topic of this thesis, the transition paths taken over the energy landscape, establishing a mathematical framework for describing transition paths in folding reactions.

2.1 Energy landscapes

The modern picture of folding takes a statistical perspective on the process. In this picture, folding is viewed as a diffusive search through conformational phase space, for the minimum energy native state, over an energy landscape (Fig 2.1). This landscape is a hyper-dimensional manifold, with each point in the landscape representing the free energy of a corresponding structure of the molecule, in a configuration space that spans all possible conformations (12). The dimensionality of this landscape scales linearly with the number of bonds, as any conformation of a linear polymer, is minimally described by its bond angles (12).

As the energy landscape represents the free energy in all possible conformations, it also represents the relative probability of the molecule adopting any of the myriad conformations. As was first pointed out by Anfinsen, the native

structure of the molecule is unique and generally corresponds to the structure with the lowest free energy. Conversely, when the molecule is unfolded, it exists in a state of high entropy and high energy, fluctuating rapidly through multiple conformations. Yet, folding can progress from any of these unfolded conformations, progressing energetically downhill toward the folded state. In this sense, the landscape is said to be funneled toward the native folded state. Although folding reactions result in an overall reduction in free energy, local barriers exist within the funneled landscape, owing to asynchronous decreases in enthalpy and entropy (12). The heights and locations of these barriers can vary widely within the energy landscape of a single protein, with smaller barriers manifesting on top of larger barriers, creating a rough energy landscape. Rough energy landscapes arise because the competing individual interactions among different segments of the biopolymer, conflict with each other and frustrate the energy minimization process (10). The roughness slows the diffusion over the landscape, as the molecule becomes transiently trapped within local free energy minima (13). Biomolecular folding is therefore said to follow the principle of minimal frustration, meaning that, through evolutionary pressure, nature has selected for proteins with optimized energy landscapes that allow the molecule to fold quickly and reliably.

Given their high dimensionality and overall complexity, it is impossible to visualize or measure, the full energy landscape of even a very simple biopolymer. But even a highly simplified description, such as a one-dimensional projection of the full landscape onto a practical reaction coordinate, can be sufficient to characterize folding quantitatively (14). Indeed, simulation and experimental

studies of folding, for a variety of small proteins, have shown that properties such as folding rates and transition path times can be accounted for quantitatively by a 1D energy landscape, so long as the reaction coordinate is chosen properly for the projection (15).

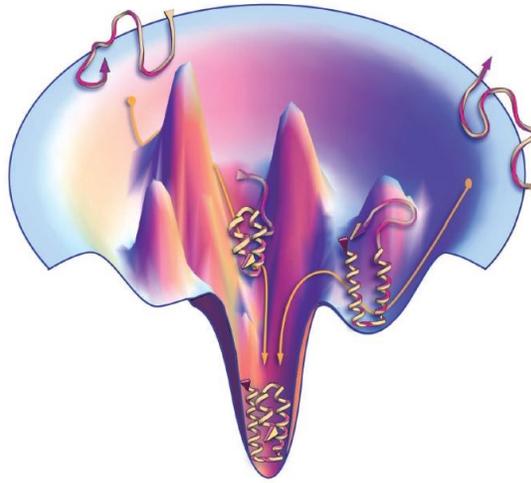


Fig 2.1: Energy landscape representation of protein folding. The high-energy unfolded molecule (top) is funneled toward the lowest energy folded state (bottom). Multiple pathways (arrows) lead from the unfolded states to the unique folded state. Adapted from Dill and MacCallum 2012 with permission.

2.2 Kramers reaction rate theory

The connection between the thermodynamics of energy landscape theory – whereby folding is described as a diffusive search for the lowest energy native state, over an energy landscape – and the folding kinetics is made via reaction rate theory in the high friction limit. Following the phenomenological observations of Arrhenius, which showed reaction rates follow an exponential temperature dependence, with a rate-limiting step determined by the activation energy or barrier height, Kramers describes the barrier crossing or escape process, as being governed by Brownian dynamics, driven by thermal fluctuations. The thermal fluctuations

are in turn connected to the temperature and friction coefficient, via the fluctuation-dissipation theorem.

Kramers' theory shows that, in the case of a 1D landscape, the reaction or folding rate, can be determined from the parameters of the energy landscape (Fig. 2.2)

$$k = \frac{\sqrt{\kappa_w \kappa_b}}{2\pi} \beta D \exp(-\beta \Delta G^\ddagger) \quad (2.1)$$

where k is the rate, κ_w and κ_b are the stiffness's of the well and barrier respectively, β is the inverse thermal energy, D is the diffusion constant and ΔG^\ddagger is the height of the barrier separating the reactants from the products. The pre-exponential term is commonly referred to as the Kramers pre-factor and describes the attempt frequency for barrier crossing due to thermal fluctuations. The theory makes the assumptions that the wells and barrier are parabolic and that D is constant throughout the barrier region.

In the case of an asymmetric energy landscape, the folding and unfolding rates will differ due to the differences in relative barrier heights and well curvatures, as depicted in Fig.2.2. For such an asymmetric landscape, the folding and unfolding rates will be given by

$$k_{fold} = \frac{\sqrt{\kappa_u \kappa_b}}{2\pi} \beta D \exp(-\beta \Delta G_u^\ddagger)$$

and

$$k_{unfold} = \frac{\sqrt{\kappa_f \kappa_b}}{2\pi} \beta D \exp(-\beta \Delta G_f^\ddagger) \quad (2.2)$$

Kramers' theory is a powerful tool for interpreting the time scales of

conformational dynamics, within the thermodynamic perspective of energy landscape theory. Moreover, equations (2.1) and (2.2) allow for the determination of D , from reconstructions of the energy landscape (16) and measurements of the folding and unfolding rates. The diffusion coefficient is an important parameter that reports on the effective roughness of the energy landscape and sets the “speed limit” for folding (13).

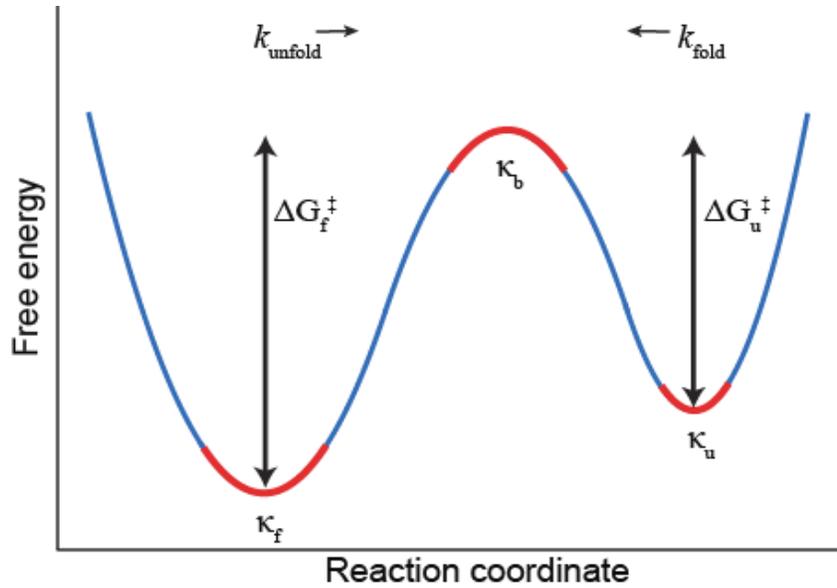


Fig 2.2: Kramers’ reaction rate theory. The parameters of the free energy landscape (curvatures $\kappa_{f/u}$, κ_b and barrier heights $\Delta G_{f/u}^\ddagger$), along with the conformational diffusion coefficient D , determine the rates for both folding and unfolding.

2.3 Transition paths

Energy landscape theory has provided a conceptual framework for folding that, for the proper choice of reaction coordinate, may allow the problem to be reduced to Brownian diffusion over a one-dimensional potential (17). When

reduced to this one-dimensional form, the folding rate is given by Kramers' reaction rate theory (18), and the inverse of the folding rate yields the folding time. This folding time however is not the time that a molecule spends actively folding, rather it is dominated by the time the molecule spends fluctuating within the stable states, as is evident by the exponential dependence on the free energy barrier height. The actual time that the molecule spends actively folding, known as the transition-path time, is much shorter (often orders of magnitude shorter) and relatively insensitive to the barrier height (19).

Transition paths (TPs) are portions of the diffusive trajectories, over the energy landscape, where the molecule passes through the high-energy transition states within the free energy barriers, separating the folded and unfolded conformations (Fig 2.3). It is along transition paths, that all of the actual folding takes place. Consequently, transition paths are the most important part of folding reactions, containing all of the critical mechanistic information about folding (17).

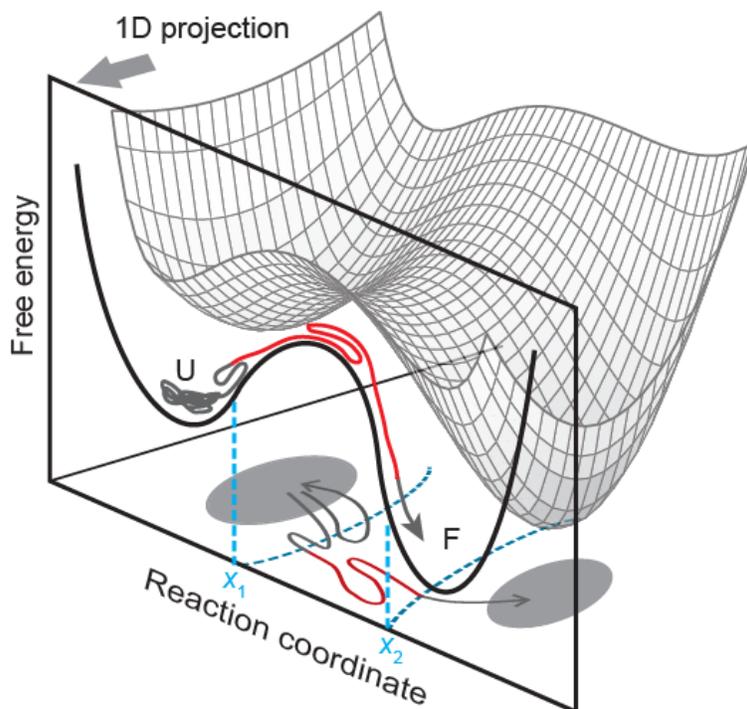


Fig. 2.3: Transition paths. Transition paths represent productive fluctuations (red) that cross the barrier between unfolded (U) and folded (F) states, and ignore non-productive fluctuations (grey). Motions over the multi-dimensional landscape are usually projected onto a 1D reaction coordinate.

2.3.1 Langevin dynamics

In this section, we build on the concept of transition paths as diffusive trajectories over the free energy landscape, by establishing a mathematical framework that describes the process. We can begin by assuming that the molecule obeys the laws of classical mechanics. This implies that the molecule's trajectory over the landscape is given by Newton's second law, $m\ddot{\vec{r}} = -\partial V(\vec{r})/\partial \vec{r}$, where m is the mass of the molecule, $V(\vec{r})$ is the landscape, and $\vec{r} = \sum_i x_i \hat{x}_i$ defines the position of the molecule on the multidimensional landscape. The molecule's total

energy is then conserved and given by $E = \frac{p^2}{2m} + V(\vec{r})$. The model can be simplified by projecting the multidimensional landscape down to a single dimension or reaction coordinate, as is done in force spectroscopy experiments. This projection comes with its own set of consequences that will be discussed later. The equation of motion on the 1D landscape is given by $m\ddot{x}(t) = -\partial V(x)/\partial x$, where x describes the molecule's position along the reaction coordinate.

An important aspect of folding that has been ignored thus far is the presence of a solvent. Biomolecules do not fold in a vacuum, rather folding reactions take place in the aqueous environment of cells. The presence of the solvent will alter the landscape by mediating intermolecular interactions and introduce frictional forces that dissipate the kinetic energy of the molecule (20). The frictional forces can be described by Stokes law and added into the model, giving $m\ddot{x}(t) = -\frac{\partial \tilde{V}(x)}{\partial x} - \gamma \dot{x}(t)$

, where $\tilde{V}(x)$ is the effective potential after adding the solvent and γ is the friction coefficient. The equation of motion still lacks one important aspect of folding, which is that the size of the molecule is very small, small enough that the motions of the individual solvent molecules will have a noticeable effect on the molecule's motion. Put more succinctly, we have neglected the diffusive part of the model. The solvent molecule motions are driven by thermal fluctuations and are therefore random in both their direction and magnitude. These solvent molecules will be constantly colliding with the folding molecule, and their effects can be modeled as a random, Gaussian distributed, time-dependent force $R(t)$, with mean zero (i.e.

$\langle R(t) \rangle = 0$). Adding this random force to the model yields

$$m\ddot{x}(t) = -\frac{\partial \tilde{V}(x)}{\partial x} - \gamma \dot{x}(t) + R(t). \quad (2.3)$$

This equation is the well-known Langevin equation, a stochastic differential equation central to the theories of transition path dynamics. As such, this equation should be explored more deeply.

First, the temporal behavior of the random force $R(t)$, must be considered. Due to its random nature and lack of any reason to believe that the value of $R(t+\delta t)$ is somehow determined from the earlier value $R(t)$, it can be modeled as a Markovian process. That is to say, that it is memoryless, with autocorrelation function $\langle R(t)R(t') \rangle = A\delta(t-t')$, where A characterizes the strength of the random force. The value of A can be determined by demanding that the model be consistent with equilibrium statistical mechanics. Specifically, demanding that the kinetic energy of the molecule be Maxwell-Boltzmann distributed (20). This requirement means that $\langle m\dot{x}^2 / 2 \rangle = k_B T / 2$ and is true even in the case where there is no potential (i.e. $\tilde{V}(x) = 0$). The Langevin equation can then be recast in terms of the velocity.

In the case of no potential, this gives $m \frac{d\dot{x}}{dt} = -\gamma \dot{x}(t) + R(t)$. Solving this equation

for the velocity yields, $\dot{x}(t) = \dot{x}(0)e^{-\gamma t/m} + \frac{1}{m} \int_0^t e^{-\gamma(t-t')/m} R(t') dt'$. In the limit of large

t , the first term decays to zero. Squaring the velocity expression in this long-time

limit and taking the average of \dot{x}^2 , gives $\langle \dot{x}^2 \rangle = A / 2m\gamma$, and comparing this to the

Maxwell-Boltzmann requirement, it becomes evident that $A = 2\gamma k_B T$ (20).

The Langevin description can be further simplified by returning to the earlier point made about the system being aqueous and the effects of the solvent on the molecule's dynamics. Folding reactions occur in the low Reynolds number limit, meaning that the dissipative frictional forces are so much larger than the inertial forces, that molecule effectively stops immediately after a force ceases to act on it. This then implies that the system is overdamped. Including this over-dampening in the Langevin model requires setting the inertial term to zero, yielding $\gamma \dot{x}(t) = -\frac{\partial \tilde{V}(x)}{\partial x} + R(t)$. The Langevin equation is now expressed in terms of the physical parameters of the molecule + solvent system, yielding an equation, albeit a stochastic one, that describes transition path trajectories during folding reactions. The caveat of this Langevin description is the fact that it is non-deterministic and therefore of little use in describing any particular transition path.

What is needed, is an expression that allows for the calculation of the probability that the molecule be at a specific position, at a specific time, given that the molecule's trajectory is generated by the Langevin equation. Such an equation exists and is the well-known Fokker-Planck equation:

$$\frac{\partial}{\partial t} P(x, t) = \frac{\partial}{\partial x} \left(\frac{1}{\gamma} \frac{\partial \tilde{V}(x)}{\partial x} P(x, t) \right) + \frac{\partial}{\partial x} \left(\frac{k_B T}{\gamma} \frac{\partial}{\partial x} P(x, t) \right). \quad (2.4)$$

This equation also conforms to the demand that the model be consistent with equilibrium statistical mechanics as its stationary solution is given by the Boltzmann distribution $P(x) \propto \exp(-\tilde{V}(x)/k_B T)$ (20). Making use of the Einstein-Smoluchowski relationship, the friction coefficient can be replaced with the more intuitive diffusion coefficient $D = k_B T / \gamma$, in both the Langevin and Fokker-Planck

equations.

The diffusion coefficient is a highly important parameter in the context of folding. It determines the speed at which the molecule samples the energy landscape and therefore the speed at which dynamical processes occur. It is through D that Kramers' theory connects the thermodynamics of landscapes to the kinetics of folding. Thus far, no assumptions have been made about the position-dependence of the diffusion (or friction) coefficient, but it is here where the first consequence of projecting the full multidimensional landscape down to a single reaction coordinate, becomes important.

On the full multidimensional landscape, the molecule samples the landscape at a constant rate, that is to say, that D is a constant everywhere, but when the landscape is projected down to a single dimension, any component of the molecule's velocity orthogonal to the projection is lost. These lost components of the velocity cause the molecule to appear to be progressing slower along the projected coordinate, at any position where the off-axis motions dominate, resulting in an apparent position dependence of the diffusion coefficient. This requires the introduction of a position-dependent diffusion coefficient, with an unknown functional form, $D(x)$. It should be noted, however, that for some molecules, with a well-chosen projection of the landscape, D may remain approximately constant (14, 15). This approximation is widely used in folding studies.

Finally, recognizing that the effective potential is actually the free energy, allows for the replacement of $\tilde{V}(x)$, with the free energy landscape $G(x)$. The two equations underlying the statistical dynamics of the transition path trajectories are

thus given by (20):

$$\dot{x}(t) = -\frac{D(x)}{k_B T} \frac{\partial G(x)}{\partial x} + \frac{D(x)}{k_B T} R(t) \quad (2.5)$$

$$\frac{\partial}{\partial t} P(x, t) = \frac{\partial}{\partial x} \left(\frac{D(x)}{k_B T} \frac{\partial G(x)}{\partial x} P(x, t) \right) + \frac{\partial}{\partial x} \left(D(x) \frac{\partial}{\partial x} P(x, t) \right) \quad (2.6)$$

These two equations are the basis of the major theories presented thus far and indeed, are the basis of nearly every theory (not including the optics) discussed and tested as part of this thesis. Kramers' reaction rate theory can be derived starting from the Langevin equation and the energy landscape theory of folding begins with the assumption that the flow between states in the landscape are described by the Fokker-Planck equation (21).

2.3.2 Most probable paths

Following the works of Faccioli et al, Makarov, and Cossio et al (22–24), we now investigate the most probable paths over an energy landscape. The solution to Eq. 2.6, subject to the boundary conditions $x(t_i) = -L$ and $x(t_f) = L$, with a constant D , can be expressed as a path integral (22)

$$P(L, t_f | -L, t_i) = e^{-G(x_f) - G(x_i) / 2k_B T} \int_{-L}^L \mathfrak{I}x(\tau) e^{-S_{eff}[x] / 2D}, \quad (2.7)$$

where $\mathfrak{I}x(\tau)$ denotes integration over all paths,

$$S_{eff}[x] = \int_{t_i}^{t_f} d\tau \left(\dot{x}^2(\tau) / 4D + V_{eff}[x(\tau)] \right), \quad (2.8)$$

is the effective action, and

$$V_{eff}(x) = \frac{D^2}{2} \left(\frac{1}{k_B T} \frac{\partial G(x)}{\partial x} \right)^2 - \frac{D^2}{k_B T} \frac{\partial^2 G(x)}{\partial x^2}, \quad (2.9)$$

is an effective potential. The most probable path of duration τ , from x_i to x_f is then the path that minimizes the effective action S_{eff} . This most probable path, known as the dominant transition path (DTP), must satisfy the energy conservation equation (23)

$$\ddot{x}(t) = \frac{dV_{eff}}{dx}. \quad (2.10)$$

For the case of a harmonic barrier, $G(x) = \Delta G^\ddagger - kx^2 / 2$, the solution to Eq.2.10 is given by (24)

$$x_{DTP}(t | \tau) = \frac{L \sinh(\beta k D(t - \tau / 2))}{\sinh(\beta k D \tau / 2)} \quad (2.11)$$

differentiating yields the velocity of the DTP

$$\dot{x}_{DTP}(t | \tau) = \frac{L \beta k D \cosh(\beta k D(t - \tau / 2))}{\sinh(\beta k D \tau / 2)} \quad (2.12)$$

2.3.3 Folding committors

The concept of folding committors is an important aspect of transition path theory and can be thought of as a measure of the kinetic distance between a given conformation and the reactant/product state (25). In the context of folding reactions, the reactants and products are the unfolded and folded states of the molecule. For this reason, the committor function is often referred to as p_{fold} , when applied to folding reactions. The value of p_{fold} at any conformation is the probability that the molecule will reach the folded state before the unfolded state. This same concept of the committor function has applications beyond folding reactions and is applicable to reactions involving two stable states A and B. The committor function for reaching state B before A is given by

$$\phi_B(x) = \frac{\int_{x_A}^x D(x')^{-1} e^{\beta G(x')} dx'}{\int_{x_A}^{x_B} D(x')^{-1} e^{\beta G(x')} dx'} \quad (2.13)$$

and for reaching state A before state B is $\phi_A(x) = (1 - \phi_B(x))$, or

$$\phi_A(x) = \frac{\int_{x_B}^x D(x')^{-1} e^{\beta G(x')} dx'}{\int_{x_A}^{x_B} D(x')^{-1} e^{\beta G(x')} dx'}. \quad (2.14)$$

Evidently then, for a two-state system separated by a barrier at position x^\ddagger the committors will be $\phi_A(x_A) = 1$, $\phi_A(x_B) = 0$, and $\phi_A(x^\ddagger) = 1/2 = \phi_B(x^\ddagger)$. It follows that the probability for a trajectory to be a transition path, given that the trajectory is at a position x , should be related to the committors. This is indeed the case and it can be shown that $p(TP|x) = 2\phi_B(x)(1 - \phi_B(x)) = 2\phi_B(x)\phi_A(x)$ for ideal diffusion in 1D (19). We can then use Bayes' theorem to relate $p(TP|x)$ to the transition path occupancy, $P(x|TP) = P(x)p(TP|x)/p(TP)$, where $p(TP)$ is the fraction of time spent in transition paths and $P(x)$ is the equilibrium position probability(19).

The Fokker-Planck equation can be also be written in terms of the committors

$$\frac{\partial}{\partial t} P(x, t) = \frac{\partial}{\partial x} D(x) \phi_B^2(x) e^{-\beta G(x)} \frac{\partial}{\partial x} \frac{e^{\beta G(x)}}{\phi_B^2(x)} P(x, t). \quad (2.15)$$

Starting from this expression of the Fokker-Planck equation, the average duration of a transition path can be calculated as (26)

$$\langle t_{TP}(x_A | x_B) \rangle = \int_{x_A}^{x_B} D(x)^{-1} e^{\beta G(x)} dx \int_{x_A}^{x_B} \phi_B(x') (1 - \phi_B(x')) e^{-\beta G(x')} dx' \quad (2.16)$$

In the case of a harmonic barrier, $G(x) = \Delta G^\ddagger - kx^2/2$, a constant diffusion coefficient, and large $\Delta x = x_B - x_A$, Eq.2.16 can be approximated as (24, 27)

$$\langle t_{\text{TP}}(x_A | x_B) \rangle \approx \frac{\ln(e^\gamma \beta k (\Delta x / 2)^2)}{D \beta k}. \quad (2.17)$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Eq.2.17 was first derived by Attila Szabo and has been utilized in several transition path studies, as such it appears several times throughout this thesis.

An interesting result can be found by evaluating Eq.2.12 for paths of duration $\tau = \langle t_{\text{TP}}(L | L) \rangle$, at $t = \langle t_{\text{TP}}(-L | L) \rangle / 2$ corresponding to the time when the barrier is crossed. This then yields an expression for the diffusion coefficient in terms of the barrier crossing velocity, the distance to the barrier, and the barrier height.

$$D = \frac{v(x^\ddagger) L \left[\left(2e^\gamma \beta \Delta G^\ddagger \right)^2 - 1 \right]}{8e^\gamma \beta^2 \Delta G^{\ddagger 2}} \quad (2.18)$$

A more elegant version of Eq.2.18 was provided by Makarov, by applying the approximation $\sinh(a) \approx \cosh(a) \approx e^a / 2$ to Eq. 2.12. This yields (23)

$$D = \frac{v(x^\ddagger) e^{\gamma/2}}{2} \sqrt{\frac{1}{\beta k}} \quad (2.20)$$

2.3.4 The Bryngelson and Wolynes energy landscape picture

The landscape view of folding describes folding as occurring through a multiplicity of routes down a funneled energy landscape. This is in contrast to the “pathway” model of folding in which folding proceeds through an obligate series of discrete intermediate steps (21, 28). A simplified view of this funneled landscape description is illustrated in Fig 2.4. The figure depicts a lattice structure where each stratum of the lattice represents an ensemble of partially folded structures and the flow between strata is impeded by kinetic trapping in local minima states on the

landscape. As the folding reaction progresses, the number of states, and hence the number of routes, within each stratum decreases until the single folded state is reached. Both the energy and entropy of the molecule are decreasing as the reaction progresses, resulting in a frustrated process. The average speed at which folding progresses down the funnel is determined by the gradient of the funnel itself (21). However, locally the stochastic motions of the molecule as it hops between local minima states, are superimposed onto the average speed. This stochastic hopping between local minima results in diffusive motion along the reaction coordinate. The depths and density of the local minima determine the local ruggedness of the landscape and are reflected in the diffusive motions through the diffusion coefficient (21). This description of folding, using diffusive coordinates along the energy landscape to account for the transient trapping within local minima, was introduced by Bryngelson and Wolynes in two landmark papers (29, 30).

The Bryngelson and Wolynes model divides the energies involved in folding into three groups: the energy associated with the state in which a residue resides, the energy associated with interactions between two residues that are close together in the polymer chain, and the energy associated with interactions between two residues that are close together in space but far apart in the polymer chain. Each of these energies is characterized by a value when the residue is in the optimal (native) configuration, and normally distributed with mean zero otherwise. This leads to a set of states whose total energy is normally distributed at any point along the reaction coordinate (29):

$$g(E, \rho) = [2\pi\Delta E(\rho)^2]^{-1/2} \exp\left\{\frac{-[E - \bar{E}(\rho)]^2}{2\Delta E(\rho)^2}\right\} \quad (2.20)$$

where ρ is the reaction coordinate, $\bar{E}(\rho)$ is the average energy, and $\Delta E(\rho)$ is the standard deviation of the energies. From its initial state, the molecule follows a metropolis dynamics scheme, meaning it can only change its state by moving to a connected state in the landscape. When the molecule reaches a local minimum energy state, this state will act as a micro-well transiently trapping the molecule. These local minima energy states are a subpopulation of the full distribution and therefore themselves have a normal distribution (29)

$$g_{LM}(E_0, \rho) \propto \frac{1}{[2\pi\Delta E(\rho)^2]^{1/2}} \exp\left\{-\frac{(E_0 - \bar{E}(\rho))^2}{2\Delta E(\rho)^2}\right\} \quad (2.21)$$

where E_0 is the energy of a local minimum state. Each local minimum will be connected to a population of states with $E > E_0$ and the distribution of these connected states is given by

$$g_{conn}(E, \rho) = \frac{g(E, \rho)}{\int_{E_0}^{\infty} g(E, \rho) dE} \quad (2.22)$$

The rate of the molecule leaving the local minimum state is the sum of the rates for the molecule going from the local minimum to the states connected to the local minimum state (29) and given by

$$R = R_0 \sum_i \exp(-\beta(E_i - E_0)) \quad (2.23)$$

where E_i is the energy of each connected state. The average value of R can be calculated from Eq.2.21 and using the asymptotic expression of the error function yields the average rate of escape from a local minimum

$$\bar{R}_{LM}(E_0, \rho) = R_0 N v \exp\left\{\beta(E_0 - \bar{E}) + \beta^2 \Delta E^2 / 2\right\} \quad (2.24)$$

where N_V is the number of states connected to the local minimum. The distribution of average escape rates from a local minimum state is

$$P_{LM}(R, \rho) = \frac{g_{LM}(E_0, \rho)}{\left. \frac{\partial \bar{R}_{LM}}{\partial E_0} \right|_{\bar{R}_{LM}(E_0, \rho) = R}} \quad (2.25)$$

Computing Eq.2.25 yields

$$P_{LM}(R, \rho) = \left(\frac{1}{(2\pi)^{1/2} R_0 \beta \Delta E(\rho)} \right) \exp \left\{ - \frac{\log^2 \left[\frac{R}{\bar{R}_{LM} \exp\{-\beta(E_0 - \bar{E})\}} \right]}{2\beta^2 \Delta E(\rho)^2} \right\}. \quad (2.26)$$

The probability density for the molecule to remain in a local minimum for a time t is given by

$$P_{LM}(t, \rho) = \int_0^\infty P_{LM}(R, \rho) R \exp(-Rt) dR \quad (2.27)$$

The associated diffusion coefficient is given by (29)

$$D(\rho) = \frac{\delta \rho^2 \lambda(\rho)}{2} \left[\left\langle \frac{1}{R} \right\rangle(\rho) \right]^{-1} \quad (2.28)$$

where $\delta \rho$ is the minimal increment along the reaction coordinate, $\lambda(\rho)$ is the probability for leaving the group of states at ρ , and $\langle 1/R \rangle(\rho)$ is given by

$$\left\langle \frac{1}{R} \right\rangle(\rho) = \int_0^\infty P_{LM}(R, \rho) \left(\frac{1}{R} \right) dR \quad (2.29)$$

Evaluating Eq.2.29 and using the asymptotic expression for the error function at large argument yields

$$D(\rho) = D_0 \exp[-\beta^2 \Delta E(\rho)^2] \quad (2.30)$$

where D_0 is the intrinsic diffusion coefficient on a smooth landscape.

The folding times and rates predicted by the theory are consistent with Kramers' theory, and hence, the theory provides a quantitative description of folding at all observable timescales, from macroscopic folding rates to microscopic rates of bond angle rotations (21). While measurements of the longer folding times have been shown to be consistent with the theory, validation of the theory at the shortest timescales remains elusive, requiring high precision measurements of the rates or waiting times, associated with crossing the microscopic barriers.

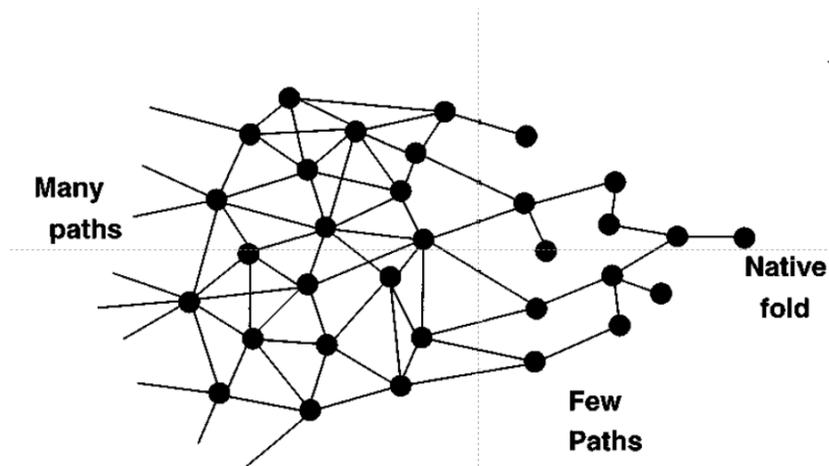


Fig. 2.4: Lattice representation of funneled energy landscapes. The points on the lattice represent the conformational states of the molecule as it folds. As the folding reaction progresses, the number of states, and hence the number of paths, decreases until the single folded state is reached. Figure adapted from Socci *et al* 1996 with permission.

Chapter 3

Optical tweezers

The original experiments performed as part of this thesis were all done using optical tweezers. This chapter provides an overview of the theory of optical trapping, the design specifics of the optical tweezers used in our lab, and how the optical tweezers are calibrated for reliable force detection.

3.1 Theory of optical trapping

An optical trap (OT) is an apparatus that makes use of the interactions between light and matter to harness the photon momentum of a laser beam, in order to hold and manipulate, microscopic dielectric particles in three dimensions. The first optical traps were pioneered by Arthur Ashkin at Bell Labs in the early 1970s and earned Ashkin the Nobel Prize in Physics in 2018. Ashkin's initial works demonstrated that optical forces could be used to push and levitate dielectric particles in both liquid and gaseous media (31, 32), eventually leading to the development of gradient force optical traps, commonly referred to as optical tweezers (OTs), capable of stably trapping and moving micron-sized dielectric particles in three dimensions (33, 34).

The basic requirement of an OT is a collimated laser beam, focused by a high-numerical-aperture (NA) lens. The focusing lens causes the collimated beam to form a double cone of light. The focal point, where the two cones meet, is a diffraction-limited spot and the gradient of the light intensity will generate a potential well at the focal point. Small dielectric particles (usually polystyrene

beads) which happen to pass close enough to the potential well will experience a net restoring force in the direction of the focal point. For small displacements from the focal point, this restoring force scales linearly with displacement and can be approximated as a Hookean-spring. Because the particle experiences this restoring force for displacements in all directions, the particle is confined or trapped, near the focal point of the laser (34). There are two regimes generally considered when calculating the force imparted by an optical trap: the Rayleigh regime, valid when the diameter of the dielectric object is much smaller than the wavelength of the laser, and the Mie regime, valid when the diameter of the dielectric object is much larger than the wavelength of the laser. Optical trapping of biological molecules places certain practical constraints on the system in order to avoid damaging the sample being studied. Near-infrared wavelengths near 1 micron are typically used for trapping biological molecules as they generally inflict minimal damage to the biological sample (35). Furthermore, the nature of the trapping force imposes constraints on the size of beads that can be trapped, effectively limiting the bead diameter to a range from a few hundred to a few thousand nanometers (34). Together these constraints imply that the criteria for neither the Rayleigh nor Mie regime are fully met. However, these two regimes can be applied approximately and provide important physical insights into the physics of optical trapping.

Let us first examine the small bead Rayleigh regime. Because the bead is very small, the electric field that it experiences due to the laser is nearly uniform over the extent of the bead. The bead then acts as a point dipole within the nearly uniform electric field of the laser. The forces acting on the dipole can be separated into two

components. The first of these is the so-called scattering force F_{scatt} , which results from the absorption and re-radiation of the electromagnetic waves of the laser. For a sphere of radius a , F_{scatt} is given by: $F_{scatt} = I_0 \sigma n_m / c$, where I_0 is the laser intensity, n_m is the index of refraction of the medium, c is the speed of light in vacuum, and σ is the scattering cross-section given by: $\sigma = \frac{128\pi^5 a^6}{3\lambda^4} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2$,

where λ is the wavelength of the light and $m = n_p/n_m$ is the ratio of the index of refraction of the particle to that of the medium (34, 36). The scattering force is in the direction of the propagation of the laser light and is proportional to the laser intensity, as well as to the radius of the bead to the sixth power. Because F_{scatt} acts only in one direction it is unable to trap an object. Stable trapping, therefore, requires a second component, the so-called gradient force F_{grad} . F_{grad} arises from the Lorentz force acting on the induced dipole within the inhomogeneous EM field of the focused laser beam and is given by: $\vec{F}_{grad} = (\vec{p} \cdot \vec{\nabla}) \vec{E}$, where $\vec{p} = \alpha \vec{E}$ is the dipole moment of the bead, with polarizability α , in electric field E . The gradient force experienced by the bead in a steady state is the time-averaged value of this force given by (34):

$$\langle \vec{F}_{grad} \rangle = \langle (\vec{p} \cdot \vec{\nabla}) \vec{E} \rangle = \frac{2\pi m_m}{c} \left(\frac{m^2 - 1}{m^2 + 2} \right) a^3 \vec{\nabla} I_0 \quad (3.1)$$

The gradient force is proportional to the gradient of the laser intensity and the radius of the bead to the third power and acts as a restoring force directed toward the focal point, (Fig 3.1) for the case of $m > 1$. Thus, when the bead is displaced from the focal point, the gradient force pushes it back toward the focal point, causing the

bead to be trapped. It is apparent from these equations that F_{scatt} and F_{grad} oppose each other along the direction of the laser propagation and hence the true location of the trap center lies slightly offset from the focal point of the laser, where the magnitude of the two forces are equal.

In the Mie regime, the particle is much larger than the laser wavelength and a treatment based on geometric optics is valid (37). Calculating the force exerted on the bead in this regime requires first calculating the force of a single ray acting on the bead and then summing over the entire laser beam (37). The result, first derived by Roosen and colleagues gives:

$$\begin{aligned} \vec{F} = & \frac{n_m P}{c} \left(1 + R \cos 2\theta - \frac{T^2 [\sin(2\theta - 2\phi) + R \cos 2\theta]}{1 + R^2 + 2R \cos 2\phi} \right) \hat{k}_{scatt} \\ & + \frac{n_m P}{c} \left(R \sin 2\theta - \frac{T^2 [\sin(2\theta - 2\phi) + R \cos 2\theta]}{1 + R^2 + 2R \cos 2\phi} \right) \hat{i}_{grad} \end{aligned} \quad (3.2)$$

where θ and ϕ are the angles of incidence and refraction, R and T are the Fresnel reflection and transmission coefficients at the bead surface, P is the power of the beam, and k_{scatt} and i_{grad} denote the scattering and gradient components of the force (38–40). In contrast to the Rayleigh regime, this result shows no dependence on the size of the bead, but shows a strong dependence on the high numerical aperture of the focusing lens, with highly converging rays contributing disproportionately to the force (40).

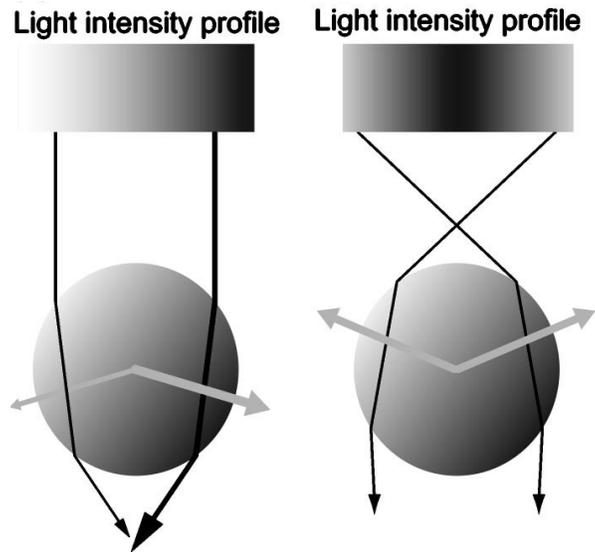


Fig. 3.1: Ray optics view of optical trapping. (left) When the light intensity is unbalanced (indicated by larger black arrow) the transfer of momentum results in a net force (indicated by larger grey arrow) pushing the bead toward the higher intensity region. (right) When the bead is displaced axially from the focal point, the transfer of momentum results in a net force toward the focal point. Adapted from Neuman and Block 2004 with permission.

3.2 Optical tweezers design

3.2.1 Trapping laser

The central component of any optical tweezers device is the trapping laser. The laser must have excellent stability in both pointing and power, enough power to provide sufficient trapping force and lase at a wavelength that will not damage the sample being studied. In optical tweezers experiments, the sample typically consists of beads and molecules in an aqueous solution, contained within a glass flow cell. The wavelength of the laser should therefore be chosen such that it has a low absorbance when passing through water, in order to avoid heating the water

and driving thermal motions within the sample. As it turns out, water has a minimum in its absorbance spectrum near one micron (Fig 3.2) (40). For this reason, as well as cost and availability, optical tweezers commonly use neodymium-doped yttrium aluminum garnet (Nd:YAG) or neodymium-doped yttrium orthovanadate (Nd:YVO₄) lasers that emit light at 1064 nm. The trapping laser used in our lab is a 5 Watt Nd:YVO₄ diode-pumped solid-state laser emitting a polarised beam in the fundamental (Gaussian TEM₀₀) mode. Measurements of the pointing stability of this laser indicate that pointing fluctuations occur on the order of 5 μ rad over the course of an hour. This stability is a highly important characteristic for optical tweezers, as fluctuations in the beams propagation direction can lead to undesired motions of the bead, resulting in artificial signals that are indistinguishable from the motions of the molecule being studied.

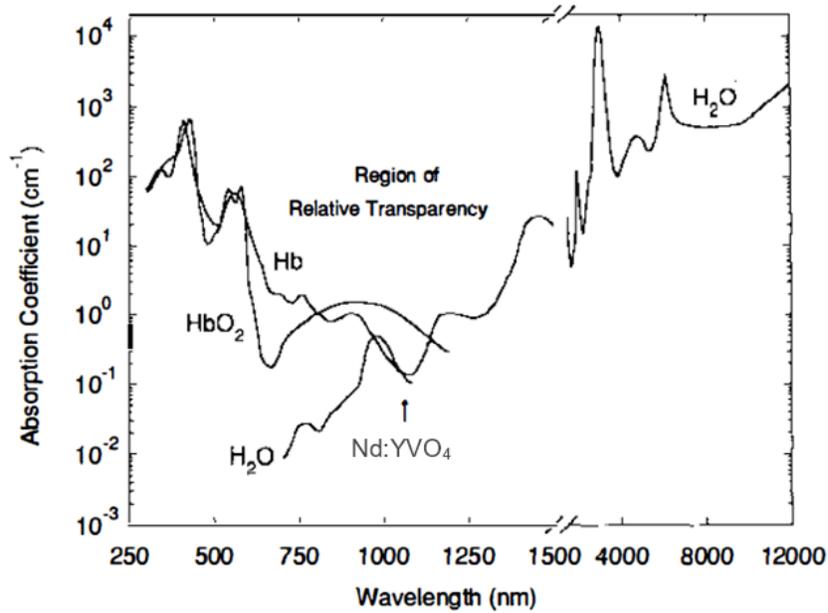


Fig. 3.2: Absorption spectrum of water. The measured absorption spectrum of water shows a local minimum in absorption in the near infrared at wavelengths close to the 1064 nm produced by an Nd:YVO₄ laser, indicating that the laser will have little effect on biological samples. Adapted from Svoboda and Block 1994 with permission.

3.2.2 Optics

Upon exiting the lasing cavity the trapping beam enters the free space of the optics box which houses most of the optical elements used to shape, direct, and steer the beam. The optics box is designed to minimize air currents and dust which may adversely affect the beam quality, as well as provide a convenient platform for attaching optical elements. The first of these optics that the beam encounters is a Faraday isolator. The isolator prevents any reflections of the beam from re-entering the lasing cavity and potentially causing mode hopping of the laser. The beam then passes through a set of telescoping lenses used both to expand the beam (2:1 expansion) and relay the position of the back focal plane. Next is a combination of

a half ($\lambda/2$) waveplate and polarizing beam splitters (pol cubes). The first waveplate rotates the polarization of the beam such that upon encountering the first pol cube the beam is split into two beams of equal power with orthogonal polarizations. The beams then briefly follow separate optical paths allowing for independent steering.

Having two independent traps allows for differential measurements and greatly reduces the effects of instrument noise. Because both traps originate from the same source and follow much of the same optical path, much of the noise introduced to the beam will affect both traps equally and hence cancel out during a differential measurement. After splitting, the beams are then steered using an electro-optic deflector (EOD) for the x-axis and an acousto-optic deflector (AOD) for the y-axis. Both deflectors act as crystals with tunable indices of refraction, allowing the beam to be deflected by prescribed angles. This angular deflection is turned into a translation of the beam by making use of two conjugate planes: the image plane located at the focus of the objective lens (i.e. the trap position) and the back focal plane located at the focus of the condenser lens. Because the two planes are conjugate, a rotation in one plane results in a translation in the other. The steering elements are therefore placed in the back focal plane, allowing for translation within the x-y image plane at the trap location. A series of Keplerian telescopes are used to relay the back focal plane throughout the instrument. The use of both EODs and AODs is done for practical reasons, EODs have much higher bandwidth and transmittance than AODs but they also have a significantly lower range of deflection angles. Furthermore, AODs suffer from non-linearity in their frequency response, known as AOD wiggles (41). These wiggles result in small

deflections during trap movements that may be interpreted as experimental signals. Although it may sound practical to build a system using EODs to steer both x and y-axis, the alignment EODs is notoriously difficult, and aligning such a system poses a technical hurdle best avoided if possible. Another advantage of having AODs is that they allow for control of the amount of light passing through them. This ability to tune the amount of light reaching the sample, and hence the stiffness of the trap, is important as different types of measurements require different trap stiffness's. For these reasons the EODs are used in the pulling axis – the axis along which force is applied to the molecule – while the AODs are used only to initialize the trap positions and control the trap stiffness, but remain stationary during measurements.

After passing through the steering elements the two beams pass through another pol cube that recombines the beams before reflecting off of a dichroic mirror and exiting the optics box. Next, the beams pass through a Keplerian telescope, used to relay the image plane through the objective lens of the microscope. After the objective lens, the beams enter the sample, where they perform their intended function of trapping. The light scattered off of the sample then passes through the condenser lens of the microscope, before finally ending at an IR filter.

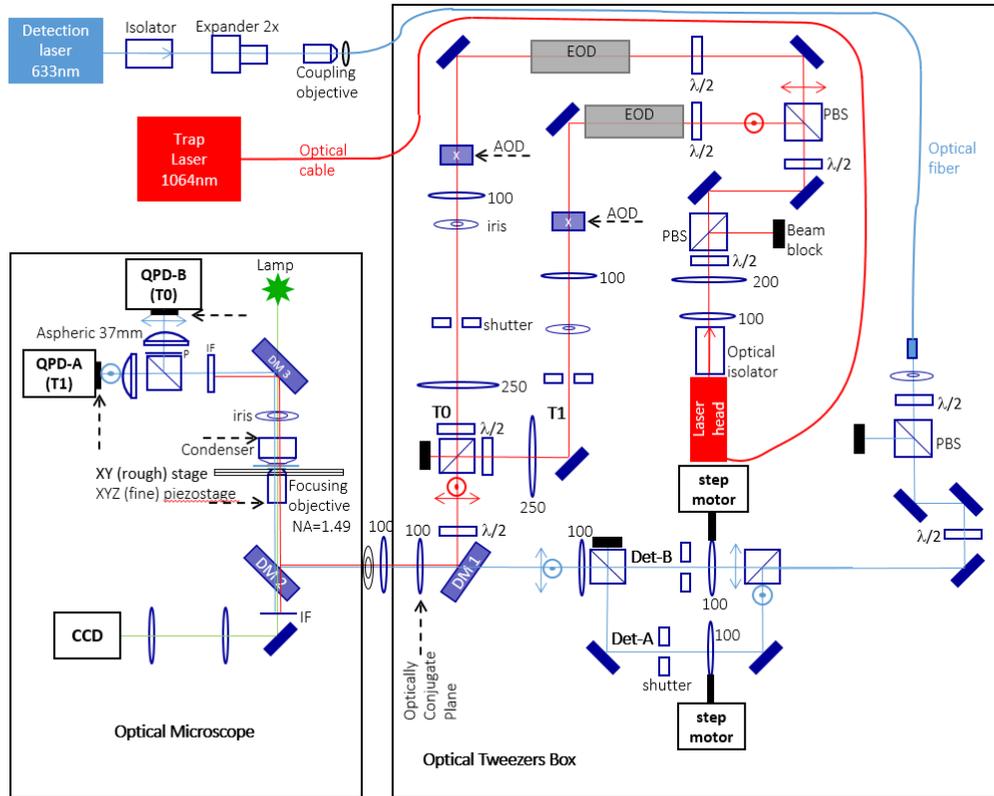


Fig 3.3: Optics layout. A schematic diagram of the optical layout of the optical tweezers instrument used for measuring transition paths.

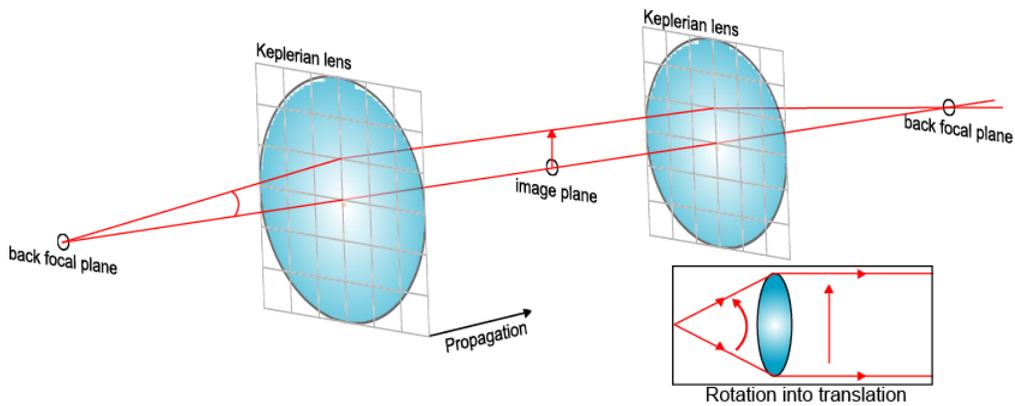


Fig 3.4: Back focal plane geometry. Rotations in the back focal plane result in translations in the conjugate image plane. A simple Keplerian telescope can be used to relay the locations of the two conjugate planes throughout the instrument. Adapted from Shaevitz 2006 (unpublished)

3.2.3 Position detection

To detect the positions of the beads within the trap, we use a separate laser. The detection laser is a 7 mW HeNe laser with a wavelength of 633 nm. The detection laser is located outside of the optics box and is fiber-coupled into a single-mode fiber using a 3-axis fiber coupler. The coupling of the laser into the fiber converts any pointing instabilities of the laser into power fluctuations which can be corrected by normalization of the detection signal. The detection beam then exits the fiber inside of the optics box where, like the trapping laser, it is split into orthogonal polarizations to allow for independent steering. Because the detection beams remain stationary during measurement, there is no need to actively steer them. Steering only needs to be done to align the detection beams, with the equilibrium positions of the trapping beams. This is done using lenses, mounted on motorized 3-axis stages, located at the back focal plane. After steering, the detection beams are recombined and pass through the dichroic mirror that reflects the trapping beams out of the optics box. At this point the detection beams follow the same optical path as the trapping beams, passing through the objective lens and scattering off of the beads in the sample chamber, before being collimated by the condenser lens. The beams then pass through the IR filter and into a pol cube that again, splits them by polarization and directs each beam toward its own detector.

The light scattered off of the beads is collected by a quadrant photodiode (QPD) in order to determine the position of the bead in the trap. These detectors are PIN diodes split into quadrants. When the incoming light strikes the active area, the PIN junction generates a photocurrent by means of the photovoltaic effect, that is

proportional to the light intensity at the incident position. This photocurrent is then converted into a voltage via Ohms law, with a known resistance value. Because the detector is split into quadrants, it will generate a separate voltage for each quadrant. The x and y positions are then determined by calculating the difference in voltage between the left and right sides of the detector as well as the top and bottom halves of the detector.

The installation of QPDs was part of a recent upgrade that I performed on the optical tweezers system, in an effort to increase the sampling rate and decrease the level of electronic noise, present in the position detection system. As mentioned in section 1.3, transition paths take place on a timescale of only a few microseconds. This brevity necessitates a fast sampling rate for measuring the positions of the beads in the trap. Previously, the system used position-sensitive diodes (PSD), rather than QPDs for position detection. A PSD basically consists of a uniform resistive layer formed on the surface of a semiconductor substrate and a pair of electrodes on both ends of the resistive layer for extracting position signals. The active area has a PN junction that generates photocurrent when light strikes the PSD. The photocurrents are inversely proportional to the distance between the incident position and each electrode, allowing for the determination of the incident position on the detector. By using two layers of PSDs, that are orthogonal to each other, both the x and y positions can be determined simultaneously. Dual-layered PSDs have one serious drawback, the orientation of two charged surfaces, with a very short distance between them, introduces a capacitance into the system which severely affects the response time of the detector. As a result, PSDs tend to have

lower bandwidths than QPDs. The PSDs previously installed in our optical tweezers had a bandwidth of 400 kHz and were replaced with QPDs with a maximum bandwidth of 2.5 MHz, allowing a sufficient number of data points to be collected, during the TPs brief duration, to properly determine the TP trajectory. The criteria used when deciding on the detector specifications come from knowledge of the timescale of transition paths and the measurements of the instrument response time. Previous studies of the average transition-path time for several different molecules (discussed in detail in chapter 5), revealed that for the molecules studied, the average transition-path times ranged from a few to tens of microseconds (17, 27, 42–44). Furthermore, measurements of the time required for the instrument to respond to large-scale perturbations in the trap position (discussed in chapter 4), yielded an average response time of 6 μs (44). Because this average response time was measured for large-scale perturbations, it represents an upper bound for the time required for the instrument to respond to much smaller length changes, associated with molecular fluctuations. It was therefore decided that the new detectors should be capable of accurately oversampling this 6 μs timescale.

Oversampling has several benefits in signal processing. It can improve the signal-to-noise ratio of the measurements and help to avoid aliasing (45). In order to avoid aliasing, the signal should be sampled at or above, the Nyquist frequency, corresponding to twice the value of the highest frequency in the signal (45). The upper bound on the instrument time response corresponds to a lower bound on the Nyquist frequency of 333 kHz. As this frequency represents a lower bound on the required detection bandwidth, it was decided that detectors with much higher

bandwidth capabilities should be installed.

3.3 Stiffness calibration of the optical trap

As discussed in section 3.1, the force applied by an optical trap is proportional to the gradient of the light intensity as a function of the displacement from the trap center. For a Gaussian beam, the intensity gradient is approximately linear within a narrow range around the trap center, resulting in a harmonic force akin to a Hookean spring. In order to determine the force being applied by the trap one only needs to know two parameters: the position of the bead relative to the trap center and the stiffness of the harmonic force. The trap stiffness is therefore a critical parameter requiring frequent and reliable calibration. There are three commonly used methods of calibration: the power spectral density of bead fluctuations, the variance of bead fluctuations within the trap, and the displacement of the bead from the trap center due to Stokes' drag (34). Because each method is susceptible to different systematic errors, it is good practice to perform all three methods and average the results (46).

3.3.1 Power spectral density calibration

The trap stiffness can be determined from the power spectral density, otherwise known as the power spectrum, of the thermal motions of a trapped bead. The power spectrum is Lorentzian (Fig. 3.4), characterized by constant power at low frequency and inverse-square behavior above a roll-off frequency f_0 :

$$P(f) = \frac{k_B T}{\pi^2 \beta (f_0^2 + f^2)}, \quad P(f) = \rho^2 P_v(f) \quad (3.3)$$

where β is the drag coefficient for the particle (34, 46, 47). By fitting the measured power spectrum to Eq.3.3 (Fig. 3.4) and determining the value of f_0 , the trap

stiffness can be determined from the relation $\kappa = 2f_0\pi\beta$. This method requires accurate knowledge of β , which depends on the bead size and geometry, the viscosity of the fluid, and the proximity to the microscope slide or coverslip. To account for variations in bead size and geometry, this calibration should be carried out for several different beads and the measured values of the stiffness averaged.

The dependence of the drag coefficient on the proximity of microscope slide and coverslip is a particular concern for optical trapping experiments as they are typically conducted within a micron of either of these surfaces. This surface effect acts to perturb the effective viscosity of the fluid, resulting in a drag coefficient that deviates from the familiar Stokes expression for a sphere. The effective drag coefficient for the spherical bead can be approximated by Faxen's Law (34, 47):

$$\beta' = \frac{6\pi\eta a}{1 - \frac{9}{16}\left(\frac{a}{h}\right) + \frac{1}{8}\left(\frac{a}{h}\right)^3 - \frac{45}{256}\left(\frac{a}{h}\right)^4 - \frac{1}{16}\left(\frac{a}{h}\right)^5} \quad (3.4)$$

where η is the fluid viscosity, a is the bead radius, and h is the height of the bead above the surface. It is easy to see that in the limit that the height above the surface goes to infinity, the Stokes expression, $\beta = 6\pi\eta a$, is recovered.

In addition to measuring the stiffness, the power spectrum is also highly sensitive to issues that can affect the quality of the trap including mechanical vibrations, optical misalignment, and electronic noise. All of these issues produce non-Lorentzian spectra, making the power spectrum a useful diagnostic measure (48).

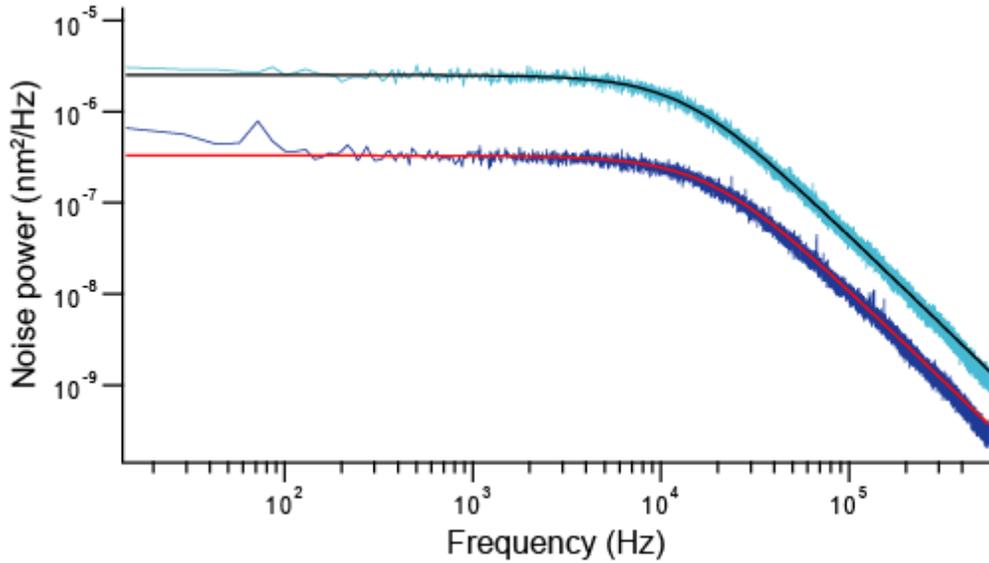


Fig 3.4: Power spectral density of a trapped bead The power spectrum for the X axis (teal) and Y axis (blue) of an optical trap are well fit by a Lorentzian (Eq. 3.3) (solid lines). The parameters of the fit determine the stiffness in each axis.

3.3.2 Variance calibration method

The variance of the beads fluctuations within the trap is simply the integral over frequency of the power spectrum and can be used to obtain the trap stiffness through the equipartition theorem. By the equipartition theorem, the thermal fluctuations of the bead in the harmonic potential of the trap are related to the thermal energy by: $\frac{1}{2}\kappa\langle x^2 \rangle = \frac{1}{2}k_B T$ (34, 40, 46, 47). This is the simplest of the three methods described as it depends only on the stiffness of the trap and the temperature, and is insensitive to the size or geometry of the bead. However, the variance is a biased estimator as the variance is always a positive value, and the presence of any extra noise (electrical, mechanical, etc.) will result in an underestimate of the trap stiffness.

3.3.3 Stokes drag calibration method

The most direct and most time-consuming method for determining the trap stiffness is the Stokes drag method (46). In this method, a trapped bead is subject to a viscous drag force produced by moving the bead through the fluid. Assuming that the bead remains in the harmonic portion of the trap potential, the drag force is given as: $F = \beta v = \kappa \cdot x(v)$, where v is the fluid velocity and $x(v)$ is the displacement of the bead out of the trap center. The fluid motion is generated by periodically driving the piezo sample stage laterally. For a bead subject to a triangular driving force of amplitude A_0 and frequency f , the motion is

$$x(t) = \frac{\beta A_0 f}{2\kappa} (1 - e^{-t\alpha/\beta}) \quad (3.5)$$

Due to the finite response time of the piezo stage, the exponential damping term is convolved with the response time of the stage. Therefore, a reliable estimate of the trap stiffness should use only the asymptotic value of Eq.3.5 (34).

Although this method can be difficult in practice, as beads can be pulled out of the trap or unwanted beads may fall into the trap mid-measurement, it also comes with certain advantages. The drag-force measurements are quite slow when compared to measurements of thermal motions and therefore require a much lower bandwidth for accurate detection. Furthermore, by increasing the amplitude or frequency of the stage motion the beads can be pushed to larger values of displacement. Measuring the trap stiffness as a function of the bead displacement provides a measure of the region over which the trap is harmonic (34, 48, 49).

3.4 Summary

This chapter describes the basic theory of optical trapping, the basic design,

and layout of the optical tweezers used in our lab, the methods used for calibrating the optical tweezers, as well as some of the upgrades that I made to the instrument in order to improve the quality of my data. It should be noted that optical tweezers can be very delicate and persnickety instruments, particularly when making measurements on the angstrom and microsecond scales, requiring near-constant adjustments to optical elements.

Chapter 4

Optical tweezers measurements

Having established a basic understanding of the physical principles governing optical tweezers, we now discuss some of the measurements typically used to study folding reactions with them, focusing on the types of measurements used in this thesis.

Optical tweezers can be used to measure systems both in and out of equilibrium. While measurements taken at equilibrium allow for interpretation with the well-known tools of equilibrium statistical mechanics, non-equilibrium measurements can be used to probe portions of the energy landscape rarely visited under equilibrium conditions.

There are two equilibrium measurement paradigms commonly employed with dual-beam optical tweezers: constant force measurements, in which the force applied to the molecule remains constant, and constant separation measurements, where the distance separating the two traps remains constant. A typical non-equilibrium measurement involves pulling on the molecule while linearly increasing the force applied. For all of these measurements, the locations of the beads are the observable that is recorded. Independent measurements of the bead locations allow for a straightforward calculation of the distance between the beads. When tension is applied to the system, changes in this distance correspond to changes in the end-to-end extension of the molecule + handle system. This end-to-end extension is the natural reaction coordinate for folding in optical tweezers measurements because it can be directly related to structural changes of the

molecule.

4.1. Equilibrium measurements

4.1.1 Constant force measurements

In constant force measurements, the molecular extension is measured while the force applied to the molecule is held at a constant value (Fig. 4.1). This can be done through either a feedback loop or by exploiting the anharmonic region of the optical trapping potential where the differential stiffness vanishes.

Feedback loops are typically implemented by measuring the instantaneous position of the trapped bead and then moving the trap to maintain a set displacement between the bead and the trap center. However, this type of force clamp is susceptible to two significant complications. First, the beam steering mechanism can exhibit nonlinear responses, leading to variable loads, which is an undesirable quality of a constant force measurement. Second, the finite response time of the feedback loop places significant restrictions on the bandwidth of measurements, greatly reducing the temporal resolution. Furthermore, the limited time response of the feedback system means that the force does not actually remain constant on short timescales. Instead, the force changes slightly in response to the very molecular motions that the experiment is attempting to measure. This ambiguity in the force makes interpretations of the statistics of the molecular motions difficult. These restrictions render the feedback loop inadequate for measuring systems requiring high bandwidth and spatial resolution, such as DNA hairpin unfolding events (49).

A better method developed by Greenleaf *et al* acts to passively maintain a constant force by exploiting the anharmonic region of the optical trapping potential

(Fig. 4.2). As discussed in the previous chapter, the force profile for a Gaussian beam optical trap is in the form of the derivative of a Gaussian curve. As the displacement of the bead from the trap center increases the force initially increases linearly (this is the harmonic region), before reaching a peak and rolling over. Near the peak of the force-displacement curve, there is a region where the force is approximately constant for small displacements. Therefore, by pulling a bead out to this zero stiffness regime, the force applied to the molecule will be effectively clamped and the value of the applied force can be tuned by adjusting the intensity of the laser light (49). In a dual optical trap setup, one bead is held in the zero stiffness regime of one trap, position measurements are made on this bead while force measurements are made on the other bead which is held in the harmonic region of the other trap.

Although constant force measurements were not performed as part of this research, the energy landscape reconstructions for DNA hairpins, obtained from passive constant force measurements, were used, prompting this brief discussion.

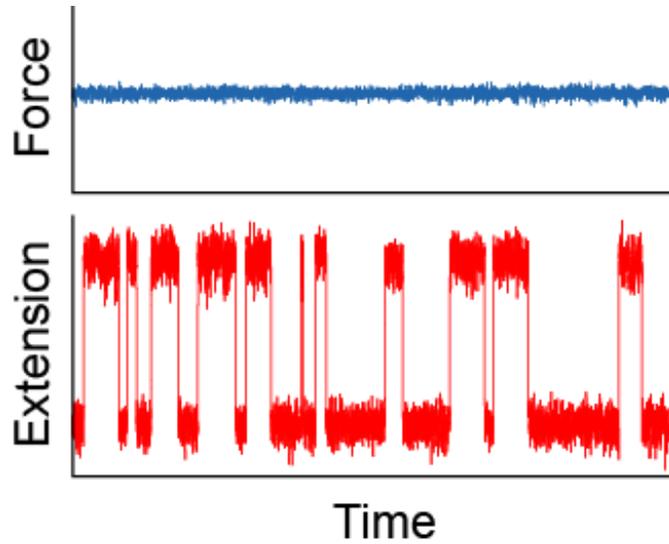


Fig 4.1 Constant force measurements. The force (blue) applied to the molecule is held constant. If the force is near $F_{1/2}$, the molecule will hop between the folded and unfolded states as evidenced by the abrupt changes in molecular extension (red). Adapted from Woodside and Block 2014 with permission

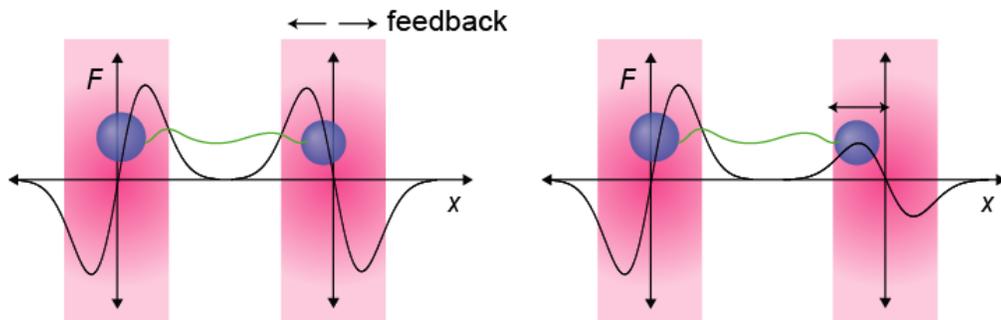


Fig 4.2 Active and passive force clamps. The force can be held at a constant value either actively via a feedback system (left) where the trap moves in response to changes in extension, or passively by keeping one bead in the zero stiffness regime of the trap (right).

4.1.2 Constant trap separation measurements

Constant trap separation measurements are conducted by maintaining a fixed distance between the centers of the two optical traps and measuring the distance

between the two trapped beads as a function of time. In this type of measurement, the applied force varies in response to changes in the molecular extension (Fig. 4.3). Constant trap measurements are better suited than constant force measurements, for measuring extremely fast events such as transition paths. While the low system stiffness of constant force measurements (owing to the requirement of operating one trap in the zero stiffness regime) leads to a large instrument response time, constant trap measurements can be done at much higher stiffness values and hence have much lower response times. The link between trap stiffness and response time is perhaps best illustrated through an analogy to a mechanical spring, where the time required for a signal to propagate from one end of the spring to the other, is proportional to the spring stiffness. The low response time of constant trap measurements makes them ideal for measuring rare and brief events such as transition paths. Hence this type of measurement makes up the bulk of the original data presented in this thesis.

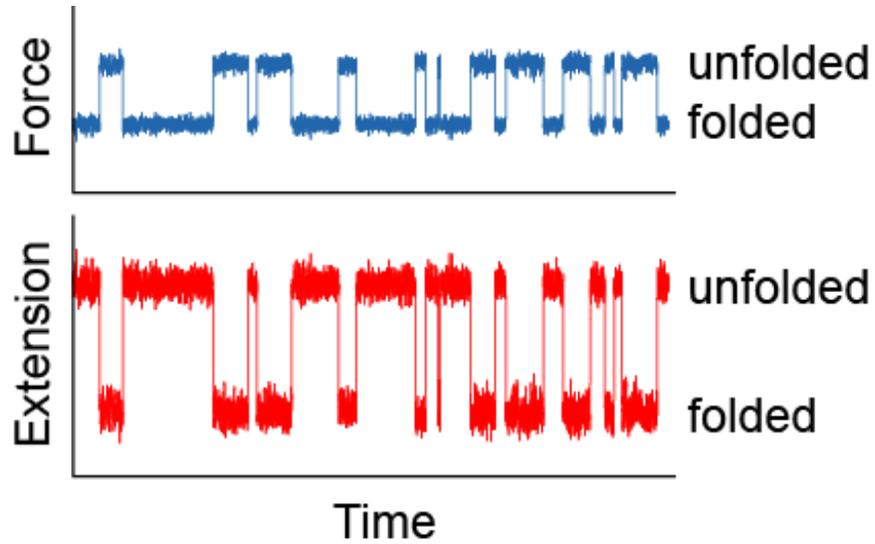


Fig 4.3 Constant position measurements. In constant position measurements the locations of the traps are held at constant positions while the force applied to the molecule is allowed to fluctuate (blue). If the average value of the force is near $F_{1/2}$, the molecule will hop between the folded and unfolded states as evidenced by the abrupt changes in molecular extension (red). Adapted from Woodside and Block 2014 with permission.

4.2 Non-equilibrium measurements

4.2.1 Force vs extension measurements

One of the most common types of measurements performed with optical tweezers is a force ramp measurement. This measurement typically involves pulling on a molecule with a linearly increasing force load while simultaneously recording both the end-to-end extension of the molecule and the instantaneous force being applied (Fig. 4.4). Plotting the applied force as a function of the molecular extension results in what is commonly known as a force-extension curve (FEC). When an FEC is performed on a folded polymer the force will increase monotonically with increasing molecular extension until structural components

begin to unfold. The structural components unfold in order from the least to most stable structures. Each time a structure unfolds it is accompanied by an abrupt increase in molecular extension, this, in turn, allows the beads to be pulled closer to the trap centers resulting in a rapid decrease in force. These unfolding events appear as distinct “rips” in the FEC. The FEC can also be run in reverse. By starting at a high force and decreasing the applied load, the refolding process is observed. When applied consecutively to the same molecule, hysteresis in the FECs may be observed as the molecule unfolds and refolds under the applied load. The presence of this hysteresis is indicative of the non-equilibrium nature of these measurements. FEC measurements reflect all elastic components of the experimental assay. This means that analysis of FEC measurements must take into account not only the stretching of the molecule under study but the dsDNA linkers and trap stiffness as well.

The choice of dsDNA for linkers was made in part because its elastic response to force is well understood in polymer physics. Indeed worm-like chain (WLC) models can account for both the elastic energy and entropy of polymer chains such as dsDNA. In the WLC model one evaluates the partition function associated with different polymer configurations, weighting each configuration by its corresponding elastic energy cost. Evaluating the partition function is highly non-trivial as it requires summing over all possible polymer configurations. Fortunately, the problem is highly simplified in both the low and high force limits, permitting a determination of the FEC properties in these limits. Furthermore, an interpolation formula connecting these limits can be derived, allowing for an evaluation of the

extension of the worm-like chain over the entire range of applied forces (50, 51).

This interpolation formula of the WLC model is given by:

$$F(x) = \frac{k_B T}{L_p} \left[\frac{1}{4 \left(1 - \frac{x}{L_c} + \frac{F}{K} \right)^2} - \frac{1}{4} + \frac{x}{L_c} - \frac{F}{K} \right] \quad (4.1)$$

where F is the applied force, x is the observed extension, L_p is the persistence length, L_c the contour length, and K the elastic modulus of the polymer. In the case of DNA hairpin unfolding the FEC is well described by Eq. 4.1 with parameters corresponding to that of dsDNA, at low force. However, when the hairpin unfolds the FEC will reflect the presence of the ssDNA of the unfolded hairpin. The persistence length, contour length, and elastic modulus of dsDNA have been measured extensively with values of, $L_p = 40\text{-}50$ nm, $L_c = 0.34$ nm/bp, and $K \sim 1000\text{-}1200$ pN (52–55). Given that these values are well known, fitting of FEC measurements to Eq. 4.1 can be used diagnostically to confirm the presence of a single molecule of interest and discriminate against the presence of multiple handle attachments or otherwise ill-formed constructs.

In the context of transition path measurements, FECs are used only for their diagnostic functionality. FECs are performed to confirm the quality of the trapped molecule before performing constant trap separation measurements to measure transition paths.

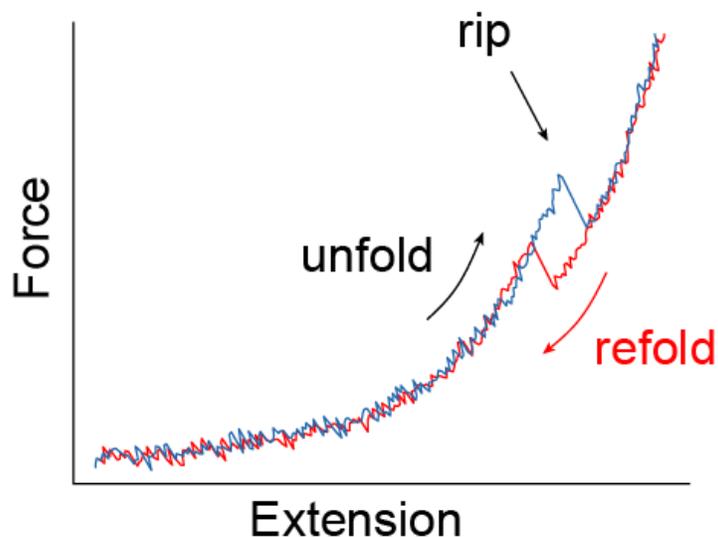


Fig 4.4 Force extension curves. In a force extension curve the force applied to the molecule is increased linearly (blue). Unfolding events result in rips in the curve. Running the process in reverse causes the molecule to refold (red) and hysteresis between the two curves indicates that the measurement is performed out of equilibrium. Adapted from Woodside and Block 2014 with permission.

4.3 Measuring transition paths

4.3.1 Requirements for measuring transition paths

Owing to their extreme brevity, transition paths have proven challenging to study experimentally. As a result, the study of TPs was until recently restricted to theory and simulation. In recent years, however, advances in single-molecule instrumentation and analysis have allowed direct measurement of TPs, opening up a new window on the microscopic conformational dynamics during folding. So what exactly are the requirements needed to measure TPs and what advances have allowed for their direct measurement? These questions are the focus of this section.

First and foremost is the fact that TPs are inherently stochastic. This poses a gigantic problem for traditional ensemble studies of folding because for one; the

folding of ensembles of molecules cannot be synchronized, and second; even if the entire ensemble could be forced to all fold/unfold at the exact same time, the stochastic nature of TPs dictates that each molecule in the ensemble would take a slightly different TP over the barrier. The result of such a synchronized ensemble experiment would be an ensemble average over all of the TPs. While such a study could still yield interesting and useful results, such as average transition-path times, it would be incapable of measuring individual transitions or yielding distributions. For these reasons, our first requirement for measuring TPs directly is the use of single-molecule approaches. This requirement is obviously satisfied by optical tweezers, but single-molecule studies of folding with optical tweezers have existed for decades, so clearly other requirements must exist.

With the use of single-molecule methods, the next requirement is very high spatiotemporal resolution (nanometer and microsecond). The spatial resolution is again satisfied by the use of optical tweezers, with sub-nanometer resolution being common for optical tweezers setups. It is the temporal resolution that poses a formidable challenge to measuring TPs with optical tweezers. TPs are extremely brief events, often only a few microseconds in duration. Measuring TPs, therefore, requires a system that has both a high sampling rate and a low time response. The upgrades to the detection system discussed in chapter 3 deal with the high sampling rate requirement. The requirement of a low time response is somewhat more difficult to achieve and subtle in its manifestation.

A time response sufficient for measuring TPs with optical tweezers requires highly stiff optical traps. This can be achieved through high-power, highly focused

lasers in conjunction with properly sized beads. Here is where things start to get slightly more complicated, recall from chapter 3 that the gradient force of an OT scales with the radius of the bead cubed. From this one might be tempted to think that larger beads should be used for measuring TPs as they lead to a higher trap stiffness, but a more massive bead will be less sensitive to the minute molecular fluctuations, leading to measurements that are dominated by the dynamics of the bead, rather than molecule (24, 56, 57). Furthermore, the frictional drag acting on the bead is proportional to the bead radius. Thus increasing the size of the bead tends to slow both the bead movement and instrument response time. This means that we need beads that are large enough to be trapped stiffly but small enough that they will not dominate the kinetics of the system. In our system we currently use an 820 nm diameter bead in one trap and a 600 nm diameter bead in the other trap, forming our dumbbell assay.

The instrument response time can be measured by inducing “transitions” in a reference construct and measuring the transit time. The reference construct consists of two beads tethered by dsDNA and is essentially the normal assay with the molecule removed. The “transitions” can be induced by abruptly jumping the position of one of the traps at a distance equivalent to the extension change in the folding of the molecule of interest. By doing this several times and measuring the time required for the molecule to transit between “states” we can build a distribution of response times. Taking the average of this distribution provides a reasonable estimate of the instrument response time. When this procedure was done on our OT the average response time was found to be 6 μ s (Fig. 4.5)(44).

Another aspect contributing to the temporal resolution of the measurements is the molecule itself. The molecule must unfold at a sufficiently high force such that a relatively high amount of tension can be maintained in the molecule + linkers construct, as higher tension results in faster propagation of the signal from the molecule, through the linkers, to the beads. For reference, the unfolding force for our poster child molecule, the DNA hairpin 30R50/T4, is approximately 14 pN (58).

As mentioned in the above sections, the transition-path measurements taken as part of this thesis were performed via the constant position method. The reason for this is twofold: first, constant position measurements allow for transition paths to be measured in equilibrium. This is an important point, as theoretical expectations for transition paths have been formulated in the context of equilibrium. Therefore, transition-path measurements must be made under equilibrium conditions if the aim is to test or draw comparisons to the established theories. Second, a relatively high trap stiffness can be maintained during constant position measurements. As we now know, this high stiffness allows for higher temporal resolution, a critical requirement for measuring TPs.

Even if all of the above considerations are accounted for, there are still potential issues that can prove disastrous when attempting to measure TPs. These issues come in the form of experimental artifacts, the effects of which must be thoroughly investigated before TP measurements can be considered as true representations of the molecular dynamics.

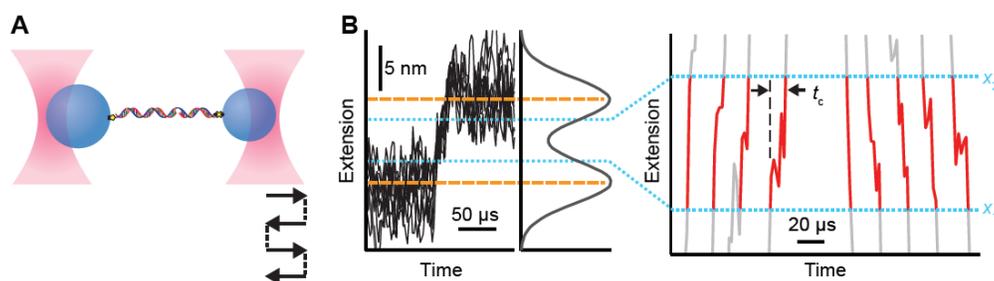


Fig 4.5 Instrument response time measurement (A) The response time of the optical tweezers can be measured using a reference construct consisting of DNA handles only. The construct is held in the traps and one trap is jumped abruptly back and forth to cause the extension of the molecule to change by an amount equivalent to the extension change in the folding of the molecule of interest (B) The time required for each “transitions” to cross the barrier region (x_1 to x_2), can be measured and averaged to determine the response time of the instrument. Adapted from Neupane et al *Science* 2016 with permission.

4.3.2 Instrumental artifacts in kinetic rates and transition-path properties

Studies of folding using SMFS generally require coupling of the molecule of interest to a force probe. For the optical tweezers measurements presented in this thesis, the force probe is the bead that is coupled to the molecule via dsDNA linkers. Due to the fact that the measurements report on the motion of the bead, rather than the motions of the molecule itself, the properties of the molecule must be inferred from the bead motion. This then implies that the observed dynamics are those of the entire bead+linker+molecule system, and may therefore be altered from the intrinsic dynamics of the molecule. Recent theoretical studies (24, 57, 59) have provided a framework that can be used to gauge the extent to which properties such as transition rates and transition-path times are altered during SMFS experiments. This framework has been applied by Neupane and Woodside (60), to two of the DNA hairpins studied as part of this thesis (30R50/T4 and 20TS06/T4) using the

same beads and linkers. The results of this study show that errors in the folding rates were only $\sim 20\%$, similar to the experimental uncertainty in the rates. They also found that measurements of transition paths using constant trap position measurements at high trap stiffness are in the low-artifact limit (60).

To quantify the errors in the observed transition rates Neupane and Woodside first performed constant force measurements of the DNA hairpins with the force set to $F_{1/2}$ and determined the observed transition rates, denoted k_{MA} . They then determined the rate k_A expected if the molecule were diffusing over the potential of mean force (PMF), via the application of Kramers' theory to the PMF. The PMF can be obtained directly from the extension trajectories via Boltzmann statistics and the implied diffusion coefficient was defined as $D_q = \langle \delta q^2 \rangle / \tau_A$, where δq is the deviation of measured extension from its average value within the stable folded or unfolded states and τ_A is the relaxation time obtained via single-exponential fits of the extension autocorrelation function within the same state. In order for the observed rate k_{MA} , to be a good estimate of the intrinsic molecular rate k_M , the expected rate k_A must be fast in comparison to k_M . This was indeed the case, with $k_A \gg k_{MA}$, indicating that $k_{MA} \approx k_M$, for both hairpins. The error in the observed rates caused by instrumental effects was then quantified using an expression derived by Cossio et al. (57)

$$[k_{MA}]^{-1} = \left[k_M \left(1 - [\beta \kappa_b \langle \delta q^2 \rangle]^{-1} \right) \right]^{-1} + [k_A]^{-1}, \quad (4.2)$$

where κ_b is the barrier curvature of the intrinsic energy landscape. This procedure showed that the instrumental effects on the rates were low with $k_{MA} \sim 20\%$ lower than k_A (60).

Instrumental artifacts can have large effects on the dynamics of transition paths, even when measurements are made in the regime where the effects on the rates are low (24). A thorough examination of these effects is therefore crucial when interpreting transition-path measurements, and failure to do so can lead to nonsensical interpretations of transition-path properties (61). Examining the extent to which instrumental effects alter the dynamics of transition-path measurements involves comparing D_q to D_x (24, 57), the diffusion coefficient along the intrinsic molecular free-energy landscape. In order for the effects to be small, the observed diffusion coefficient D_q should be fast in comparison to D_x . In order to examine the instrumental effects on the transition path measurements, Neupane and Woodside performed constant trap position measurements on the DNA hairpins and recalculated D_q from the extension fluctuations and autocorrelation function. It should come as no surprise that D_q was found to be much higher than in the constant force case, owing to the significant increase in the trap stiffness of the constant position measurements. D_x was then calculated through two independent methods: from the application of Kramers' theory to the energy landscape obtained after removal of instrument compliance through deconvolution; and the average transition-path time. Both methods returned the same values for D_x , which was found to indeed be lower than D_q for both hairpins, suggesting that the observed transition paths should be minimally distorted (60).

	30R50/T4	20TS06/T4
Constant Force		
D_q ($\times 10^5$ nm ² /s)	2.4 ± 0.2	4.0 ± 0.3
k_A (s ⁻¹)	105 ± 7	700 ± 100
k_{MA} (s ⁻¹)	3.3 ± 0.2	11 ± 3
k_M (s ⁻¹)	4 ± 1	15 ± 3
D_x ($\times 10^5$ nm ² /s) (via Kramers')	4.6 ± 0.5	5 ± 3
Constant Trap Position		
D_q ($\times 10^5$ nm ² /s)	6.0 ± 0.3	5.7 ± 0.3
D_x ($\times 10^5$ nm ² /s) (transition time)	4.4 ± 0.4	3.4 ± 0.8

Table 4.1: Kinetic parameters of DNA hairpins. Values represent the average of folded and unfolded values and errors represent the s.e.m. Adapted from Yu et al. 2012 with permission.

4.3.3 Reaction coordinate quality

Last but not least, is the requirement that molecular extension accurately reflects the dynamics of the molecule being studied. In the context of energy landscape theory, folding reactions are described as a diffusive search over a highly multidimensional energy landscape for the lowest energy folded state. When folding is monitored in experiments, only a single reaction coordinate is observed. In optical tweezers measurements, this single reaction coordinate is the end-to-end extension of the molecule. Thus, the observed reaction coordinate is a one-dimensional projection of the full multidimensional landscape. Although only a representative approximation of the full complexity, these 1D descriptions are widely used (as they are generally all that is accessible to experiments) and often give an adequate representation of the full dynamics of the reaction (14). Despite the success of these approximations, their suitability to any particular system is difficult to know without performing tests of the reaction coordinate quality.

Moreover, the use of a poor reaction coordinate can have critical implications for interpreting experimental results and may lead to incorrect conclusions (15). Testing of the reaction coordinate quality can only be done after TP measurements have already been taken and are one of the more difficult tests to perform, as it requires accurately reconstructing the molecular free-energy landscape and deconvolution of the effects of the beads and linkers from the landscape.

One such test of reaction coordinate quality was developed specifically for folding reactions. This test involves calculating the committor, $p_{\text{fold}}(x)$ – the probability that when the molecule is at a given position along the reaction coordinate x , it will reach the folded state before the unfolded state. In the case of a two-state system, a good reaction coordinate requires that $p_{\text{fold}}(x) = 1/2$ at the top of the barrier. Hence, if this requirement is not met, then the reaction coordinate is considered poor. However, the inverse of this statement is not true. In other words, all good reaction coordinates have $p_{\text{fold}}(x) = 1/2$ at the barrier, but so do some bad reaction coordinates.

A more definitive test of reaction coordinate quality is based on transition path statistics. This states that a reaction coordinate is good if the conditional probability that the molecule is on a transition path when it has extension x , $p(\text{TP}|x)$, should be highly peaked around the location of the free energy barrier, x^\ddagger , ideally reaching a value of $1/2$ at x^\ddagger (15, 19). The conditional probability $p(\text{TP}|x)$, can be calculated from the equilibrium trajectory by the Bayesian relation $p(\text{TP}|x) = P(x|\text{TP})p(\text{TP})/P(x)$, where $P(x)$ is the equilibrium distribution of extension values in the complete trajectory, $P(x|\text{TP})$ is the distribution of extension values

along only transition paths, and $p(\text{TP})$ is the fraction of time in the trajectory spent on transition paths. For a purely 1D diffusive process, this conditional probability can also be related to $p_{\text{fold}}(x)$, as $p(\text{TP}|x) = 2 p_{\text{fold}}(x)[1 - p_{\text{fold}}(x)]$.

This test was applied to transition path measurements of the DNA hairpin 30R50/T4 by Neupane et al (15). In their first attempt, they found that while $p(\text{TP}|x)$ was indeed peaked near the location of the barrier, its peak value was only 0.04, far lower than the expected value of 0.5. While this result may suggest that extension is a poor reaction coordinate for the hairpin, Neupane et al. explain that the test fails to take into account the effects of experimental noise, such as bead and linker fluctuations, potentially altering the result of the test.

In an effort to explore how the bead and linker fluctuations alter $p(\text{TP}|x)$, they then performed simulations of a molecule compliantly linked to a bead on a purely 1D landscape. Because the landscape is 1D, the reaction coordinate is by definition good and any deviations in $p(\text{TP}|x)$, from the expected result, can be attributed to the bead and linker effects. Performing the simulation with a range of different linker stiffness's, they found that the amplitude of the peak in $p(\text{TP}|x)$ was suppressed as the stiffness decreased. Noting that $p(\text{TP}|x)$ is primarily affected by the addition of statistical weight to $P(x)$ through the addition fluctuations of the bead and compliant linker, they then sought to correct for the compliance effects by using the intrinsic molecular extension distribution, $P_i(x)$ – which can be calculated by empirical deconvolution of the equilibrium extension distribution.

After applying the corrective procedure and recalculating $p(\text{TP}|x)$ for the hairpin data, they found that $p(\text{TP}|x)$ reached a maximum of ~ 0.45 at an extension

close to the barrier peak, x^\ddagger , thus validating the molecular extension as a good reaction coordinate for the DNA hairpin data. Applying the same test to the DNA hairpin 20TS06/T4, Neupane et al. found similar results.

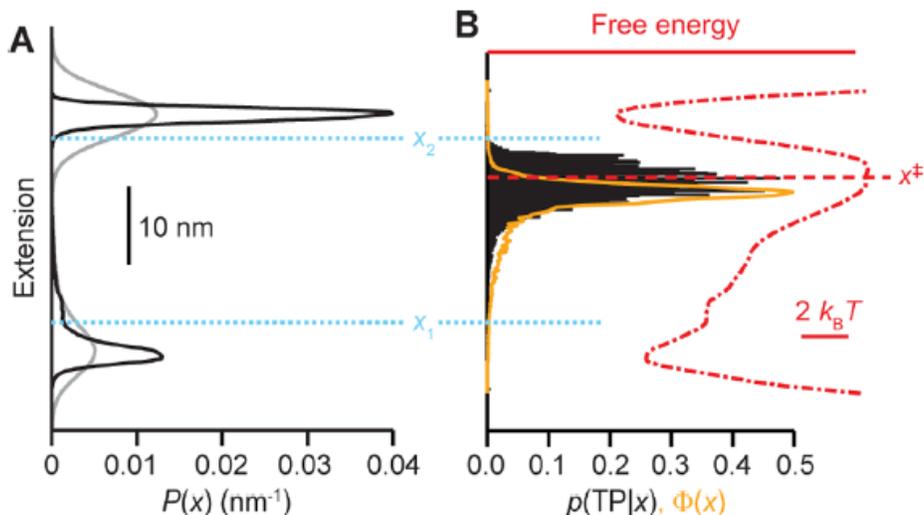


Figure 4.6: Reaction coordinate quality tests for DNA hairpin 30R50/T4.(A) The deconvolved extension probability (black) is peaked more sharply than the raw extension probability (grey). (B) The conditional probability distribution (black) and p_{fold} (yellow) are sharply peaked at x^\ddagger , the location of the barrier in the free energy landscape (red) indicating that the extension is a good reaction coordinate. Adapted from Neupane et al. 2015 with permission.

4.3.4 Specifics of transition-path measurements

The specifics of the transition-path measurements performed in this thesis are as follows: transition-path measurements were made at equilibrium at a force near $F_{1/2}$, the force at which the occupancies of the folded and unfolded states were similar, under conditions of constant trap separation with high trap stiffness (0.75–1.1 pN/nm in one trap, and 0.56–0.63 pN/nm in the other) (Fig 4.7). The data were sampled at 125–1,000 kHz and filtered online at the Nyquist frequency (one-half of the sampling rate) in order to avoid aliasing. The variation in sampling rates

corresponds to the time at which measurements were taken and the capability of the instrument at the time of measurement. The majority of the transition paths measured were sampled at 1 MHz and took place after upgrades to the detection system were put in place.

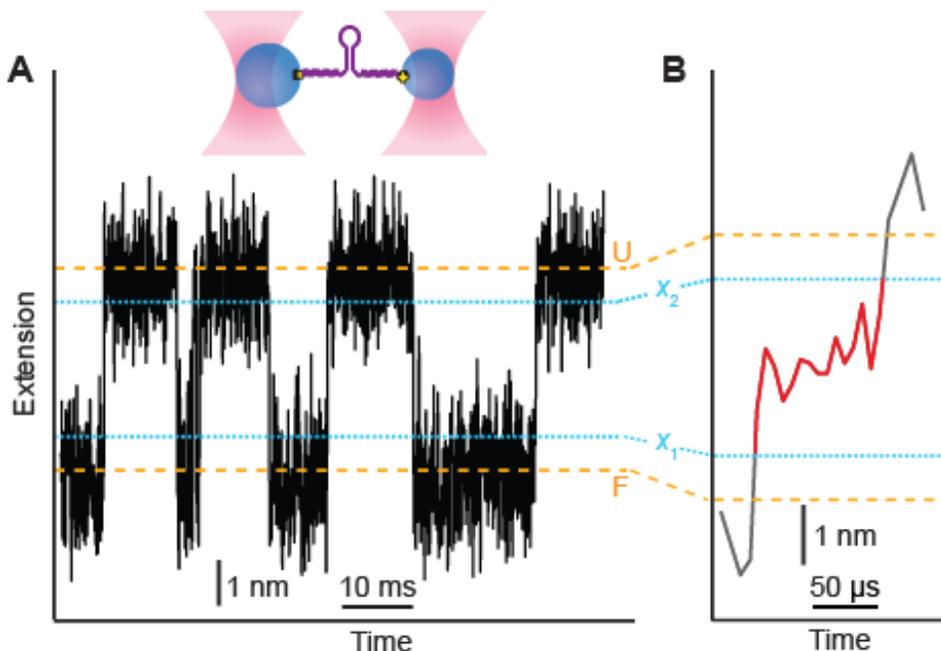


Fig 4.7 Transition-path measurements. (A) Transition paths were measured with dual beam optical tweezers (inset) in the constant trap separation paradigm, near $F_{1/2}$, the force at which the molecule is equally likely to occupy the folded (F) or unfolded (U) states. (B) High spatiotemporal resolution allows for observations of individual transition paths where the molecule transits the barrier region (x_f to x_u).

4.4 Summary

This chapter describes the types of optical tweezers measurements used in this thesis as well as what each type of measurement was specifically used for. Constant force measurements were not performed as part of this thesis, however,

the energy landscape reconstructions obtained from constant force measurements provided landscape parameters that were used in several aspects of the analysis presented in future chapters. FEC measurements were used to determine the quality of trapped molecules prior to measuring transition paths using constant trap separation measurements.

The requirements for accurately measuring TPs with OT are: high spatiotemporal resolution, low artifacts, and a proper reaction coordinate, are discussed and all shown to be satisfied in the case of the TP measurements presented in this thesis. Having discussed how TP measurements are performed, the remaining chapters focus on what can be learned from these measurements, looking at different TP properties and what these properties tell us about folding reactions.

Chapter 5

Transition-path studies

5.1 smFRET measurements of average transition-path times in folding

The first direct measurements of TPs were done in pioneering work by Eaton and colleagues using single-molecule Förster resonance energy transfer (smFRET) (Fig. 5.1), to determine the average transition-path time, τ_{TP} . In smFRET studies of folding, the transitions between the folded and unfolded states are generally induced through changes in the concentration of chemical denaturants. The progress of the reaction is then monitored via two (or sometimes 3) differently colored fluorophores, attached at different positions on the molecule (Fig. 5.1, 5.2A). When the donor fluorophore is excited by a photon, it will either emit a photon or transfer energy to the acceptor fluorophore through nonradioactive dipole-dipole coupling. When energy is transferred to the acceptor, the acceptor will emit a photon at a lower wavelength than those emitted by the donor. The efficiency of this energy transfer is strongly dependent on the distance between the donor and acceptor, allowing the distance between the two fluorophores to be determined by analysis of the emitted light (62) (Fig. 5.1). The collection of many photons is needed in order to accurately determine this efficiency, making high temporal resolution a challenge for FRET. Because of this, traditional FRET analysis has insufficient time resolution to detect individual TPs, but by applying a maximum-likelihood analysis of photon-by-photon trajectories(63), τ_{TP} could be deduced as the lifetime of a virtual intermediate state (Fig. 5.2B). It was measured

for two small proteins, a fast-folding β -structured WW domain (64) and the engineered helical protein α_3 D (65), and an upper bound was placed on τ_{TP} for the $\alpha\beta$ -structured protein GB1 (64) and a DNA hairpin (66). For the WW domain, τ_{TP} was estimated as $\sim 2 \mu\text{s}$ in a standard aqueous solvent, very close to the result obtained in atomistic simulations (67) and not far from the putative protein folding ‘speed limit’ of $\sim 0.1\text{--}1 \mu\text{s}$ for small proteins (68), whereas, for α_3 D, $\tau_{\text{TP}} \sim 13 \mu\text{s}$ (Fig. 5.2C); the upper bound for GB1 was $10 \mu\text{s}$, and that for the hairpin was $2.5 \mu\text{s}$.

Crucially, these results clustered closely in the range $\sim 1\text{--}10 \mu\text{s}$. In contrast, the folding rates of these molecules varied by as much as four orders of magnitude. The fact that τ_{TP} varies much less than the folding time, τ_{f} , reflects a key property of TPs: since by definition a molecule on a TP always receives sufficient thermal energy to cross the barrier, τ_{TP} is relatively insensitive to barrier height, as opposed to τ_{f} , which is dominated by the time a molecule must wait for an energy fluctuation large enough to allow barrier crossing. These concepts are captured theoretically by Kramers’ expression for τ_{f} (69) and Szabo’s expression for τ_{TP} (27) in the harmonic-barrier approximation:

$$\tau_{\text{f}} = \frac{2\pi k_{\text{B}}T}{D\sqrt{\kappa_{\text{w}}\kappa_{\text{b}}}} \exp(\Delta G^{\ddagger}/k_{\text{B}}T) \text{ and } \tau_{\text{TP}} = \frac{k_{\text{B}}T \ln(2e^{\gamma} \Delta G^{\ddagger}/k_{\text{B}}T)}{D\kappa_{\text{b}}}, \quad (5.1)$$

where $k_{\text{B}}T$ is the thermal energy, ΔG^{\ddagger} the barrier height, $\kappa_{\text{b/w}}$ the stiffness of the energy barrier/well, and D the diffusion coefficient. These expressions show that τ_{f} is exponentially sensitive to ΔG^{\ddagger} , while τ_{TP} varies little with ΔG^{\ddagger} . Moreover, the weaker dependence of τ_{TP} on ΔG^{\ddagger} , makes it much more sensitive to D .

This sensitivity of τ_{TP} to D was exploited by Chung & Eaton to show that the large τ_{TP} of $\alpha_3\text{D}$ arises from substantial internal friction, reflecting roughness in the energy landscape (65). By comparing τ_{TP} measurements directly to all-atom molecular dynamics simulations of the TPs, the mechanism underlying the internal friction was deduced and found to involve the formation of non-native salt bridges (70) (Fig. 5.2D). This mechanism was confirmed by showing that at low pH, where the salt bridges were disrupted by protonation, τ_{f} and τ_{TP} both decreased in roughly the same proportion, reflecting an increase in D from the removal of the salt bridges (Fig. 5.2E). Intriguingly, $\alpha_3\text{D}$ is thus a counterexample to previous findings, that non-native interactions do not play key roles in folding mechanisms for small proteins (71). This result suggests that the folding of engineered proteins may differ from that of naturally evolved molecules by having rougher landscapes (72) leading to slower diffusion. Combining measurements and atomistic simulations of TPs as done here, where the timescales of both experiment and simulation overlap directly, provides an exciting and uniquely powerful approach for gaining mechanistic insight into the microscopic dynamics of folding.

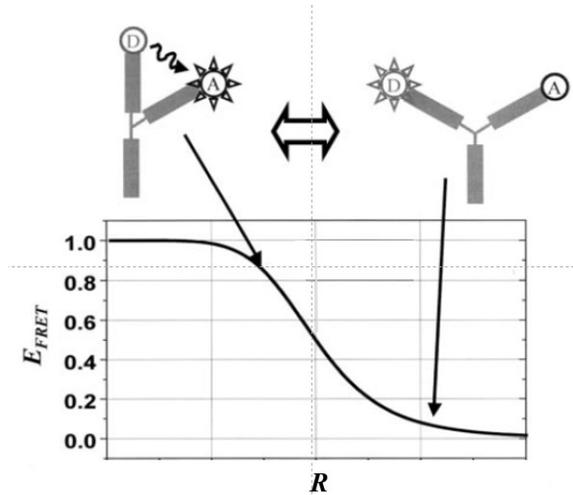


Figure 5.1: smFRET measurements. Top: Conformational changes in single molecules are detected with smFRET. When the donor and acceptor are in close proximity (left) energy is transferred to the acceptor, resulting in excitation and photon emission from the acceptor and weak emission from the donor. When the donor and acceptor are far from each other, little energy is transferred, resulting in primarily donor emission. Bottom: The energy transfer efficiency is a function of the distance R , between the two dyes. Figure from Ha *METHODS* 2001 with permission.

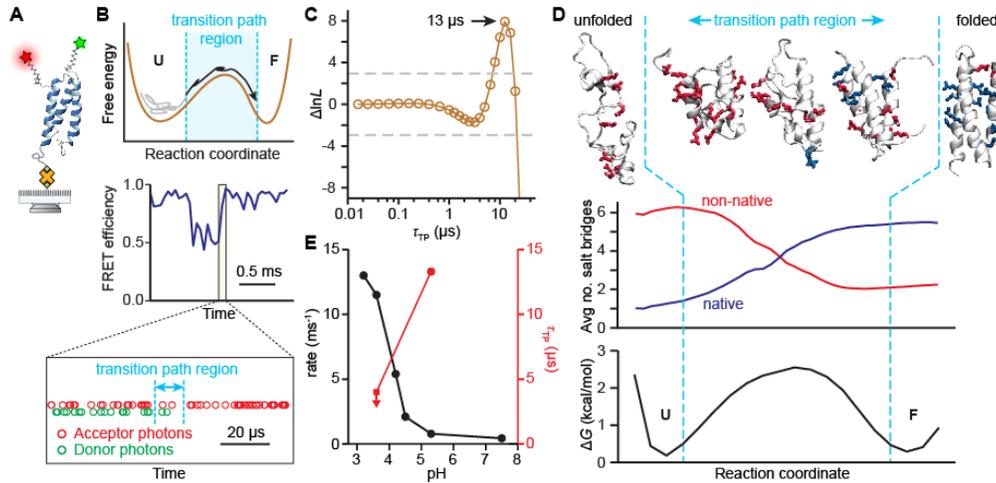


Figure 5.2: smFRET measurements of transition-path times. (A) Schematic of measurement showing FRET dye pair attached to the protein α_3D . (B) To measure the average time to cross the barrier region (top, cyan), FRET trajectories (middle, blue) typically have insufficient resolution because the minimum averaging window is too large, but analyzing photon arrival statistics (bottom) can identify the most likely τ_{TP} . (C) Maximum-likelihood analysis of photon statistics for α_3D yields $\tau_{TP} \sim 13 \mu s$. (D) All-atom simulations of α_3D folding show non-native salt bridges play a key role in generating internal friction slowing τ_{TP} ; the total number of salt bridges remains close to constant as native bridges (blue) replace non-native ones (red) along the transition paths. (E) The transition path time decreases when salt bridges are disfavored by lowering pH, in roughly the same proportion as the folding rate increases, indicating the changes arise from reducing the internal friction.

5.2 Force spectroscopy measurements of transition-path times

Two alternative approaches for determining transition-path times were developed by Woodside and colleagues for use with SMFS measurements. The first method relies only on the parameters of the free-energy landscape and hence involves no actual measurements of transition paths, while the second uses transition path measurements to calculate transition path times directly.

The first method uses the Szabo expression in order to determine the average transition path time and like the smFRET measurements, it is indirect and incapable of determining individual transition path times and is therefore only capable of measuring average transition-path times. This method requires the parameters of the free energy landscape, the kinetic rates, and the diffusion coefficient. These values can be found either by using constant force trajectories to reconstruct the full landscape via inverse Boltzmann transforms, then using Kramers' theory to find the kinetic rates and diffusion coefficient, or by measuring FECs and using the theory of Dudko et al.(73) to obtain the landscape parameters and rates, then using Kramers' theory to find the diffusion coefficient. Woodside and colleagues applied this landscape-based method to determine the average transition path time for DNA hairpins of varied size and sequence (42), the prion protein PRP (16), and various RNA pseudoknots (42). The measured average transition path times varied between the different molecules, ranging from a few microseconds to tens of microseconds.

The second method exploits SMFS ability to measure and visualize individual transition-path trajectories. Measuring transition paths as described in chapter 4, individual transition path times were taken directly from the trajectories by measuring the time required for each individual TP to completely transit from the unfolded to folded state or vice versa. This approach was used to observe TPs in both DNA hairpin folding (43, 44) and prion protein misfolding (44, 74). These measurements revealed a very broad distribution of TP times, $p(t_{TP})$, reflecting statistically independent and highly variable but time-reversal symmetric behavior in each transition as the barrier was crossed diffusively (Fig. 5.3). Values for τ_{TP} ,

calculated from $p(t_{TP})$, matched those predicted from the energy landscape reconstruction method (12, 75). Interestingly, τ_{TP} was orders of magnitude larger for prion protein misfolding—500 μs —than for native folding—2 μs , as estimated from Eq. 5.1 (16)—implying that misfolding landscapes are much rougher, possibly because evolution selects for efficient native folding but not misfolding.

Having $p(t_{TP})$ allowed for the comparison of the observed distribution to that expected for transit over harmonic barriers in the high-barrier limit in the Kramers regime, where $p(t_{TP})$ is expected to take the form(76):

$$p(t_{TP}) \approx \frac{\omega_K \sqrt{\beta \Delta G^\ddagger}}{1 - \text{erf} \sqrt{\beta \Delta G^\ddagger}} \frac{\exp[-\beta \Delta G^\ddagger \coth(\omega_K t_{TP} / 2)]}{\sinh(\omega_K t_{TP} / 2) \sqrt{2\pi \sinh(\omega_K t_{TP})}} \quad (5.2)$$

where $\omega_K = \beta D \kappa_b$ is the decay constant of the exponential tail. This long-time exponential tail can itself be approximated by

$$p(t_{TP}) \approx 2\omega_K \beta \Delta G^\ddagger \exp(-\omega_K t_{TP}) \quad (5.3)$$

The observed distribution was found to follow the same functional form of the equations above and fitting the observed distribution (Fig.5.3) to the above two equations yielded a value of $\omega_K = 6 \pm 3 \times 10^4 \text{ s}^{-1}$, corresponding to a diffusion coefficient $D = 2 \pm 1 \times 10^5 \text{ nm}^2/\text{s}$, close to the values calculated from Kramers' theory and the Szabo expression. The value of the barrier height taken from the fits was perplexing as it was found to be $\Delta G^\ddagger \sim 0.4 k_B T$, much lower than the value measured from landscape reconstructions of the same hairpin.

In the same study, Woodside and colleagues also measured $p(t_{TP})$ for the prion protein PRP (Fig. 5.3C). After fitting the distribution they again found that while the diffusion coefficient matched what was found via Kramers' theory, the

barrier height implied by the fits was much too low. This discrepancy has been attributed to various effects, including the role of entropy in the barrier (77), the dependence of TPs on a different type of barrier than rates (78), anomalous diffusion at short time scales (79), and memory effects in the dynamics (80) and/or non-thermal noise (81). Regardless of the true underlying cause for the discrepancy, this study demonstrated the ability of SMFS transition path measurements to probe folding reactions with unprecedented detail.

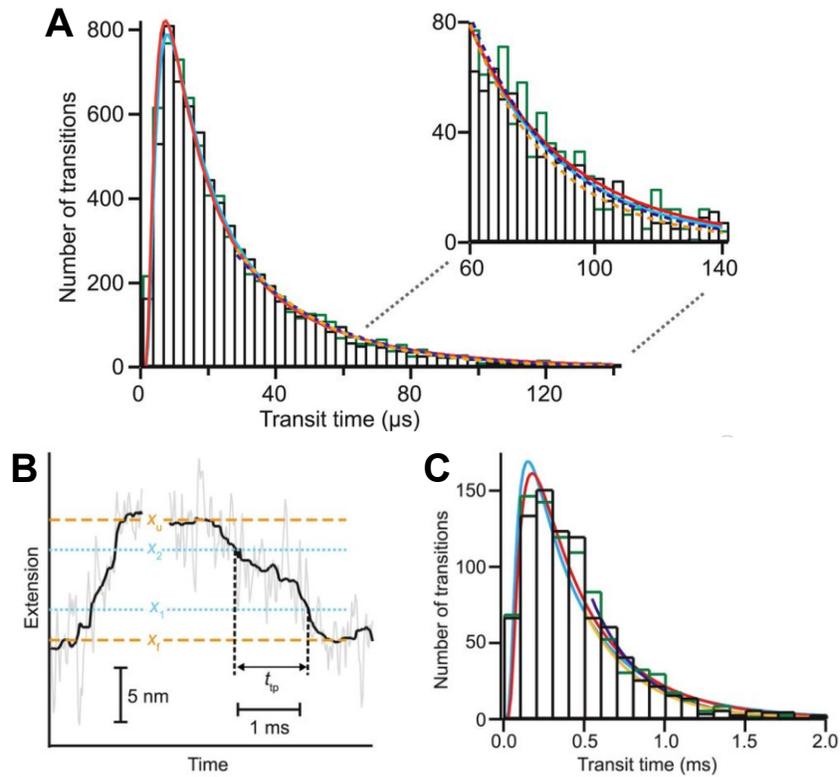


Figure 5.3: SMFS measurements of transition-path times. (A) Distribution of transition-path times measured for the DNA hairpin 30R50/T4. The distribution is the same for folding (green) and unfolding (black). The full distribution is well fit by Eq. 5.2 and the exponential tail is well fit by Eq. 5.3. (B) An unfolding transition (left) and a refolding transition (right) measured for the prion protein PRP. The transition path time is measured as the time required to cross from x_1 to x_2 (cyan). (C) Distribution of transition-path times measured for the prion protein PRP. The full distribution and the exponential tail are well fit by equations $_$ and $_$ respectively. Adapted from Neupane *et al* Science 2016 with permission.

5.3 Transition paths in protein-protein complexes

Although most TP measurements have focused on unimolecular folding reactions, recent work has extended the methods described above to study coupled binding and folding—where two molecules bind and then reconfigure to form the final folded structure—and characterize the resulting encounter complex.

Sturzenegger *et al.* (82) used smFRET (Fig. 5.4A) to examine the association of two intrinsically disordered proteins (IDPs) that form stable folded dimers upon binding, applying maximum-likelihood analysis to discern a transition intermediate wherein the proteins were bound but not yet folded. They found that this complex was much longer-lived than the transition paths observed in unimolecular folding, with $\tau_{\text{TP}} \sim 80 \mu\text{s}$, allowing the most likely duration of individual TPs to be estimated from the FRET signal (Fig 5.4B). Ruling out internal friction effects as the origin of the large τ_{TP} , $p(t_{\text{TP}})$ was found to be most consistent with local barriers around a high-energy intermediate (Fig. 5.4C), where encounter-complex formation was driven electrostatically but subsequent folding dominated by hydrophobic effects. In a similar study, Kim *et al.* (83) also used photon-by-photon analysis to compare the lifetime of encounter complexes for two sets of IDPs. The first set, TAD, and NCBD require binding in order to fold while the second set, barnase, and barstar, are folded independent of binding. The measured lifetime of the encounter complex for TAD/NCBD was found to be at least two orders of magnitude longer than that of barnase/barstar. The results suggest that the encounter complex is stabilized by non-native interactions which are far less prevalent in the already folded barnase/barstar complex. These works highlight the potential to extend TP measurements to illuminate the microscopic dynamics of the many multi-protein complexes that play essential roles in biological processes.

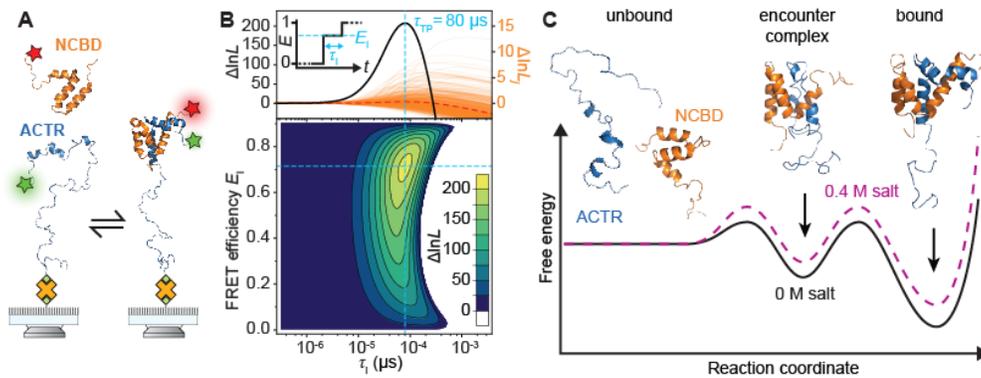


Figure 5.4: Measuring TPs in an encounter complex. (A) Measurement schematic showing FRET dye labels attached to two disordered proteins that bind and then fold. (B) Analysis of photon arrival statistics treating TP as a virtual intermediate (top, inset) reveals most likely TP times for each transition (top, orange), overall average $\tau_{TP} = 80 \mu\text{s}$ (top, black), and distribution of TP time and FRET value likelihoods (bottom). (C) The long TP times arises from a high-energy intermediate with salt-dependent barriers rather than high internal friction.

Chapter 6

Transition-path velocities

In this chapter, we move beyond measurements of the transition-path time and begin to probe the local properties of transition paths. Specifically, we measure the local velocity along transition paths in DNA hairpin folding, finding that the distribution of velocities agrees well with diffusive theories, yielding the diffusion coefficient. We use the average velocity to calculate the transmission factor in transition-state theory (TST), finding observed rates that are $\sim 10^5$ -fold slower than predicted by TST. This work quantifies the importance of barrier recrossing events and highlights the effectiveness of the diffusive model of folding

The material in this chapter was originally published as two separate manuscripts. Sections 6.1- 6.4 were published as K. Neupane, N.Q. Hoffer, M.T. Woodside “Measuring the velocity along transition paths during the folding of single biological molecules.” *Phys. Rev. Lett.*, 121:018102 (2018). This paper was chosen as an “editor’s suggestion” by the editors of *Phys. Rev. Lett.* For this paper KN and NQH took the data, NQH performed the analysis, and KN, NQH, and MTW wrote the paper. Sections 6.5- 6.7 were published as K. Neupane, N.Q. Hoffer, M.T. Woodside “Testing Kinetic Identities Involving Transition-Path Properties using Single-Molecule Folding Trajectories.” *J. Phys. Chem. B*, 122:11095-11099 (2018). For this paper KN and NQH took the data, NQH and KN performed the analysis, and KN, NQH, and MTW wrote the paper.

6.1 Transition-path velocity measurements

The transition path studies discussed so far have focused primarily on the

time required for the transition to take place: the average transition-path time was measured for both proteins (16, 64, 65, 70, 74) and nucleic acids (42, 84) using advanced single-molecule fluorescence and force spectroscopy methods, and the variations in time for individual transition-path crossings was measured in proteins and nucleic acids with force spectroscopy (43, 44). In many ways, however, the local velocity along the transition paths is more interesting than global properties like the transit time. The velocity reflects the dynamics within the crucial transition states, a feature of folding reactions that had not been possible to observe previously. This chapter focuses on the first measurements of transition-path velocities, studying the folding of DNA hairpins as a model system for ‘two-state’ folding. Properties of the local velocity along the transition paths are characterized and related to the physical picture of folding as a diffusive search over an energy landscape (5, 29).

Single DNA hairpins having different sequences and energy landscapes (specifically, hairpins 20R0/T4, 20R25/T4, 20R55/T4, 20R100/T4, 30R50/T4, and 20TS06/T4 from Appendix A) were attached to beads held in high-resolution optical traps via kilobase-long linkers of double-stranded DNA (Fig. 6.1(a), upper inset) as described previously. Hairpins were held under tension near $F^{1/2}$, the force at which the folded and unfolded states were equally occupied, at constant trap separation. The end-to-end extension of the molecule was measured as the hairpin fluctuated in equilibrium between folded and unfolded (Fig. 6.1(a)), yielding a total of ~8,500–46,000 transitions for each hairpin. High trap stiffness (0.75–1.1 pN/nm in one trap, 0.56–0.63 pN/nm in the other) was maintained to maximize the time

resolution of the measurement; under these conditions, the time resolution of the measurement was 6–9 μs , and the kinetic artifacts from beads and handles were small (57, 60). Transition paths were identified from the extension trajectories (Fig. 6.1(b)) as those parts of the trajectory passing between two boundaries (x_1 and x_2) demarking the barrier region. In order to capture as much of the dynamics between the folded (F) and unfolded (U) states as possible, the barrier region was defined as the middle 2/3 of the distance traversed between F and U. We found the velocity of the transition paths, $v(t)$, from the local slope of the trajectory, $x(t)$. To reduce the effects of random measurement noise, the trajectories (Fig 6.1(b), black) were first smoothed with a smoothing spline interpolation (Fig. 6.1(b), red) before numerical differentiation (Fig. 6.1(b), upper inset). The velocity at each position along the reaction coordinate, $v(x)$, was then found directly by combining $v(t)$ and $x(t)$, taking the average value at any extension where re-crossing events occurred.

The velocity showed distinct local variations as the molecule crossed the barrier region, reflecting complex dynamics within the transition states. In some transitions, the motion was fast across the whole barrier (Fig 6.1(c), left), but in others, the fast motions were interrupted by periods of slower motion (Fig 6.1(b)) and even transient reversals in the direction of motion (Fig 6.1(c), center and right). Episodes of fast, slow, and reversed motion were distributed roughly randomly along the transition path from one transition to the next. Indeed, the distribution of velocity as a function of position within the barrier region measured from all transitions for each hairpin (Figs. 6.2(a) and 6.3(a)) showed velocities ranging from roughly -0.5 to 1.5 mm/s at all positions. The overall distribution of velocities

(Figs. 6.2(b) and 6.3(b)) was close to Gaussian in all cases but slightly skewed; for each hairpin, negative velocities indicating reversed motion were observed at every position within the barrier region, providing direct evidence of barrier re-crossing events as expected for a diffusive process. Finding the average velocity profile along the transition paths from the mean of the distribution at each position in the barrier region, the result was the same for both folding (Figs. 6.2(c), black) and unfolding (Figs. 6.2(c), red) transitions, as expected from the time-reversal symmetry of the process. The velocity profiles varied by $\sim 10\text{--}40\%$ across the barrier region and were noticeably different for different hairpin sequences.

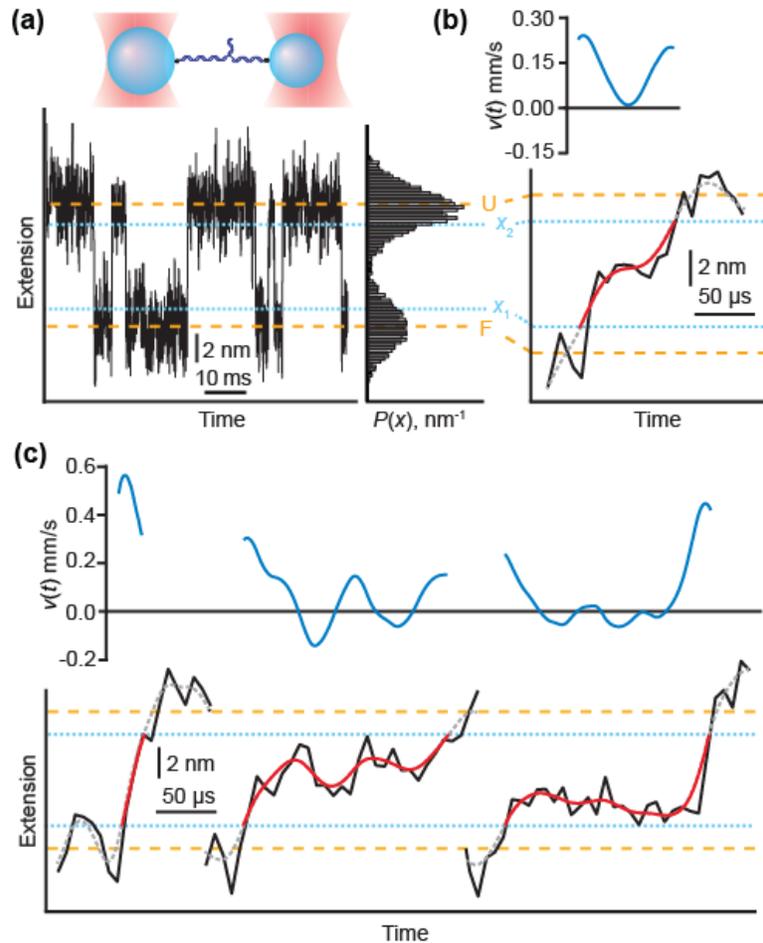


Fig 6.1: Measuring transition path velocity with optical tweezers. (a) The end-to-end extension of the hairpin held under tension fluctuates between the folded and unfolded states (dashed lines). Upper inset: Cartoon of measurement. A hairpin attached to DNA handles is held between two beads trapped by laser beams. Right inset: Probability distribution of hairpin extension. (b) Zooming in on a single unfolding transition, the transition path is seen as that part of the trajectory crossing between the boundaries defining the barrier region (dotted lines). Inset: The velocity along the transition path is found by differentiating the smoothed transition path. (c) Selected unfolding transitions and corresponding velocity profiles showing a wide range of velocities. The trajectory sometimes reverses course along the transition path, reflecting barrier re-crossing.

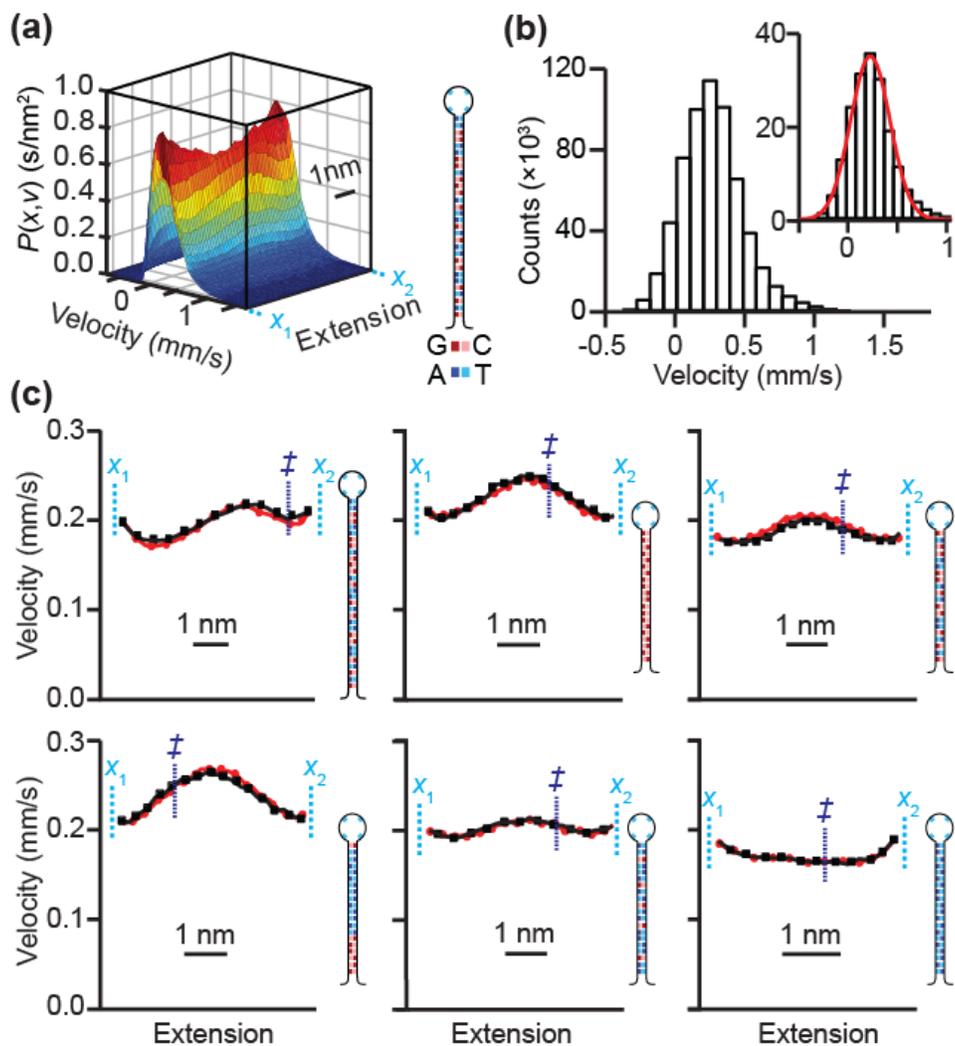


Fig 6.2: (color online): Transition-path velocity distributions. (a) The probability density for the velocity as a function of position within the barrier region for the hairpin 30R50T4. Inset: sequence of the hairpin 30R50T. (b) The distribution of velocities measured at all positions in the barrier region for the hairpin 30R50T4 is close to Gaussian. Inset: Fitting the distribution of velocities observed at the barrier peak to a Gaussian yields D . (c) The average velocity profile within the barrier region is the same for folding (black) as for unfolding (red) but generally non-uniform. Cyan: barrier region boundaries (x_1, x_2). Blue: position of barrier peak (\ddagger). Insets: hairpin sequences. Error bars represent standard error of the mean.

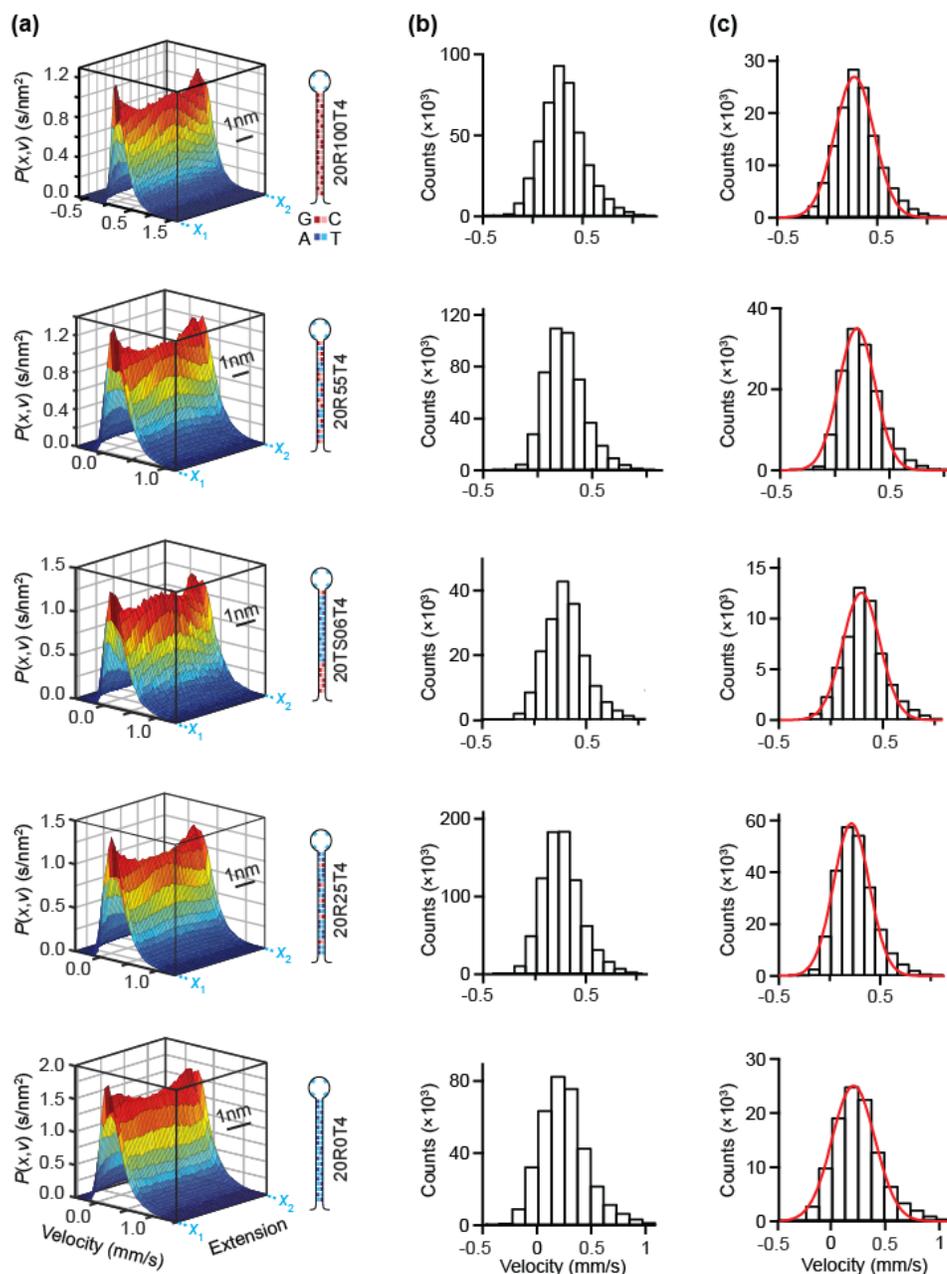


Figure 6.3: Transition-path velocity distributions. (a) Probability density for the velocity as a function of position within the barrier region for different hairpins. Inset: hairpin sequences. (b) The distribution of velocities observed at all positions in the barrier region is close to Gaussian for each hairpin. (c) Fitting the distribution of velocities observed at the barrier peak to a Gaussian yields D .

6.2 Diffusion coefficients from transition path velocities

The diffusion coefficient D , is the crucial parameter that relates the kinetics of the folding to the thermodynamics of the energy landscape, reflecting the microscopic motions made by the molecule in its conformational search (29) and the ‘internal friction’ that gives rise to speed limits in folding (68). It is difficult to measure D using traditional approaches based on analysis of rates via Kramers’ theory because of its exponential sensitivity to errors in barrier energies (16, 42, 85), but transition-path measurements provide a more sensitive and robust way to probe it (43, 56, 65). Theoretical work has shown that a 1D harmonic barrier will produce a Gaussian distribution of velocities at the barrier peak with average velocity

$$\langle v(x^\ddagger) \rangle \approx 1.5D(\beta\kappa_b)^{1/2}, \quad (6.1)$$

where β is the inverse thermal energy and κ_b the barrier stiffness (23). The barrier-peak location for each hairpin (Figs. 6.2(c), \ddagger symbol) was taken as the average of the results found previously from the force-dependence of the rates (58, 75) and from energy-landscape reconstructions using committor analysis of the extension trajectories (43, 86).

The velocity distribution at the barrier peak was indeed reasonably well fit by a Gaussian for each hairpin (Figs. 6.2(b), inset, and 6.3(c)). Calculating D from $\langle v(x^\ddagger) \rangle$ via Eq. 6.1, we found good agreement with values for D obtained previously in independent ways: from the average transition path time (τ_{TP}), from the exponential decay of the individual transit-time distribution (43, 44), and from the folding rates via Kramers’ theory (42) (Table 6.1). The trend of increasing D with

increasing G:C content found in previous studies (43) was also recapitulated in the results from $\langle v(x^\ddagger) \rangle$, enhancing confidence that the velocity measurements are reliable.

DNA Hairpin	$\langle v(\ddagger) \rangle$ ($\times 10^2$ $\mu\text{m/s}$)	D from $\langle v(\ddagger) \rangle$ (\times $10^5 \text{ nm}^2/\text{s}$)	D from τ_{TP} ($\times 10^5$ nm^2/s)	D from $P_{\text{TP}}(t)$ (\times $10^5 \text{ nm}^2/\text{s}$)	D from rates (\times $10^5 \text{ nm}^2/\text{s}$)
30R50/T4	2.0 ± 0.3	2.5 ± 0.4	3.5 ± 0.3	1.8 ± 0.2	4.6 ± 0.5
20R100/T4	2.4 ± 0.2	3.1 ± 0.4	4.1 ± 0.3	2.2 ± 0.2	-
20R55/T4	2.0 ± 0.2	2.6 ± 0.5	3.6 ± 0.3	1.5 ± 0.2	-
20TS06/T4	2.5 ± 0.2	2.6 ± 0.7	3.1 ± 0.3	1.6 ± 0.2	5 ± 3
20R25/T4	2.0 ± 0.2	2.6 ± 0.5	2.6 ± 0.3	1.3 ± 0.2	-
20R0/T4	1.7 ± 0.3	2.2 ± 0.3	2.5 ± 0.2	1.0 ± 0.2	-

Table 6.1: Comparison of D from $\langle v(x^\ddagger) \rangle$, transition times, τ_{TP} , and rates.

Results based on average transition time (τ_{TP}) and distribution of transition times ($P_{\text{TP}}(t)$) are from Ref. 54 and Ref. 71. Results based on rates are from Ref. 70. Errors represent standard error on the mean.

6.3 Deviations from 1D harmonic theories

The quantitative consistency of the obtained values of D from four independent physical properties of the folding, each of which assumes a single dominant harmonic barrier and constant diffusivity, shows that these assumptions are reasonable, at least for DNA hairpins. They are not ideal, however, as suggested by the slight skew in the velocity distributions at x^\ddagger (skew ~ 0.6 – 1.6). Indeed, Brownian-dynamics simulations of transition path trajectories (Fig 6.4) show that anharmonicity, position-dependence of D , and the presence of multiple types of transition paths can all generate skew in the velocity distribution (Fig 6.4). The

observed skew likely reflects contributions from each of these effects, since the barriers are not strictly harmonic, D should vary at least somewhat with position (89, 90), and 1D descriptions of hairpin folding are incomplete (91) even if they work reasonably well (15).

The average velocity profile, $\langle v(x) \rangle$, shows even more substantial deviations from the theoretical expectations. For transition paths with a 1D harmonic barrier, $\langle v(x) \rangle$ is expected to have a minimum at the top of the barrier (92). In most cases, however, the velocity is not near a minimum at the barrier top (Fig. 6.2(c)). This disagreement could arise from anharmonicity in the barrier, reflecting the sensitivity of $\langle v(x) \rangle$ to the barrier shape as found previously in simulations (23), since energy-landscape reconstructions show that the barriers are not completely harmonic (43, 75). Another possibility is that D is not constant (as assumed in the theory) but rather depends on the position, as may arise from the projection of the full multi-dimensional landscape onto a 1D reaction coordinate (89); a modest position dependence is also expected from the sequence-dependence of diffusion in DNA duplexes (88). The likely influence on $\langle v(x) \rangle$ of the position-dependence of D can be deduced by comparing the experimental velocity profiles to those obtained from Brownian dynamics simulations of transitions over the energy landscapes measured for each of the hairpins made under the assumption of a constant D . Although $\langle v(x) \rangle$ from the simulations (Fig 6.5) recapitulates some of the qualitative features observed experimentally (Fig 6.2c), the details of the spatial variations differ, suggesting that at least some of the variations come from position-dependence of D .

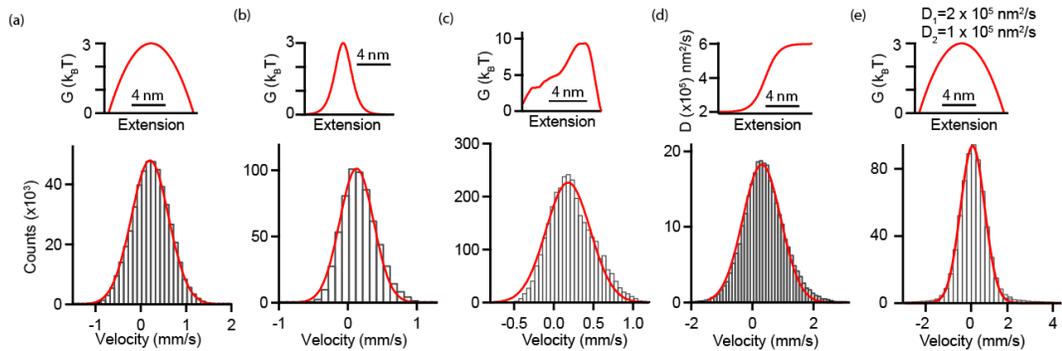


Fig 6.4: Simulated transition-path velocity distributions. The factors contributing to skew in the transition-path velocity distribution were explored using Brownian dynamics simulations. (a) Simulations of transitions over a harmonic barrier (inset) with constant D yield a Gaussian distribution with skew ~ 0 , as expected. (b) Simulations using an anharmonic Eckart barrier (inset) with constant D yield a skewed distribution (here, skew ~ 0.4). (c) Simulations using the barrier measured for hairpin 30R50/T4 (inset) with constant D yield a slight skew (~ 0.2). (d) Simulations using a harmonic barrier as in (a) but with a sigmoidal $D(x)$ (inset) yield skew (here, ~ 0.3). (e) Simulations including two types of transition paths having harmonic barriers of the same height (inset) but D differing by a factor of 5 yield skew (here, ~ 1); the faster transition path was occupied $1/6^{\text{th}}$ of the time.

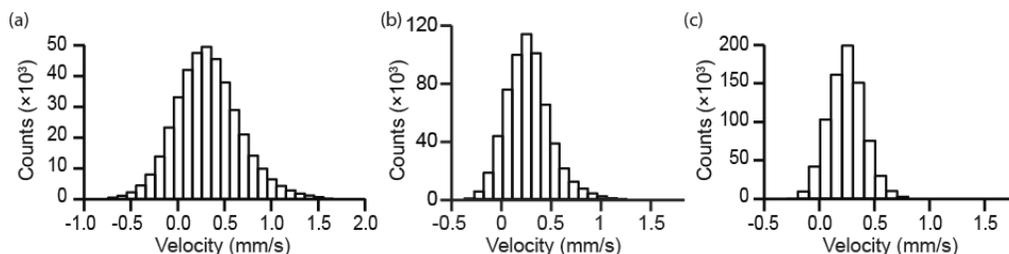


Fig 6.5: Effects of smoothing on transition-path velocity distributions. Velocity distributions resulting from data smoothed by different amounts (a) The data is smoothed at half of the optimal amount. (b) The data is optimally smoothed. (c) The data is smoothed at double the optimal amount. The distribution of velocities is Gaussian in all cases. The width of the distribution decreases as the smoothing factor increases.

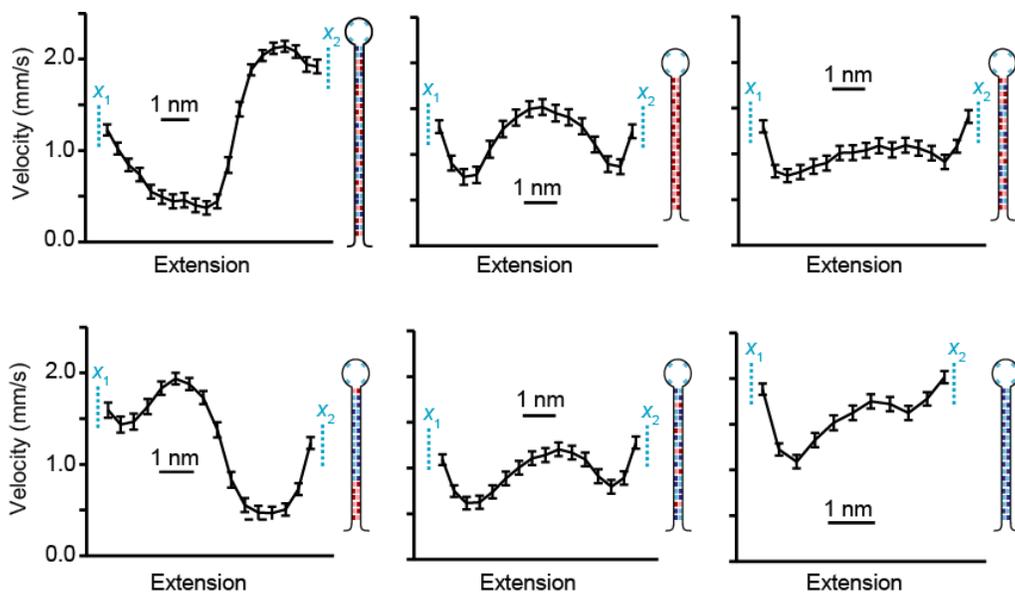


Fig 6.6: Simulated velocity profiles. The average transition-path velocity profiles expected from the experimentally measured energy landscapes under the assumption of constant D were determined from Brownian dynamics simulations for hairpins (a) 30R50/T4, (b) 20R100/T4, (c) 20R55T/4, (d) 20TS06/T4, (e) 20R25/T4, and (f) 20R0/T4. The simulated profiles share some qualitatively similar features with the observed velocity profiles (Fig 6.2) while differing in detail, suggesting that the observed profiles reflect position-dependence in D . Error bars represent standard error of the mean.

6.4 Comparing transition-state theory to Kramers' reaction rate theory

An intriguing aspect of transition-path velocities is that they illuminate the role of barrier re-crossing events in folding reactions, by showing how re-crossing alters the observed rates and velocities from what would be expected if instead classical transition-state theory (TST), which neglects re-crossing, held true. In TST, the microscopic velocity is the thermal velocity of the molecule, $v_{\text{th}} = (2k_{\text{B}}T/\pi m)^{1/2}$, where m is the mass. If barrier re-crossing does not occur, then the rate should be related to the velocity across the barrier by

$$k_{U/F}^{\text{TST}} = P(x^\ddagger)v_{\text{th}}/2P_{F/U}, \quad (6.2)$$

where $k_{U/F}^{\text{TST}}$ is the transition-state theory rate for unfolding/refolding, $P(x^\ddagger)$ the Boltzmann-weighted equilibrium occupancy at the barrier peak, and $P_{F/U}$ is the fractional occupancy of the folded/unfolded state(19, 93). If barrier re-crossing does not occur frequently, then the observed rates ($k_{U/F}$) should be similar to $k_{U/F}^{\text{TST}}$. However, if re-crossing events are significant then one would expect $k \ll k^{\text{TST}}$; the transmission factor $\kappa = k/k^{\text{TST}}$ quantifies the importance of re-crossing as the factor by which the rates are depressed compared to TST. The transmission factor can also be determined from the velocity, using relations between the observed velocity, the probability distribution along the transition paths, and the committor probability (94):

$$\kappa = \langle v(x^\ddagger) \rangle / 2v_{\text{th}}. \quad (6.3)$$

We evaluated κ from both the rates and the velocities. First, the rates $k_{F/U}$ were determined from single-exponential fits (Fig 6.6) to the distributions of unfolded/folded-state lifetimes, as measured directly from the extension trajectories by partitioning the trajectories into the two states via thresholding (58), $P_{F/U}$ was found from the fraction of time spent in each state, and $P(x^\ddagger)$ was evaluated from the free-energy landscape via $P(x^\ddagger) = A\exp(-\Delta G^\ddagger/k_B T)$, where A is the normalization constant ensuring $\int P(x)dx = 1$. The thermal velocity was calculated by estimating m as the mass of the part of the hairpin remaining unfolded when the barrier was crossed at x^\ddagger . The TST rates implied by Eq. 6.2 were found to be much higher than the rate observed directly in the extension trajectories for every hairpin, by a factor of roughly 100,000-fold, yielding $\kappa \sim 10^{-5}$ (Table 6.2). The values for κ

obtained independently from the velocities via Eq. 6.3 were in excellent agreement with those found from the rates (Table 6.2).

These results indicate that barrier re-crossing plays a central role in folding transitions, reducing rates by ~ 5 orders of magnitude. Classical transition-state theory thus does not provide a good description of folding, and using it to estimate kinetic pre-factors or activation energies as sometimes done (95) will lead to substantial overvaluation. Instead, diffusive theories like Kramers' approach are more appropriate. In fact, κ can be evaluated directly from Kramers' theory as $\kappa = (m\kappa_b)^{1/2}D/k_B T$ (85). Using the average values of D from Table 6.1, we obtained very similar values for κ as found from the rates and velocities (Table 6.2). Although it has previously been shown that Kramers' theory accounts well for the observed kinetics in both simulations (21, 89, 96) and experiments (16, 97, 98), the transmission factor reflecting the influence of barrier re-crossing had not previously been measured.

DNA Hairpin	κ from rates ($\times 10^{-6}$)	κ from velocities ($\times 10^{-6}$)	κ from Kramers' theory ($\times 10^{-6}$)
30R50/T4	10 ± 4	10 ± 2	11 ± 2
20R100/T4	10 ± 1	9 ± 2	9 ± 1
20R55/T4	7 ± 1	7 ± 2	7 ± 2
20TS06/T4	4.4 ± 0.4	7 ± 3	7 ± 2
20R25/T4	5.4 ± 0.5	7 ± 2	7 ± 2
20R0/T4	0.9 ± 0.2	6 ± 2	5 ± 1

Table 6.2: Comparison of κ from rates, velocities, and Kramers' theory.
Errors represent standard deviation.

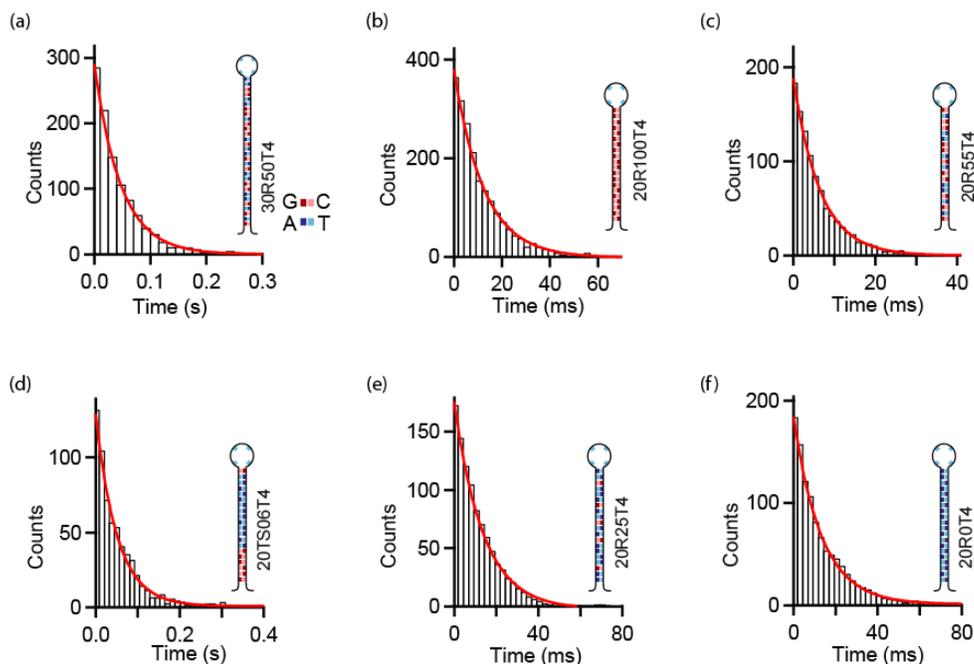


Fig 6.7: Rate measurements. Folding rates were determined from single-exponential fits (red) to the distribution of unfolded-state lifetimes (black) measured directly from extension trajectories. Representative examples are shown for hairpins (a) 30R50/T4, (b) 20R100/T4, (c) 20R55T/4, (d) 20TS06/T4, (e) 20R25/T4, and (f) 20R0/T4. The same analysis applied to folded-state lifetimes yielded the unfolding rates.

6.5 Transition path velocity and position probability distribution

The difficulty in measuring transition paths has led to a situation where many theories of transition paths have remained untested experimentally. One such relationship that has been proposed is that the average velocity profile along the transition paths should be inversely proportional to the position probability distribution within them:

$$P(x|TP) = [\langle v(x) \rangle \tau_{tp}]^{-1}, \quad (6.4)$$

where $P(x|TP)$ is the probability density for being at a value x of the reaction

coordinate when on a transition path, $\langle v(x) \rangle$ is the average velocity at x along the transition path and τ_{tp} is the average transition-path time (23). This relation arises from a simple physical intuition: considering a single transition, the time needed to move from x to $x + dx$ is just $dx/v(x)$, hence the fraction of the transition-path time spent in this reaction-coordinate interval is just $dt/t_{\text{tp}} = dx/[v(x)t_{\text{tp}}]$, where t_{tp} is the time to complete that particular transition path. Since this relation is true for each transition, the analogous expression holds for the average over all transition paths, and hence the probability for being between x and $x + dx$ is $P(x|\text{TP})dx = dx/[\langle v(x) \rangle \tau_{\text{tp}}]$.

Already having found $\langle v(x) \rangle$ for several DNA hairpins, testing this relationship only required us to calculate the position probability distributions. The distributions were obtained by isolating the transition paths from the rest of the trajectory and calculating the extension probability distribution at each point of the reaction coordinate within the barrier region. We then calculated $P(x|\text{TP})$ (Fig 6.7, grey), finding that the different hairpins had somewhat different transition-path occupancies: for hairpin 20R0/T4 it was effectively flat across the barrier region, whereas for the other hairpins it was lowest near the top of the barrier (indicated by the ‡ symbol). Despite these differences, when comparing $P(x|\text{TP})$ to $[\langle v(x) \rangle \tau_{\text{tp}}]^{-1}$ (Fig 6.7, black), we found good agreement for all hairpins, validating Eq. 6.4.

Although this relationship is seemingly quite simple, it should be noted that the agreement between the velocity calculated from the smoothed trajectories and the probability distribution measured directly from the original data indicates that the smoothing procedure is not introducing artifacts into the transition-path shape:

the average velocity profile is the same as what would be expected based on the local occupancy statistics. More interestingly, however, this work shows that measurements of the transition-path occupancy are sufficient to determine the average velocity profile across the transition paths, since these two quantities are inversely proportional. $P(x|TP)$ thus offers an alternative route to measuring $\langle v(x) \rangle$ directly, one that is less stringent technically: it could in principle be used if the trajectories are too noisy to obtain $v(x)$ via differentiation, or even if it is not possible to sample the transition paths sufficiently finely to define their shapes adequately (*e.g.*, owing to sampling-rate limitations). Of course, $P(x|TP)$ is more limited than direct measurement of velocities, as it yields only the average velocity profile. If the folding is dominated by a single type of transition path, this approach may be sufficient to characterize the transition behavior, but if multiple, distinct types of transition paths are present, then it may provide a misleading picture because it reflects the average behavior.

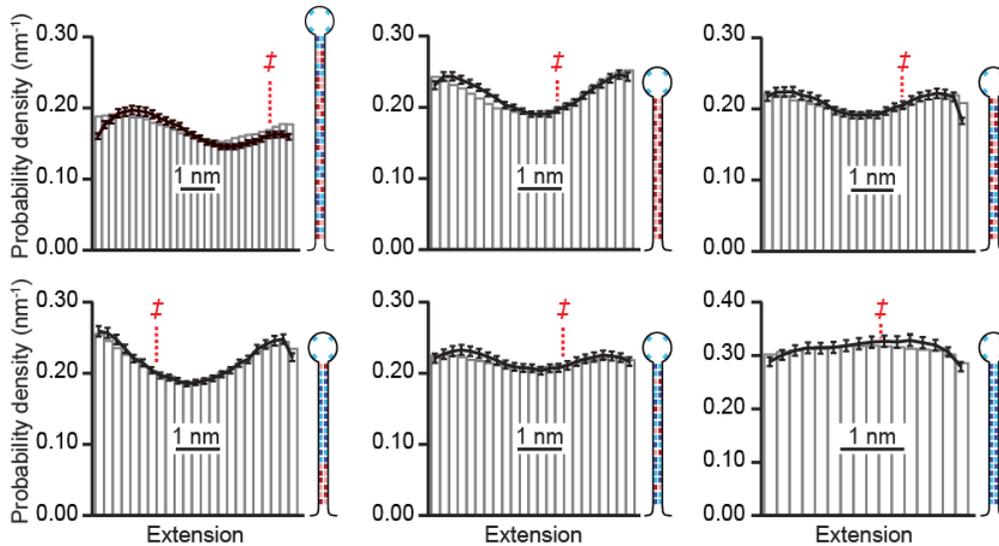


Fig 6.8: Test of relation between transition-path velocity and occupancy. The average transition path occupancy (grey) agreed well with the expectation from the transition-path velocity (black) via Eq 7.4 for all hairpins. The location of the barrier top is indicated for each hairpin by the double dagger. Insets: hairpin sequences. Error bars represent s.e.m.

6.6 Calculating the committor from transition-path velocities

An intriguing consequence of Eq. 6.4 is that $\langle v(x) \rangle$ can be used to evaluate the committor probability, $p_{\text{fold}}(x)$, given by

$$p_{\text{fold}}(x) = \frac{\int_{x_a}^x D(x')^{-1} e^{\beta G(x')} dx'}{\int_{x_1}^{x_2} D(x')^{-1} e^{\beta G(x')} dx'}$$

(6.5)

through the relationship of the latter to $P(x|\text{TP})$. By Bayes' theorem, $P(x|\text{TP}) = P(x)p(\text{TP}|x)/p(\text{TP})$, where $p(\text{TP}|x)$ is the conditional probability of being on a transition path at extension x and $P(x)$ is the equilibrium extension probability. Given that we also have $p(\text{TP}|x) = 2p_{\text{fold}}(x)[1 - p_{\text{fold}}(x)]$ for ideal diffusion (99), we

obtain:

$$p_{\text{fold}}(x)[1 - p_{\text{fold}}(x)] = N/[2T\langle v(x) \rangle P(x)], \quad (6.6)$$

where N is the number of transitions observed in the trajectory and T is the total trajectory duration. We verified this relation using transition-path measurements of hairpin 30R50/T4. Calculating p_{fold} from the average transition-path velocity via Eq. 6.6 (Fig 6.8, red), we compared the result to the p_{fold} calculated directly from the extension trajectories using the definition of p_{fold} as done in previous studies (86) (Fig 6.8, black). The agreement was very good within error. We note that the relation between the committor and $p(\text{TP}|x)$ underlying Eq. 6.6 assumes folding is a diffusive process but Eq.6.4 does not, hence this validation of Eq. 6.6 underlines the consistency of the diffusive model of folding with the identities being tested.

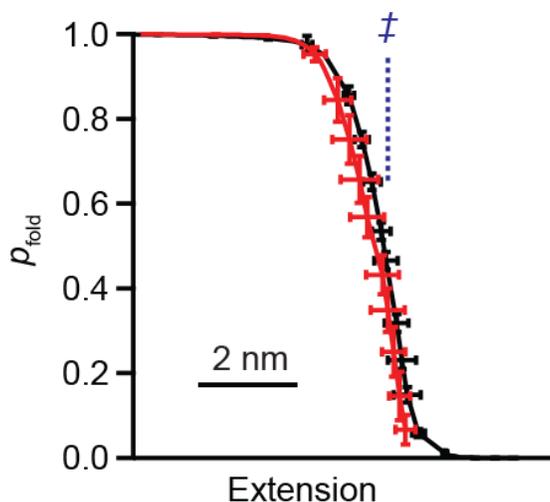


Fig 6.9: Comparison of empirical p_{fold} and p_{fold} calculated from transition-path velocity. The committor determined empirically from the full extension trajectories (black, error bars represent s.e.m.) matches the result for p_{fold} obtained from transition-path velocities via Eq. 6.6 (red, error bars represent standard error from bootstrapping analysis).

6.7 Discussion of transition-path velocity results

This study of the velocity in transition paths highlights how well the diffusive model of folding reactions works to describe and predict folding phenomena. Remarkable quantitative consistency is seen across a wide range of experimental observables at various scales. Analyzing different kinetic properties of folding, for example, leads to consistent values for the diffusion coefficient whether looking at rates, the much-smaller transition times(43, 44), or the speed of motion along the transition paths. The statistics of transition-path occupancy also quantitatively match what is expected for diffusive motions over the energy landscape (14, 15), whether reconstructed under equilibrium or non-equilibrium conditions (75, 86, 100). This consistency across numerous observables and many different types of measurement gives strong confidence in the quantitative validity of the diffusive picture of folding.

Chapter 7

Transition path properties in the limit of small barriers

A great deal of theory has been developed for transition paths in the limit of large harmonic barriers. While these approximations have been relatively successful at predicting transition path behavior, they are only applicable to molecules that have large harmonic barriers and are of little use to the many molecules with small or nonexistent barriers. This chapter addresses the need for establishing theory in the limit of small and or anharmonic barriers and establishes criteria for when to apply either the large or small barrier approximations.

The material presented in this chapter was originally published as A.G.T. Pyo, N.Q. Hoffer, K. Neupane, M.T. Woodside, Transition-path properties for folding reactions in the limit of small barriers. For this paper AGTP developed the theory, NH and KN took the data, AGTP performed the analysis, AGTP and MTW wrote the paper.

7.1 Large barrier approximations

Experimental studies of transition paths have made use of a few central results from theoretical work on transition paths obtained under the assumption of a large, harmonic barrier. Initial work(42, 64) measuring the average transition-path time, τ_{TP} , for example, made use of an expression for τ_{TP} valid for harmonic barriers in the large-barrier limit:(27, 76)

$$\tau_{\text{TP}} \approx \frac{\ln(2e^\gamma \beta \Delta G^\ddagger)}{\beta D \kappa_b}, \quad (7.1)$$

where γ is the Euler-Mascheroni constant, $\beta = 1/k_{\text{B}}T$ is the reciprocal of the thermal

energy, ΔG^\ddagger is the barrier height, D is the diffusion coefficient, and κ_b is the curvature of the barrier. An expression for the distribution of transition times, $P_{\text{TP}}(t)$, was also derived in the same limit(76, 101) and used to fit the distribution of times observed experimentally for individual transition paths in proteins and nucleic acids:(43, 44)

$$P_{\text{TP}}(t) \approx \frac{\omega_K \sqrt{\beta \Delta G^\ddagger}}{1 - \operatorname{erf} \sqrt{\beta \Delta G^\ddagger}} \frac{\exp\left[-\beta \Delta G^\ddagger \coth(\omega_K t/2)\right]}{\sinh(\omega_K t/2) \sqrt{2\pi \sinh(\omega_K t)}}, \quad (7.2)$$

where $\omega_K = \beta D \kappa_b$ is the inverse of the relaxation time. More recently, theoretical investigations into the shape of transition paths(23, 24, 26) led to an expression for the average velocity along a transition path in the limit of a large harmonic barrier:(23)

$$\langle v_{\text{TP}}(x) \rangle \approx \langle v(x^\ddagger) \rangle [1 + (x \omega_K / \langle v(x^\ddagger) \rangle)^2]^{1/2}, \text{ with } \langle v(x^\ddagger) \rangle = 2(e^{-\gamma} D \omega_K)^{-1/2}, \quad (7.3)$$

where x^\ddagger is the location of the top of the barrier. Eq. 7.3 was used to analyze recent measurements of the average velocity on transition paths in DNA hairpins. (102)

7.2 The need for small barrier approximations

These theoretical descriptions of transition-path properties in the limit of large harmonic barriers have generally described the experimental data quite well. Using them to deduce the value of D for various molecules, for example, returned results that were broadly self-consistent and similar to the values obtained from analysis of kinetic rates using Kramers' theory(42–44, 74, 102), which also assumes a harmonic barrier. (18) However, high-resolution measurements of the shape of the energy barrier(43, 75, 86) reveal that although approximating the barriers as harmonic is reasonable, anharmonicity is usually present to some degree.

Indeed, small but systematic differences between the estimate of D obtained from equations (6.1), (6.2), and (6.3) were attributed in part to barrier anharmonicity. (43, 102) Furthermore, many molecules have barriers that are small or even non-existent,(103) for which equations (6.1)–(6.3) would not be applicable. These considerations motivated work to find expressions for the transition-path times and velocities that are both simple enough to use for analyzing experimental data and valid under more general barrier conditions.

Berezhkovskii and Makarov(94) derived expressions for the average time and velocity as a function of position along the transition path without assuming large harmonic barriers. Here we extend this previous work, obtaining expressions that can be used in the limit of small harmonic barriers. We compare the effectiveness of the approximations (high- and low-barrier limits) and exact theory when applied to simulated folding trajectories over harmonic and anharmonic barriers with a range of heights, finding that the cross-over to the low-barrier regime occurs at barrier heights of $\sim 4 k_B T$. Finally, we apply the theory to experimental data that was previously analyzed with Eqs. (7.1) and (7.3), re-evaluating the diffusion coefficient extracted from the data to quantify the error in the estimate of D introduced by approximating the barriers as large and harmonic.

7.3 Derivation of the small barrier limit equations

We first re-derive the general expressions for $\langle v_{TP}(x) \rangle$ and $\langle t_{TP}(x) \rangle$ presented in Ref. (94) in terms of the committor function, using a slightly different approach that starts by assuming that folding is described by a 1D Smoluchowski equation:

$$\frac{\partial}{\partial t} \rho(x, t) = \frac{\partial}{\partial x} D(x) e^{-\beta G(x)} \frac{\partial}{\partial x} e^{\beta G(x)} \rho(x, t). \quad (7.4)$$

Here $x(t)$ is the trajectory of the reaction coordinate, $\rho(x, t)$ is the probability density function, and $G(x)$ is the free-energy landscape governing the folding. An advantage of this approach is that it describes the time evolution of the probability density, which is in principle an experimental observable. We re-express the reaction dynamics in terms of the committor function (also known as the splitting probability), $\phi_B(x)$, the probability that the molecule will reach state B as a function of the reaction coordinate, which is given by(104)

$$\phi_B(x) = \frac{\int_{x_A}^x D(x')^{-1} e^{\beta G(x')} dx'}{\int_{x_A}^{x_B} D(x')^{-1} e^{\beta G(x')} dx'}, \quad x \in (x_A, x_B). \quad (7.5)$$

Here x_A and x_B define the boundaries between the barrier region and respectively the folded and unfolded states. Following previous work,(105) the Smoluchowski equation conditioned for a transition event expressed in terms of ϕ_B becomes

$$\frac{\partial}{\partial t} \rho_{\text{TP}}(x, t) = \frac{\partial}{\partial x} D(x) \phi_B^2(x) e^{-\beta G(x)} \frac{\partial}{\partial x} \frac{e^{\beta G(x)}}{\phi_B^2(x)} \rho_{\text{TP}}(x, t), \quad (7.6)$$

where $\rho_{\text{TP}}(x, t) \equiv \rho(x, t | x_A \rightarrow x_B)$, with an absorbing boundary at $x = x_B$ and subject to the initial condition $\rho(x, t_0 | x_A \rightarrow x_B) = \lim_{\varepsilon \rightarrow 0^+} \delta(x - x_A - \varepsilon)$. Eq. 7.6 yields the time evolution of the probability density of the transition paths, which can be readily solved numerically. (106)

The average velocity as a function of position along the transition path, $\langle v_{\text{TP}}(x) \rangle$, is then given by

$$\langle v_{\text{TP}}(x) \rangle = \frac{\int_{t_0}^{\infty} J(x,t) dt}{\int_{t_0}^{\infty} \rho_{\text{TP}}(x,t) dt}, \quad (7.7)$$

where $J(x,t) \equiv J(x,t|x_A \rightarrow x_B)$ is the probability current. Since the probability distribution for transition paths is given by(19)

$$\int_{t_0}^{\infty} \rho_{\text{TP}}(x,t) dt = \frac{\phi_B(x)(1-\phi_B(x))e^{-\beta G(x)}}{\int_{x_A}^{x_B} \phi_B(x)(1-\phi_B(x))e^{-\beta G(x)} dx}, \quad (7.8)$$

we can combine Eqs. 7.6 and 7.8 with the continuity equation for particle conservation to solve explicitly for the time integral of J , which is just the reciprocal of τ_{TP} :

$$\int_{t_0}^{\infty} J(x,t) dt = \left[\int_{x_A}^{x_B} \phi_B(x)(1-\phi_B(x))e^{-\beta G(x)} dx \int_{x_A}^{x_B} D(x)^{-1} e^{\beta G(x)} dx \right]^{-1} = \tau_{\text{TP}}^{-1} \quad (7.9)$$

as found previously. (94) From here, we recapitulate the results from Ref. (94). In terms of the committor function, the velocity is given by

$$\langle v_{\text{TP}}(x) \rangle = D(x) \frac{\frac{d}{dx} \phi_B(x)}{\phi_B(x)(1-\phi_B(x))}, \quad (7.10)$$

which for constant diffusivity yields D in terms of the velocity at the barrier top:

$$D = \frac{1}{4} \langle v(x^\ddagger) \rangle e^{\beta G(x^\ddagger)} \int_{x_A}^{x_B} e^{\beta G(x)} dx. \quad (7.11)$$

Lastly, as shown previously,(94) by integrating the reciprocal of the transition velocity, the average time to reach the position x along the transition path can be expressed as

$$\langle t_{\text{TP}}(x | x_A \rightarrow x_B) \rangle = \int_{x_A}^x \phi_B(x') (1 - \phi_B(x')) e^{-\beta G(x')} dx' \int_{x_A}^{x_B} D(x)^{-1} e^{\beta G(x)} dx \quad (7.12)$$

In the case of constant D , the diffusion coefficient and the mean transit time are related by(106)

$$D = \int_{x_A}^{x_B} e^{-\beta G(x)} \left(\int_{x_A}^x e^{\beta G(x')} dx' \right) \left(\int_x^{x_B} e^{\beta G(x')} dx' \right) dx \left[\langle \tau_{\text{TP}} \rangle \int_{x_A}^{x_B} e^{\beta G(x')} dx' \right]^{-1} \quad (7.13)$$

Given that $\langle v(x^\ddagger) \rangle$ and τ_{TP} are experimental observables in transition-path measurements(44, 102) and the energy-barrier profile can sometimes be measured directly, as with the single-molecule force spectroscopy methods,(12) Eqs. 7.11 and 7.13 provide a more precise way to determine D , the fundamental parameter that connects the thermodynamics of the energy landscape to the dynamics of the folding.

Because energy barriers in folding reactions are often approximated as harmonic and D is typically assumed to be constant, it is useful to re-express the equations above for this special case. For a parabolic potential $G(x) = -\frac{1}{2}\kappa x^2$, where κ is the curvature of the barrier and the transition region spans the range $x = -L$ to L , the transition-path velocity for constant D is(94)

$$\langle v_{\text{TP}}(x | -L \rightarrow L) \rangle = \frac{4D\alpha \operatorname{erf}(\alpha L) e^{-\alpha^2 x^2}}{\sqrt{\pi} (\operatorname{erf}^2(\alpha L) - \operatorname{erf}^2(\alpha x))}, \quad (7.14)$$

where $\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-x'^2} dx'$ is the error function and $\alpha = \sqrt{\frac{\beta\kappa}{2}}$ (hence $(\alpha L)^2$ is

the barrier height). D is then given by

$$D = \frac{\sqrt{\pi}}{4\alpha} \operatorname{erfi}(\alpha L) \langle v(x^\ddagger) \rangle. \quad (7.15)$$

These results are valid for harmonic barriers regardless of the barrier height. Approximations have already been computed in the high-barrier limit,(23, 24, 26, 76) but not yet in the low-barrier limit. For small barriers, we use the approximation $\operatorname{erf}^2(x) \approx 1 - \exp(-\pi^2 x^2/8)$ for small x , leading to

$$\langle v_{\text{TP}}(x | -L \rightarrow L) \rangle = \frac{2\sqrt{2}D\alpha}{\sqrt{\pi}} \frac{\sqrt{\sinh\left(\frac{\pi^2}{16}\alpha^2 L^2\right)} e^{\frac{\pi^2}{32}\alpha^2 L^2}}{\sinh\left(\frac{\pi^2}{16}\alpha^2(L^2 - x^2)\right) e^{\left(1 - \frac{\pi^2}{16}\right)\alpha^2 x^2}} \quad (7.16)$$

and

$$\langle t_{\text{TP}}(x | -L \rightarrow L) \rangle = \frac{\pi \left[\lambda (\operatorname{erf}^2(\alpha L) - 1) (\operatorname{erfi}(\alpha x) + \operatorname{erfi}(\alpha L)) + \operatorname{erf}(\lambda \alpha x) + \operatorname{erf}(\lambda \alpha L) \right]}{8D\lambda\alpha^2 \operatorname{erf}(\alpha L)} \quad (7.17)$$

where $\operatorname{erfi}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{x'^2} dx'$ and $\lambda = \sqrt{\frac{\pi^2}{8} - 1}$.

Finally, by evaluating Eq. 7.17 at $x = L$, the average transition-path time can be found:

$$\tau_{\text{TP}} = \frac{\pi \left[\lambda (\operatorname{erf}^2(\alpha L) - 1) \operatorname{erfi}(\alpha L) + \operatorname{erf}(\lambda \alpha L) \right]}{4D\lambda\alpha^2 \operatorname{erf}(\alpha L)}. \quad (7.18)$$

We note that these results are still valid as the barrier height goes to zero. Taking the limit $\kappa \rightarrow 0$, Eq. 7.16 for the average transition-path velocity simplifies to $\langle v_{\text{TP}}(x) \rangle \approx 2DL/(L^2 - x^2)$, consistent with the exact solution; similarly, Eq. 7.18

for the average transition-path time simplifies to $\tau_{\text{TP}} \approx 2L^2/3D$, consistent with the expression found previously by Kim and Netz. (26)

7.4 Application to Brownian dynamics simulations

To test these equations and discern the practical effects of using them to analyze data from molecules with different barriers, we first applied them to analyze transition paths in Brownian dynamics simulations of motion with constant diffusion over parabolic barriers of varying barrier heights. We started by examining the average velocity profile, $\langle v_{\text{TP}}(x) \rangle$, for two barrier heights: $15 k_{\text{B}}T$ (high-barrier limit) and $3 k_{\text{B}}T$ (low-barrier limit). We found that the average velocity calculated directly from the simulated trajectories (Fig 7.1, black) matched the analytical predictions from Eq. 7.14 (Fig 7.1, red) very well for both barriers. In contrast, although the high-barrier approximation, Eq. 7.3, matched the empirical result for the high-barrier case (Fig 7.1A, cyan), as expected, it systematically underestimated the velocity in the low-barrier case (Fig 7.1B, cyan). Likewise, the low-barrier approximation, Eq. 7.16, matched the empirical result well for the low-barrier case (Fig 7.1B, blue), but systematically overestimated the velocity in the high-barrier case (Fig 7.1A, blue).

We next considered the average transition-path time, τ_{TP} . Transitions were simulated over harmonic barriers with heights ranging from 1 to $6 k_{\text{B}}T$. The values for τ_{TP} obtained empirically for each barrier height from the simulated transitions (Fig 7.2, black) were compared to the values predicted by Eq. 6.1 in the high-barrier limit (Fig 7.2, cyan), to the values predicted by Eq. 7.18 in the low-barrier limit (Fig 7.2, blue), and to the values predicted by the exact solution, Eq. 7.12 (Fig 7.2,

red). Whereas the exact solution matched the empirical results very well over the whole range of barrier heights, the high-barrier approximation overestimated τ_{TP} for barriers lower than $\sim 4 k_{\text{B}}T$, and the low-barrier approximation underestimated τ_{TP} for barriers higher than $\sim 4 k_{\text{B}}T$. The cross-over between high-barrier and low-barrier approximations was thus around $4 k_{\text{B}}T$. The under- and over-estimations resulting respectively from the low- and high-barrier approximations were not very large in the range of barrier heights studied, however, being less than 10–15%.

Many of the experimental studies of transition paths to date have used the properties of transition paths to investigate the diffusion coefficient D , (43, 44, 65, 70, 102) because transition-path properties are far less sensitive to the height of the energy barrier (which is difficult to measure precisely) than are other kinetic properties like rates. (17, 56, 65) We therefore, examined the reliability of the various estimates of D that can be obtained from transition paths using the formulae above, in the different limits. Looking first at the value of D obtained from $\langle v(x^\ddagger) \rangle$, we used simulations over harmonic barriers with heights varying from 1 to $6 k_{\text{B}}T$ to compare the actual value of D imposed in the simulations (Fig 7.3A, black) to the value calculated from the exact solution via Eq. 7.15 (Fig 7.3A, red). We found that Eq. 7.15 returned the expected value of D over the whole range of barrier heights. Repeating the analysis using the low-barrier approximation (Eq. 7.16) yielded very similar results (Fig 7.3A, blue), indicating that this approximation is reasonable for barrier heights up to at least $6 k_{\text{B}}T$. However, the large-barrier approximation (Eq. 7.3) tended to overestimate D (Fig 6.3A, cyan), with the estimate worsening noticeably (albeit not dramatically) for barrier heights below

$\sim 3 k_B T$. We note that the $\sim 6\%$ overestimate of D from the high-barrier approximation persists even in the high-barrier limit. Applying the same kind of analysis to τ_{TP} , we found that the values of D returned by the exact solution, Eq. 7.13 (Fig 7.3B, red), again agreed very well with the value imposed in the simulations over the range of barrier heights used (Fig 7.3B, black), whereas the high-barrier approximation (Eq. 7.1) tended to overestimate D , with the overestimation increasing noticeably but not dramatically for barrier heights $\sim 4 k_B T$ and below (Fig 7.3B, blue). In contrast, the low-barrier approximation (Eq. 7.18) tended to underestimate D somewhat for barrier heights $\sim 5 k_B T$ and above (Fig 7.3B, blue).

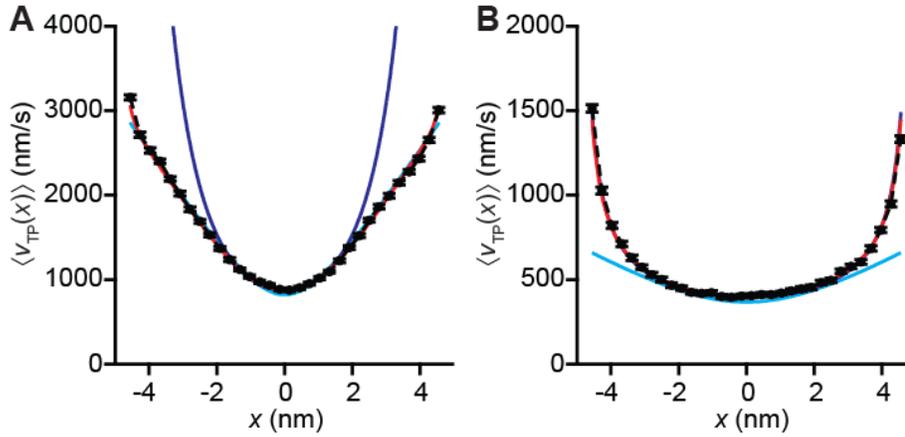


Fig. 7.1: Average transition-path velocity from simulated transitions over harmonic barriers. (A) The average transition-path velocity $\langle v_{TP}(x) \rangle$ found directly from simulations of transitions over a large, $15-k_B T$ harmonic barrier (black) matches very well with the result predicted by the exact solution in Eq. 7.14 (red) and the high-barrier approximation in Eq. 7.3 (cyan), but is poorly described by the low-barrier approximation in Eq. 7.16 (blue). (B) The average velocity found directly from simulations of transitions over a small, $3-k_B T$ harmonic barrier (black) agrees well with the predictions of both the exact result (red) and the low-barrier approximation (blue), but is poorly described by the high-barrier approximation (cyan).

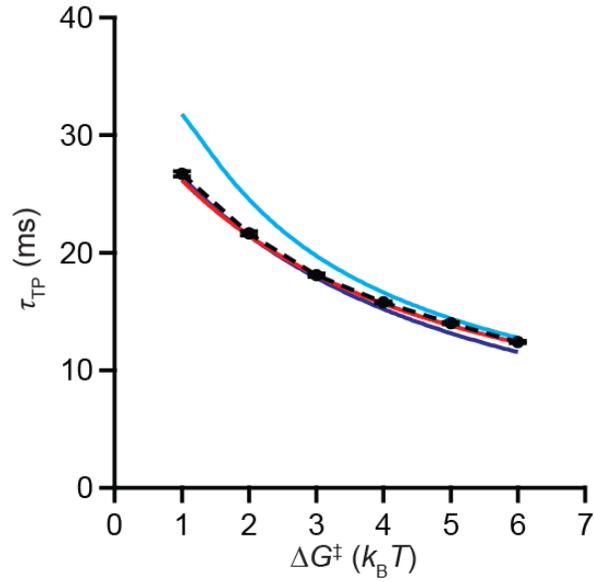


Fig. 7.2: Average transition-path time from simulated transitions over harmonic barriers. The average transition-path time determined directly from simulations over a harmonic barrier with height ranging from 1 to 6 $k_B T$ (black) agrees well with the exact solution from Eq. 7.13 (red) over the whole range of barrier heights. The high-barrier approximation from Eq. 7.1 (cyan) agrees well at barrier heights of $\sim 5 k_B T$ and higher, whereas the low-barrier approximation from Eq. 7.18 (blue) agrees well at barrier heights of $\sim 4 k_B T$ and lower.

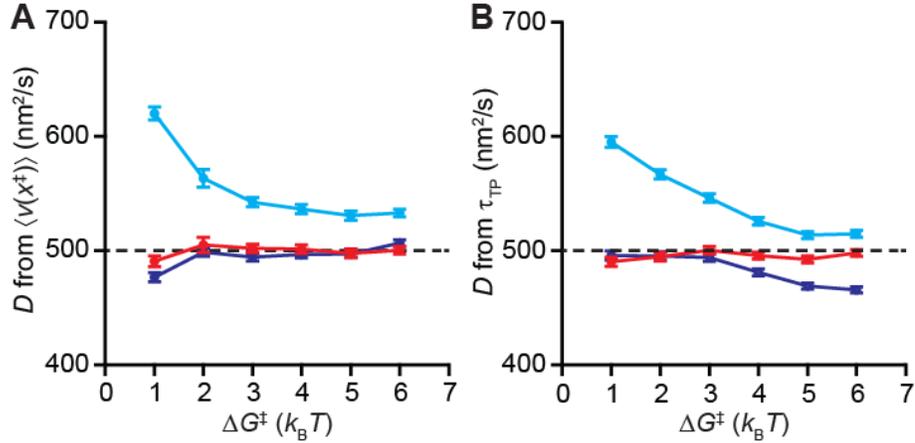


Fig. 7.3: Estimates of the diffusion coefficient for simulations over harmonic barriers. (A) The estimate for D from the exact solution for $\langle v(x^\ddagger) \rangle$ via Eq. 7.15 (red) agrees with the value imposed in the simulation (black) across the whole range of barrier heights, whereas the high-barrier solution (cyan) systematically overestimates D , especially for barriers below $\sim 4 k_B T$. (B) The estimate for D from the exact solution for τ_{TP} (red), determined from the simulation results by solving Eq. 7.13 for D , agrees with the value imposed in the simulation (black) across the whole range of barrier heights, whereas the high-barrier solution (cyan) overestimates D for barriers below $\sim 5 k_B T$ and the low-barrier solution (blue) underestimates D for barriers above $\sim 4 k_B T$.

7.5 Application to anharmonic barriers

Next, we looked at the effects of relaxing the assumption that the barrier is harmonic by simulating transitions over two types of anharmonic barriers: one with a barrier that is sharper than in the harmonic case, described by $G(x) = -\Delta G^\ddagger [\tanh(x/2)/\tanh(L/2)]^2$ (Fig 7.4A, inset), where ΔG^\ddagger is the barrier height; and the other with a barrier that is broader than in the harmonic case, described by $G(x) = -\Delta G^\ddagger [\sinh(x/2)/\sinh(L/2)]^2$ (Fig 7.4C, inset). In each case, ΔG^\ddagger was varied from 1 to 6 $k_B T$, as for the harmonic barriers above. Considering first the sharper barrier, we found excellent agreement between the D values predicted from $\langle v(x^\ddagger) \rangle$ by Eq.

7.11 (Fig 7.4A, red) and the imposed value (Fig 7.4A, black). However, approximating the barrier as harmonic and using Eq. 7.3 for the high-barrier limit yielded values that were systematically too low (Fig 7.4A, cyan), diverging from the imposed value for barrier heights lower than $\sim 5 k_B T$. The same held true when using Eq. 7.16 for the low-barrier limit—it was no better even at low barrier heights. The value of D recovered from τ_{TP} from the exact solution, Eq. 7.13 (Fig. 7.4B, red) agreed well with the imposed value (Fig 7.4B, black), but the harmonic approximations in the high-barrier (Fig 7.4B, cyan) and low-barrier (Fig 7.4B, blue) limits both yielded values that were several-fold lower for all barrier heights examined. Turning next to the barrier that was broader than harmonic, the value of D recovered from $\langle v(x^\ddagger) \rangle$ by Eq. 7.11 (Fig 7.4C, red) agreed well with the imposed value, as above, but now the approximation as a large harmonic barrier overestimated D by 2–3-fold (Fig 7.4C, cyan), whereas the small harmonic barrier approximation overestimated D by only 10–40% (Fig 7.4C, blue). Looking instead at D from τ_{TP} , the exact solution (Fig. 7.4D, red) recovered the imposed value (Fig 7.4D, black), and the two harmonic barrier approximations again tended to overestimate D , except in the case of the large-barrier approximation at the smallest barrier heights, where the solution yielded an unphysical negative value (Fig 7.4D, cyan).

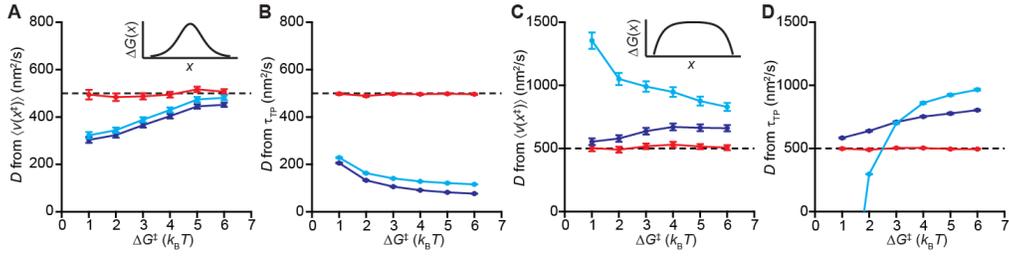


Fig 7.4: Estimates of the diffusion coefficient for anharmonic barriers. (A) The estimate for D from the exact solution for $\langle v(x^\ddagger) \rangle$ via Eq. 7.11 (red) agrees with the value imposed (black) in simulations of transitions over a barrier with higher curvature at the top than a harmonic barrier of the same width and height (inset) across the whole range of barrier heights, whereas the high-barrier (cyan) and low-barrier (blue) harmonic approximations both underestimate D for barriers below $\sim 5\text{--}6 k_B T$. (B) The estimate for D from the exact solution for τ_{TP} via Eq. 7.13 (red) agrees with the value imposed in the simulations (black) across the whole range of barrier heights, whereas the high-barrier (cyan) and low-barrier (blue) harmonic approximations both significantly underestimate D for all barrier heights. (C) For simulations over a barrier with lower curvature at the top than harmonic (inset), estimates of D from the exact solution for $\langle v(x^\ddagger) \rangle$ (red) agree with the imposed value (black), whereas the high-barrier (cyan) and low-barrier (blue) harmonic approximations both overestimate D ; the low-barrier approximation performs markedly better, especially for small barriers. (D) The estimate for D from the exact solution for τ_{TP} (red) agrees with the value imposed in the simulations (black) for all barrier heights, whereas the high-barrier (cyan) and low-barrier (blue) harmonic approximations overestimate D ; at the smallest barriers, where the high-barrier approximation yields unphysical negative values.

7.6 Application to experimental data

The results from analysis of the simulated transitions show the kind of errors that can arise from approximating the barrier shape. Indeed, it was previously noted

that applying Eqs. 7.1–7.3, which assume large harmonic barriers, to transition paths measured for DNA hairpins using optical tweezers(43, 102) led to small but systematic differences between the values of D recovered from different physical quantities (average transition-path times, decay of the transition-time distributions, and barrier-top velocities). A breakdown in the assumptions underlying Eqs. 7.1–7.3 was proposed to explain the discrepancies in the results. We therefore applied the formulae above to test the extent to which the assumption of a large harmonic barrier distorted the results of the analysis.

We used the constant- D expressions for $\langle v(x^\ddagger) \rangle$ and τ_{tp} in Eqs. 7.11, 7.13, 7.15, and 7.18 to re-evaluate the value of D implied by the measurements of two hairpins, based on the shape of the energy barrier that had previously been measured for each hairpin.(43, 75) For hairpin 20R25/T4, which has a barrier that is close to harmonic but relatively small (height measured with respect to the boundaries of the barrier region of $2.3 k_{\text{B}}T$),(43) we found that applying Eq. 7.15 to estimate D from τ_{TP} yielded a new value of $D = 1.9 \pm 0.5 \times 10^5 \text{ nm}^2/\text{s}$, down from the previous estimate(43) using Eq. 7.1 of $D = 2.6 \pm 0.3 \times 10^5 \text{ nm}^2/\text{s}$. Applying instead Eq. 7.18 to estimate D from $\langle v(x^\ddagger) \rangle$ yielded a new value of $D = 2.1 \pm 0.3 \times 10^5 \text{ nm}^2/\text{s}$, again reduced from the previous estimate(102) using Eq. 7.3 of $D = 2.6 \pm 0.5 \times 10^5 \text{ nm}^2/\text{s}$. These new estimates were both closer to the value obtained from Eq. 7.2 ($D = 1.3 \pm 0.2 \times 10^5 \text{ nm}^2/\text{s}$) than the previous estimates, reducing the variance between the three estimates over 3-fold compared to the previous work.(102) Note that these values on the order of $10^5 \text{ nm}^2/\text{s}$ are consistent with previous measurements and modeling of nucleic acid folding (42).

We repeated this analysis for hairpin 30R50/T4, which has a barrier that is roughly twice as large as the barrier for hairpin 20R25/T4 but more anharmonic. Applying Eq. 7.13 to estimate D from τ_{TP} led to a new value of $D = 2.4 \pm 0.3 \times 10^5$ nm²/s, instead of $3.5 \pm 0.3 \times 10^5$ nm²/s using Eq. 7.1,(43) whereas applying Eq. 7.11 led to a new estimate from $\langle v(x^\ddagger) \rangle$ of $D = 2.0 \pm 0.4 \times 10^5$ nm²/s, instead of $2.5 \pm 0.4 \times 10^5$ nm²/s using Eq. 7.3.(102) Again, both of these new estimates are more consistent with the value found from Eq. 7.2, $D = 1.8 \pm 0.2 \times 10^5$ nm²/s, in this case reducing the variance between the different estimates roughly 8-fold compared to what was found previously (43). For both hairpins, then, going beyond the large harmonic barrier approximation helped deliver improved, more consistent estimates of D .

7.7 Discussion of results

The harmonic approximation for energy barriers is almost always used to interpret folding data, not only in the context of transition paths as discussed here but also in terms of rates, such as in the celebrated expression of Kramers for diffusive barrier crossing.(18) The large harmonic barrier approximation has worked reasonably well in initial studies of transition paths in proteins and nucleic acids, for example yielding values of D that are consistent within factors of order unity with values obtained via other means like Kramers' theory.(44, 74) However, as future experiments aim to increase the precision with which fundamental quantities like D are determined from transition-path measurements, expressions for key observables like $\langle v_{\text{TP}}(x) \rangle$ and τ_{TP} that enable analysis to go beyond the limit of large harmonic barriers and incorporate the effects of small barrier heights or

anharmonicity must be applied. Our work shows that the general expressions for $\langle v_{\text{TP}}(x) \rangle$ and τ_{TP} derived in Ref. (94) and the low-barrier limits derived herein can indeed help improve the precision of the data analysis.

The results from analyzing the simulations indicate some of the trends that can be expected from using specific approximations when analyzing experimental data. For harmonic barriers, the large-barrier assumption generated only modest errors, even down to barriers as small as $1 k_{\text{B}}T$, where the actual D , for example, was $\sim 20\text{--}25\%$ smaller than estimated. Nevertheless, the small-barrier approximation provided distinct improvements at the lowest barrier heights. The presence of anharmonicity, on the other hand, led to much larger errors when applying the large and small harmonic barrier approximations. For barriers that have higher curvature at the top than harmonic barriers of the same height and width would have, the estimates of D are systematically lower than they should be, because the high curvature of the harmonic approximation implies a much higher barrier than is actually the case with the anharmonic barrier, and hence a faster velocity and shorter transit time than actually occurs. This explanation also accounts for the poor performance of the low-barrier approximation: even for nominally small barriers, the implied harmonic barrier height is still large. For barriers that are flatter than harmonic, on the other hand, D is overestimated by applying the harmonic approximation, because the actual barrier is higher than implied by the harmonic approximation and hence the velocity and transit time are faster than would be expected in the approximation. Again, this explanation also accounts for the improved performance of estimates using the small-barrier

approximation. Taken together, these results imply that estimates of D using the harmonic approximation are considerably more sensitive to the presence of anharmonicity than they are to the height of the barrier. Note that the fact that barrier anharmonicity has a greater effect on D than barrier height could be expected based on Eq. 7.1, where the transition path time has a linear dependence on the barrier's curvature but a much weaker, logarithmic dependence on its height.

From the simulation results, we can also determine the cross-over point at which the high-barrier approximations start to generate noticeable errors. In almost every case, the high-barrier approximations began to diverge from the low-barrier or exact results at a barrier height of $3\text{--}5 k_B T$, allowing the high-barrier limit to be defined fairly confidently as anything above this level. Previous measurements of transition paths have examined molecules with barrier heights close to but sometimes slightly below this boundary, explaining why the high-barrier approximation worked reasonably well but generated minor discrepancies between estimates of quantities like the diffusion coefficient that were obtained from different methods. As seen above, applying more appropriate approximations or exact formulas helped to resolve these discrepancies.

The main challenge with applying the equations for τ_{TP} and $\langle v_{\text{TP}}(x) \rangle$ is that they require detailed knowledge of the shape of the energy barrier. Energy-barrier shapes can in some cases be measured directly, as with the single-molecule force spectroscopy methods used here,(12) or else they may be estimated from simulations and/or modeling.(21, 107–109) Ideally, the position-dependence of the diffusion coefficient should also be known, since the common approximation that

D is constant is not generally correct.(89, 110) This position dependence can in some cases be deduced by combining experiments and modeling,(98) but reliable methods for direct measurement are exceedingly difficult to implement(111). Indeed, the first successful measurement of the position dependence of D is presented in chapter 8 of this thesis. Hence the constant- D approximation may often be unavoidable in practical applications.

Although we have used nucleic acid folding to demonstrate the application of Eqs. 7.10–7.18, these results are, of course, equally applicable to protein folding. Theoretical considerations and experimental measurements(16, 28, 108, 112–114) both suggest that energy barriers for protein folding are generally small compared to the total free-energy change. In many cases, they may be close to the $3\text{--}5 k_B T$ limit for assuming a large barrier, making the results presented here useful for obtaining better estimates of the transition-path properties. The small-barrier limit will be particularly relevant for studies of molecules with very small or even non-existent barriers, such as fast-folding or downhill-folding proteins, which continue to be the target of considerable interest because their small barriers allow transition states to be populated more easily.(115)

Chapter 8

Transition state dynamics

This chapter builds on an interesting observation from the velocity measurements presented in chapter 6, that being the distinct pauses present in the velocity profiles. These pauses are predicted in the Bryngelson and Wolynes landscape picture and are thought to arise as the molecule encounters micro-barriers or micro-wells in the energy landscape. When such a barrier is encountered the molecule momentarily pauses until it receives sufficient thermal energy to cross the barrier. In this chapter we use the statistics of the pauses in order to characterize the micro-barriers and validate the Bryngelson and Wolynes landscape model on the scale of individual nucleobases.

The material in this chapter was originally written in the manuscript “Direct observation of the residue-by-residue search for native structure during folding of a single molecule”, Noel Q. Hoffer, Krishna Neupane, and M.T. Woodside. At the time of writing this thesis, the manuscript remains unpublished. For this paper NQH and KN took the data, NQH performed the analysis, NQH and MTW wrote the paper.

8.1 Pausing along transition paths

In this study we build on the success of transition path velocity measurements, identifying periods of extremely slow transition path velocity, where the molecules are briefly paused. In these pauses, the molecule was captured fleetingly in a transition state, allowing individual high-energy states within the

barrier to be observed and studied directly. We used velocity thresholding(116) to identify the location and duration of pauses in over 326,000 transition paths for hairpin 30R50/T4. As discussed in chapter 6, the velocity profile of each path (Fig. 8.1, upper panel, blue) was calculated from the local slope of the trajectory, after first applying a smoothing spline interpolation to reduce the effects of measurement noise (Fig. 8.1, lower panel, red), and then pauses were defined as those parts of the trajectory where the speed dropped below a threshold equal to 10% of the average transition-path velocity (Fig. 8.1, magenta). We verified that these pauses arose from the motions of the hairpin itself, not as artifacts from the dynamics of the handles and beads attached to the hairpin, by measuring reference constructs lacking the hairpin, which we abruptly stretched and relaxed (Fig. 8.3A) to generate extension changes similar to those seen during hairpin folding transitions (Fig. 8.3B). Analyzing the ‘transitions’ in these measurements exactly as for the hairpin constructs, we found that ‘pauses’ were shorter and detected much more rarely (Fig. 8.4).

From the probability density for finding pauses of a given duration as a function of position within the barrier region (Fig. 8.2A), we found that although pauses were not detected in all transitions (an average of ~ 0.4 pauses per transition), they occurred ubiquitously across the entire barrier. The distribution of pause locations was the same for both folding (Fig. 8.2B, black) and unfolding (Fig. 8.2B, red), reflecting the microscopic reversibility of the transition paths; it showed a noticeable dip in the number of pauses detected in the highest-energy states near the top of the energy barrier (Fig. 8.2B, purple).

These pauses in the transition paths are especially interesting because by capturing the molecule briefly in small kinetic traps in the barrier region, we can study how the molecule moves into and out of the transition states. Pauses could reflect the activation energy needed to form or break base-pairs, which presumably generates local corrugations in the energy landscape. They might also reflect periods where the molecule is re-configuring in ways that leave the molecular extension effectively unchanged, such as through rotations of the bases or changes in the sugar pucker (Fig 8.5): projecting the full, multi-dimensional dynamics onto a one-dimensional (1D) reaction coordinate gives rise to local diffusivity changes and effective landscape roughness(89). To probe the motions into and out of the microstates captured during the pauses in greater detail, we analyzed the distribution of pause durations.

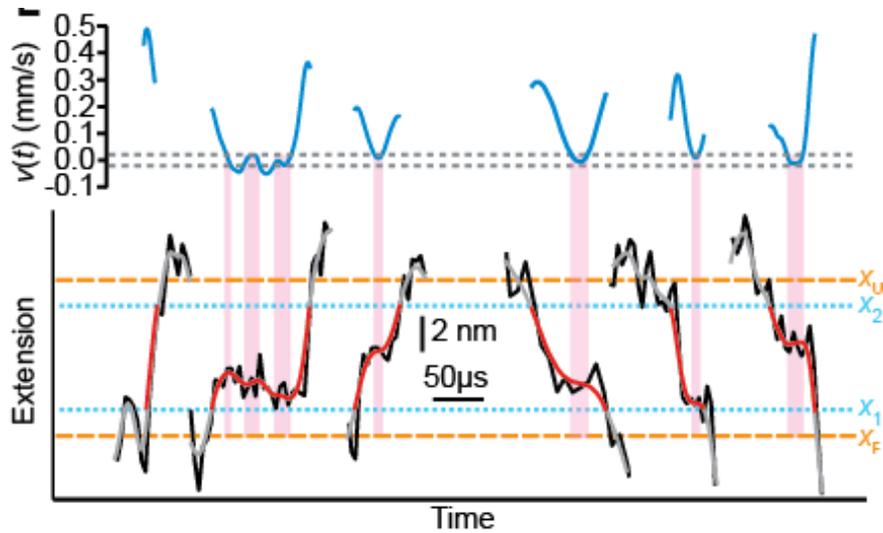


Fig 8.1: Observing pausing along transition paths. Transition-path trajectories (lower panel, black) showed a wide range of behavior. The velocity profiles of individual trajectories (upper panel, blue) were obtained by numerical differentiation of the spline-smoothed trajectories (lower panel, grey) of the extension. Pause locations and durations were identified from the portions of the trajectories in which the speed remained under a threshold (upper panel, grey) equal to 10% of the average transition speed (magenta bands).

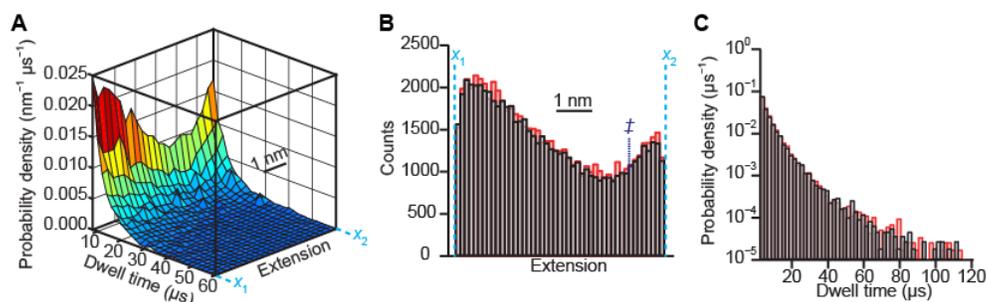


Fig 8.2: Pausing within the transition states. (A) Probability density for finding pauses of a given duration at a given location within the barrier region for hairpin 30R50/T4. (B) The distribution of pause locations is the same for folding (black) and unfolding (red). Pauses occurred ubiquitously across the entire transition region but were least likely to occur near the barrier top (\ddagger). (C) The distribution of pause durations is the same for folding (black) and unfolding (red), dropping sub-exponentially in each case.

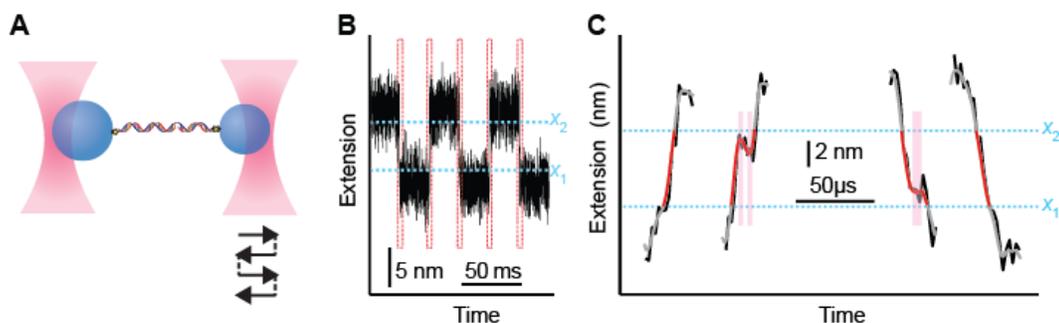


Fig 8.3: Measurements of reference construct. (A) Schematic of reference construct measurements. A construct consisting of two DNA handles without any hairpin was attached two beads held under tension between two traps, at the same force as the hairpin measurements. The traps were abruptly moved away and back together, causing the beads to be displaced by the same distance as observed in folding trajectories. (B) Extension trajectory resulting from bead motions (black) showing ‘transitions’ (red boxes). (C) Samples of trajectories from the reference construct (black) showing ‘transition paths’ (red) within the boundaries x_1 and x_2 (cyan), identifying ‘pauses’ (magenta) exactly as for the hairpin folding trajectories.

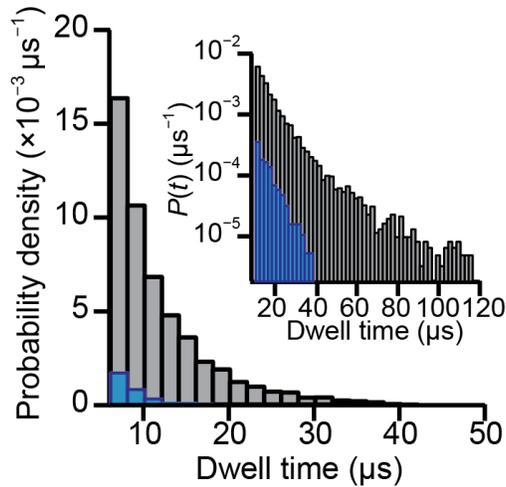


Fig 8.4: Distribution of ‘pauses’ in reference construct trajectories. Comparing the absolute probability of finding ‘pauses’ of a given duration in the ‘transitions’ for the reference construct (blue) to the probability in samples with the hairpin present (black) shows that the dynamics of the handles and beads did not contribute significantly to the pause distribution: the ‘pauses’ in the absence of the hairpin were much shorter and much rarer.

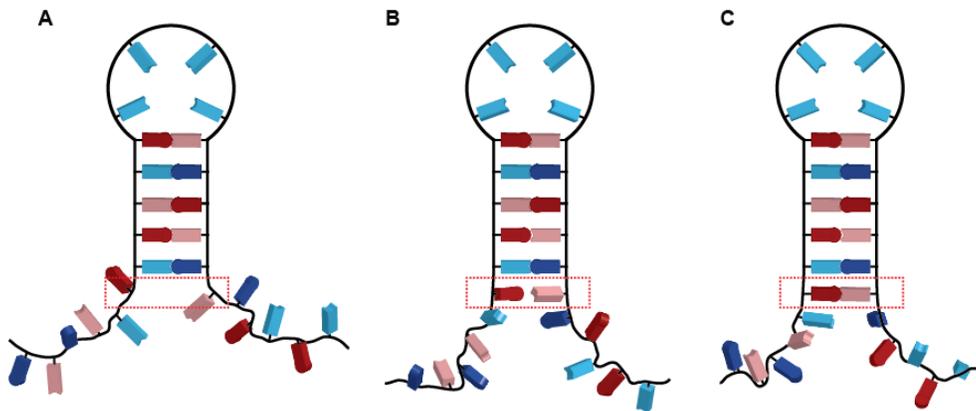


Fig 8.5: Microscopic motions during hairpin folding. To extend a partially folded helix (A), the next set of nucleotides (red box) must make several bond-angle rotations in order to be positioned for successful pairing. Even when the two bases fluctuate into close proximity (B), the glycosidic bond angles may still be unfavorable, requiring additional bond rotations before (C) the base-pair is formed and the attempt to form the next base-pair begins.

8.3 Pause durations

The distribution of pause durations, $p(t_{\text{pause}})$, was also the same for folding (Fig. 8.2C, black) and unfolding (Fig. 8.2C, red), dropping sub-exponentially in each case. The decay of $p(t_{\text{pause}})$ is clearly not single-exponential (Fig. 8.6A, cyan), hence the local energy fluctuations trapping the molecule cannot all have the same size. Indeed, at least two different micro-barrier heights should be expected, since each type of base-pair in the hairpin (A:T and G:C) has its own activation barrier, leading to at least two distinct rates for escaping the micro-wells generated by native base-pairing. A double-exponential decay fits $p(t_{\text{pause}})$ quite well (Fig. 8.6A, blue), consistent with a purely 1D zippering model in which activation energies for base-pair formation create local kinetic traps (Fig. 8.7A). To test if this model is correct, we repeated the pausing analysis for measurements of hairpins containing only A:T base-pairs (hairpin 20R0/T4, ~146,000 transition paths) or G:C base-pairs (hairpin 20R100/T4, ~64,000 transition paths). For both hairpins, pausing was again seen throughout the barrier region (Figs. 8.6B & C, inset). However, because only one type of base-pair is present in these hairpins, the 1D zippering model would predict a single-exponential decay for $p(t_{\text{pause}})$ (Fig. 8.7B). The fact that single-exponential fits are poor for both hairpins (Figs. 8.6B & C, blue) indicates that the micro-barriers inducing pauses must arise from more than just simple 1D zippering.

To account for these observations, we turned to a classic microscopic theory of folding that models it in terms of a residue-by-residue search through both native and non-native conformations(29). This theory assumes a continuous range of

micro-well depths (ΔE) owing to energy fluctuations from both native and non-native interactions, multiple possible conformations per residue (ν), and a constant transition rate between micro-conformations (R_0). It predicts a log-normal distribution of rates for escaping micro-wells within the energy landscape:

$$P(k) \propto \exp[-(\ln k - \mu)^2 / 2(\beta \Delta E)^2], \quad (8.1)$$

where k is the rate, β is the inverse thermal energy, and μ is a function of ν , R_0 , ΔE , and the number of residues. Fitting $p(t_{\text{pause}})$ to Eq. 8.1, we found excellent agreement for all three hairpins (Fig. 8.6, red), indicating that the microscopic theory is consistent with all of the measurements.

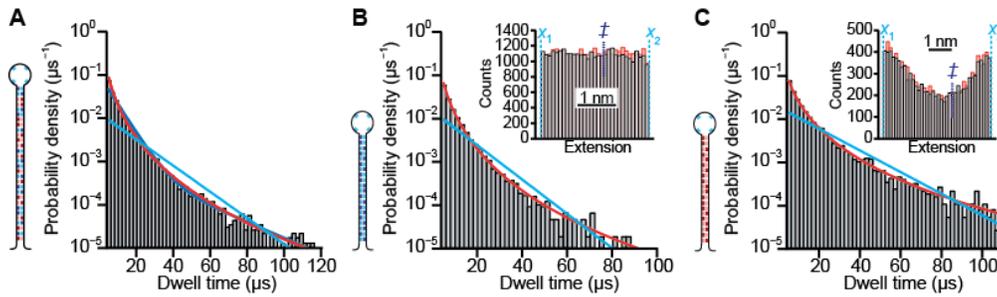


Fig 8.6: Pause durations. (A) The distribution of pause durations for hairpin 30R50/T4 (black) was poorly fit by a single-exponential decay reflecting a single rate constant (cyan) but was well fit by both the double-exponential decay expected if pauses arise from micro-barriers associated with A:T and G:C base-pair formation (blue) and by the log-normal distribution of rate constants expected from a microscopic theory of folding as a search through non-native states (red). (B) The pause duration distribution for hairpin 20R0/T4, containing only A:T base-pairs, was fit well by a log-normal distribution of rate constants (red) but not by a single-exponential decay (cyan). Inset: pauses occurred at the same locations for both folding (black) and unfolding (red). (C) Similar results were found for hairpin 20R100/T4, containing only G:C base-pairs.

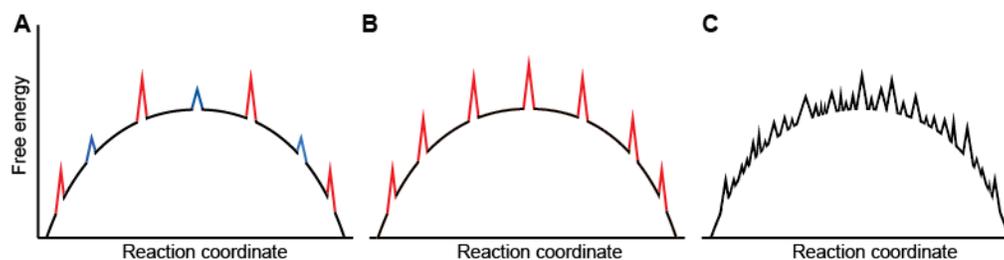


Fig 8.7: Models of microwells in the barrier region. (A) If the activation energy for forming and breaking base-pairs dominates the energy fluctuations in the barrier region, the presence of both G:C and A:T base-pairs will lead to at least two distinct sizes of micro-barriers (respectively red and blue) on top of the underlying free energy barrier (black). (B) If only a single type of base-pair is present, the micro-barriers from the base-pairing activation energy (red) will all have a similar height. (C) The microscopic model for searching through many non-native conformations predicts a random, Gaussian-distributed set of micro-barriers.

8.4 Diffusion coefficients from pausing statistics

As an additional test that this model describes the pauses well, we used the results of the fitting (Table 8.1) to estimate the diffusion coefficient, D . D is one of the key physical parameters in the energy landscape description of folding, because it encapsulates the speed at which the landscape is explored on the microscopic scale, relating the kinetics of structure formation to the thermodynamics of the landscape and reporting on the internal friction in the molecule(117). In terms of the fits to Eq. 8.1, we have(29):

$$D \approx \frac{\delta x^2}{4} \exp\left[\mu - (\beta\Delta E)^2 / 2\right], \quad (8.2)$$

where δx is the extension change associated with forming/breaking a single base-pair (~ 0.9 nm at the forces used here), the minimal increment for changing the

number of native contacts. For hairpin 30R50/T4, we found $D = 1 \times 10^{5 \pm 0.4} \text{ nm}^2/\text{s}$, in good agreement with the values $D \sim 2\text{--}4 \times 10^5 \text{ nm}^2/\text{s}$ estimated previously from multiple approaches (rates(118), average transition time(44), transit-time distribution(44), and transition-path velocity(102)). This result also agreed well with the value estimated crudely from the average pause duration, $\langle t_{\text{pause}} \rangle$, via(29) $D \sim \langle \delta x^2 \rangle / 2 \langle t_{\text{pause}} \rangle$: $D = 1 \times 10^{5 \pm 0.3} \text{ nm}^2/\text{s}$. Similarly reasonable agreement was found estimating D for hairpins 20R0/T4 and 20R100/T4 from $p(t_{\text{pause}})$ by fitting to Eq. 8.1, respectively $D = 1 \times 10^{5 \pm 0.5}$ and $5 \times 10^{4 \pm 0.6} \text{ nm}^2/\text{s}$, and from $\langle t_{\text{pause}} \rangle$, respectively $D = 8 \times 10^{4 \pm 0.3}$ and $7 \times 10^{4 \pm 0.4} \text{ nm}^2/\text{s}$. An intriguing aspect of this analysis is that it does not require any information about the shape of the free energy landscape, allowing us to obtain what is, to our knowledge, the first estimate of D for barrier crossing in a folding reaction without the need to reconstruct the energy landscape.

Hairpin	$e^\mu \text{ (s}^{-1}\text{)}$	$\sigma \text{ (} k_B T \text{)}$
20R0/T4	1.3 ± 0.4	1.1 ± 0.2
20R100/T4	0.5 ± 0.4	1.1 ± 0.2
30R50/T4	0.7 ± 0.3	1.0 ± 0.1

Table 8.1: Results from fitting pause-duration distributions to Eq. 8.1. The fitting parameters σ and μ are used in Eq. 8.2. Errors represent the standard error from bootstrap analysis.

Finally, we examined how D varies within the barrier region based on the position-dependence of the pausing. Despite the fact that D is known to vary with position along the reaction coordinate(89), it is almost universally assumed to be

constant when analyzing experiments, because its position dependence is difficult to measure. Several methods to measure $D(x)$ from single-molecule experiments have been proposed, but experimental artifacts make them unreliable(111), leaving $D(x)$ experimentally inaccessible. By fitting $p(t_{\text{pause}})$ to Eq. 8.1 for the pauses in each 1-nm segment of the barrier region, we found that $D(x)$ for hairpin 30R50/T4 was not constant, but rather had a distinct peak near the top of the barrier, where the pauses were least frequent (Fig. 8.8A, black). The same pattern was seen when determining $D(x)$ from $\langle t_{\text{pause}} \rangle$ (Fig. 8.8A, red). Intriguingly, the diffusivity profile was somewhat different for each hairpin (Fig. 8.8B), being close to constant for 20R0/T4, sharply peaked for 20R100/T4, and in between for 30R50/T4, again mirroring the frequency of pausing.

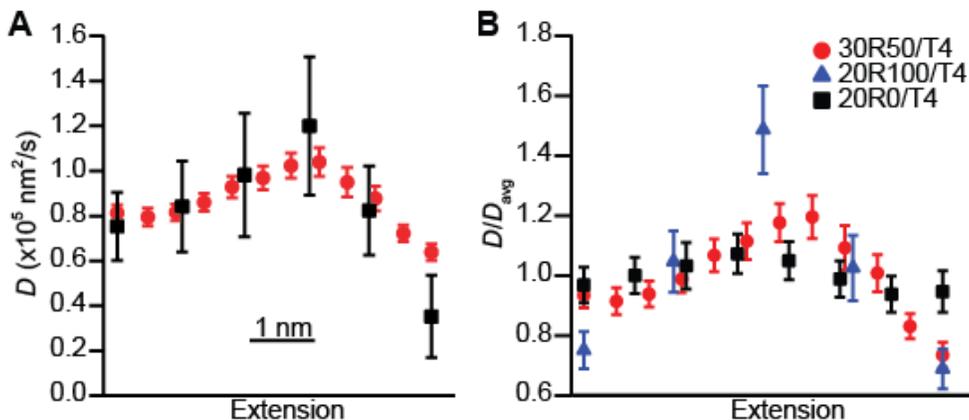


Fig. 8.8: Position dependence of D . (A) Calculating D as a function of position along the reaction coordinate from the average pause duration at different positions (red) reveals that D is not constant, but rises to a peak near the middle of the barrier region. A similar pattern is seen from Eq. 8.2 using fits of the position-dependent pause durations (black). (B) D showed a different position dependence for the 3 hairpins studied: flat for hairpin 20R0/T4 (black), moderately peaked for hairpin 30R50/T4 (red), and strongly peaked for hairpin 20R100/T4 (blue). Error bars represent standard error of the mean.

8.5 Discussion of transition path pausing results

The results presented here provide powerful insight into the microscopic processes occurring during folding. The fact that the distributions of transition-path pauses are best fit by the microscopic kinetic model implies that these pauses reflect the requirement for each residue to search through multiple non-native conformations in order to find the correct structure for base-pairing. Indeed, it is the energy fluctuations involved in exploring these non-native conformations—rather than the activation barriers for native base-pair formation—that appear to dominate the microscopic molecular motions. Crucially, such a residue-by-residue search for the correct conformation reveals the previously undetectable elementary steps underlying folding reactions. By showing that folding can be described consistently across all timescales, from the rapid rates for the fleeting motions between transition states to the vastly slower rates for the macroscopic folding/unfolding transitions, these results quantitatively validate the microscopic theory of folding.

The symmetry in the pausing behavior for unfolding and refolding might at first glance be puzzling, since conceptually unfolding and refolding appear to be different. In folding reactions, it is intuitive to imagine how a pause might manifest, for example with the molecule transiently arrested in a conformation with bond angles that prevent base-pairing: before the next base-pair could be formed, the energy barrier for bond rotation would need to be crossed, involving motions orthogonal to the observed reaction coordinate that produce a brief pause. In contrast, native base-pairs are being broken during unfolding, hence naively one

might expect different behavior. However, even in a transition path that ultimately leads from folded to unfolded state, the microscopic behavior involves ubiquitous reversals of motion leading to re-pairing of bases(102), resulting in the microscopic reversibility that is both observed and expected.

Our focus in this study on DNA hairpins was driven by the fact that their zippering mechanism of folding is particularly simple and straightforward to interpret: under applied tension, base-pairs are constrained to zip/unzip sequentially and produce uniform length changes. Such hairpins likely come as close as possible to realizing an ideal 1D system in folding, and indeed the transition-path properties of hairpins studied to date have all reflected close to ideal 1D behavior(14, 15, 84, 102, 118–121). Examining the transient pauses in the transition paths, however, reveals the influence of motions in axes orthogonal to the zippering, reflecting the fundamental multidimensionality of the folding reaction.

The sensitivity of pausing to the multidimensionality of folding likely accounts for its ability to reveal the position-dependence of the diffusivity, which is otherwise difficult to detect reliably(111). The diffusivity profiles observed here involve relatively modest variations in D , generally within the range $\pm 50\%$, consistent with previous work suggesting that the variations are not large(111). The diffusivity profiles are also all peaked near the middle of the reaction coordinate, consistent with the results of simulations of more complex molecules like proteins when using the fraction of native contacts (equivalent to end-to-end extension here) as the reaction coordinate(89). Curiously, however, the amplitude of the variation differs for the 3 hairpins studied. These differences may reflect the effects of force,

since higher force is expected to make extension a better reaction coordinate(99) and 20R100/T4 unfolds at twice the force of 20R0/T4(122). We speculate that at higher forces there are fewer excursions into hidden coordinates not captured by the extension, leading to reduced pausing, especially near the barrier top that dominates the folding mechanism (and hence reaction coordinate quality), thereby accounting for a more dramatic increase D near the barrier for transitions occurring at higher force.

To conclude, this work illustrates how measurements of the motions within transition paths provide a powerful tool for probing the elementary, microscopic events underlying folding—including the brief pauses associated with finding the correct conformation for stacking and pairing of bases—elucidating previously inaccessible properties and opening a new frontier in folding studies.

Chapter 9

The average shapes of transition paths

While the previous chapters have focused on the average values and distributions of quantities such as the transition-path times and velocities, this chapter examines the shapes of transition paths. Transition-path shapes encode key information about folding mechanisms because they represent the sequence of microscopic motions followed while traversing the critical transition states within the energy barrier. This chapter presents the first measurements of transition-path shapes, determining the average shapes for the folding of DNA hairpins. We find that time-domain averages are sensitive to instrumental artifacts, but extension-domain averages match the dominant path shape predicted theoretically for diffusion over a 1D landscape. Path-shape variances reveal sequence-dependent changes in the diversity of transition paths, suggesting that multiple parallel pathways through the transition states can be detected simultaneously in a single molecule.

The material in this chapter was originally published in the manuscript N.Q. Hoffer, K. Neupane, A.G.T. Pyo and M.T. Woodside, “Measuring the average shape of transition paths during the folding of a single biological molecule” Proc. Natl. Acad. Sci. 2019. For this paper NQH and KN took the data, NQH and AGTP performed the simulations, NQH performed the analysis and NQH and MTW wrote the paper.

9.1 Motivation for studying transition- path shapes

The studies discussed thus far have yielded important insight into the

microscopic motions involved in folding, but they have been limited to characterizing the average values and distributions of properties like transition-path times and velocities. Such averages and distributions, while reflecting key aspects of the folding, nevertheless do not retain any information about the sequence of events taking place during the transitions. The ability to study sequential relationships between the local motions that constitute the folding reaction promises to unlock a wealth of new information for characterizing folding mechanisms by directly observing the statistical ensemble of pathways populated during a folding reaction. Such direct observation has not yet been possible in experiments, restricting the study of pathway ensembles to simulations (123–130). However, the desired sequential information can in principle be accessed through the shapes of the transition paths, which capture the time spent within each part of the reaction coordinate and how it varies as the energy barrier is crossed.

Recent theoretical work has begun to explore the properties of transition-path shapes, focusing on characterizing the average shape. Kim and Netz (26) considered the shape defined by $\langle t(x) \rangle$, the average of the transition paths in the time domain as a function of position along the reaction coordinate (which effectively reflects the distribution of times required to reach each point x along the transition), deriving expressions for the average path shapes expected for different types of barrier potential. Makarov (23) and Cossio *et al.* (24) took a somewhat different approach, focusing instead on the average shape of the dominant or most probable path, predicting the path shape expected for transitions crossing a harmonic barrier. This work showed that the transition-path shapes are more sensitive to the shape of

the underlying energy barrier than are properties like the transition time, suggesting that path shapes should indeed be a powerful tool for probing the microscopic details of folding.

These theoretical predictions had not been compared to experimental observations, however, because measurements of transition-path shapes had not yet been reported. As a first step towards experimental characterization of transition-path shapes, here we have used single-molecule force spectroscopy measurements of DNA hairpins held under tension in high-resolution optical tweezers as a model system to determine the average shapes of the transition paths computed by averaging in the time domain and in the extension domain. The transition-path shapes were obscured in time-domain averages by the effects of the experimental force probes, but for each hairpin the extension-domain average matched the shape of the dominant transition path predicted for diffusion over a harmonic barrier, yielding diffusion coefficients similar to previous results. We also explored the fluctuations in the path shapes, finding evidence that more than one type of transition path may be present for some hairpins.

9.2 Time domain averaged transition path shapes

Isolating a representative set of individual transitions from within the full folding trajectory for a single hairpin and aligning the transitions to start at the same time so as to characterize the transition-path shapes (Fig. 9.1), a large variety of path shapes was seen for both unfolding (Fig. 9.1A, grey) and refolding (Fig. 9.1B, grey). This wide range of shapes represents many different patterns of motion for the hairpin as it crosses the energy barrier. We first found the average shape of the

transitions for each hairpin by expressing the transition-path trajectories as $t(x)$, aligning them all on the boundaries of the barrier region (x_1 and x_2), and then averaging in the time domain to obtain $\langle t(x) \rangle$. Because many trajectories involved re-crossing events where the hairpin extension turned back on itself before completing the transition (102), we took the average of all the times at which a given x value was crossed in a particular transition path before averaging the values across all different paths, a procedure shown to avoid artefactual over-weighting of transitions that include re-crossing events (24). The averages for the unfolding (Fig. 9.1A, black) and refolding (Fig. 9.1B, red) transition paths for hairpin 30R50/T4, shown overlaid on 30 individual path trajectories (Fig. 9.1A–B, grey), were found to be time-reversal symmetric (Fig. 9.1C), as expected from the microscopic reversibility of transition paths (23). Similar results were found for the other hairpins (Fig. 9.2). Despite the wide variation in the individual path shapes, the average shapes computed from $\langle t(x) \rangle$ were all very similar and quite simple: almost straight lines, but with a very slight sigmoidal curvature where the path-shape curved up on the side of the barrier region nearest the folded state and down on the side nearest the unfolded state.

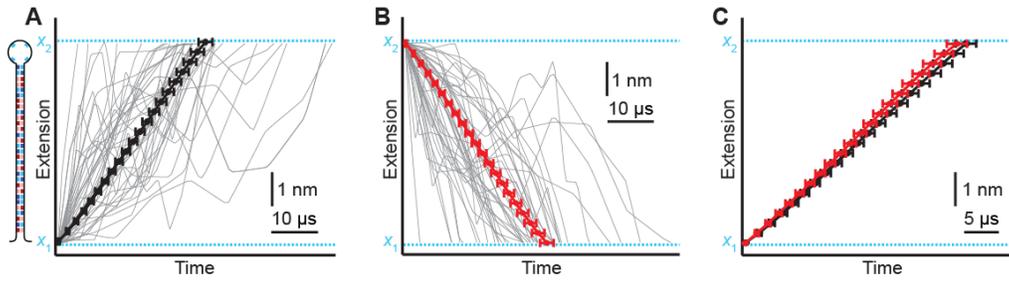


Fig 9.1: Time-domain average transition-path shape $\langle t(x) \rangle$ for hairpin 30R50/T4. (A) The path shape for unfolding averaged in the time domain (black), shown along with 30 individual unfolding transitions (grey), is almost straight but with a slight sigmoidal curvature. (B) The average path shape $\langle t(x) \rangle$ for refolding transitions (red), shown with 30 individual refolding transitions (grey), is the same but time-reversed. (C) The time-reversed path shape for refolding transitions (red) matches the shape for unfolding transitions (black). Error bars represent s.e.m.

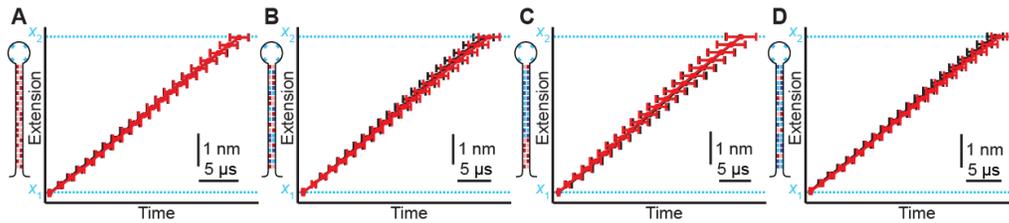


Fig 9.2: Time-domain average transition-path shapes $\langle t(x) \rangle$ for different hairpins. The path shape averaged in the time domain, $\langle t(x) \rangle$, is shown for unfolding (black) and refolding (red, time-reversed) of hairpins (A) 20R100/T4, (B) 20R55/T4, (C) 20TS06/T4, and (D) 20R25/T4.

9.3 Extension domain averaged transition path shapes

The conceptual motivation for calculating path shapes by averaging in the time domain at each extension value, as above, is that every transition passes through the same range of extension values, from x_1 to x_2 , making the simple average $\langle t(x) \rangle$ well-defined. In contrast, a simple average in the extension domain at each time

value—*i.e.*, $\langle x(t) \rangle$ —is ill-defined, because the range of t values varies widely between transitions (44, 76, 101). Experimentally, however, averaging in the extension domain is more natural, since x is the observed variable containing experimental noise that should be averaged, not t . In order to compute an extension-domain average while accounting for the differences in the transition-path times, we averaged all the transitions having a given transition-path time τ_i to obtain $\langle x(t) | \tau_i \rangle$, illustrated in Fig. 9.3A for hairpin 30R50/T4, and then averaged the resulting curves from all τ_i (Fig. 9.3B, grey) in the time domain, weighting by the relative probability for each τ_i value. The average shapes computed in this way (Fig. 9.3B, red), denoted $\langle x(t | \tau) \rangle_\tau$, were the same for unfolding (Fig. 9.3C, black) and refolding (Fig. 9.3C, red), again reflecting the expected time-reversal symmetry. They showed a pronounced curvature that differed from the shape of $\langle t(x) \rangle$: whereas both were sigmoidal, the curvature was opposite in sign and noticeably higher for the extension-domain average (Fig. 9.3D, black) than the time-domain average (Fig. 9.3D, red). Similar results were observed for the other hairpins (Fig. 9.4).

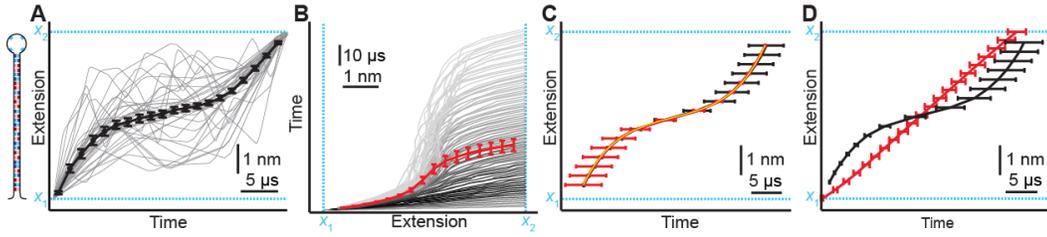


Fig 9.3: Extension-domain average transition-path shape $\langle x(t|\tau) \rangle_\tau$ for hairpin 30R50/T4. (A) The path shape averaged in the extension domain for unfolding transitions with duration τ_{TP} (black), shown along with 30 individual unfolding transitions (grey). (B) The average path shape $\langle x(t|\tau) \rangle_\tau$ for unfolding (red) was computed by averaging the curves $\langle x(t|\tau_i) \rangle$ (grey) obtained from transition paths with duration τ_i , for all τ_i values. Grey-scale coloring indicates the proportion of transitions with duration τ_i . (C) The path shapes $\langle x(t|\tau) \rangle_\tau$ for unfolding (black) and refolding (red, time-reversed) are time reversal symmetric and well fit by Eq.9.1 (yellow). The fits for unfolding and refolding are visually indistinguishable. (D) The time-domain average shape $\langle t(x) \rangle$ (red) differs noticeably from the extension-domain average shape $\langle x(t|\tau) \rangle_\tau$ (black), with the latter having much more pronounced curvature. Error bars represent s.e.m.

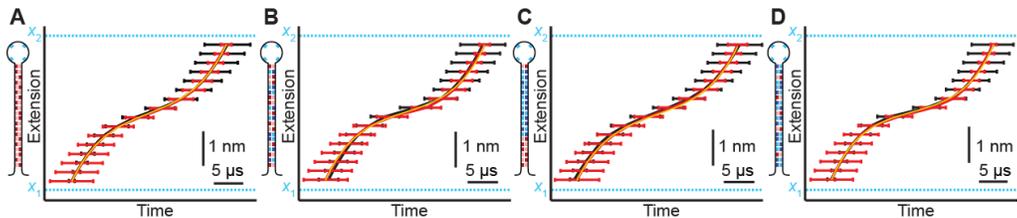


Fig 9.4: Extension-domain average transition-path shapes $\langle x(t|\tau) \rangle_\tau$ for different hairpins. The path shape averaged in the extension domain, $\langle x(t|\tau) \rangle_\tau$, is shown for unfolding (black) and refolding (red, time-reversed) of hairpins (A) 20R100/T4, (B) 20R55/T4, (C) 20TS06/T4, and (D) 20R25/T4. The fits to Eq. 9.1 for unfolding (yellow) and refolding curves (not shown) are visually indistinguishable.

9.4 Exploring the discrepancy between the time and extension domain averages

To understand why these two methods of computing the average shapes yielded different results, we explored Brownian-dynamics simulations of transitions paths. Simulating diffusive crossing of a one-dimensional (1D) harmonic barrier as described previously (102), we found that both $\langle t(x) \rangle$ and $\langle x(t|\tau) \rangle_\tau$ had similar shapes (Fig. 9.5A red and black, respectively), in contrast to the experimental results. Furthermore, both averages closely matched the shape of the dominant transition path (Fig 9.5A, blue) calculated analytically from the potential (22–24). However, previous work has shown that attaching a molecule to a large probe like a bead or cantilever via a compliant linker and monitoring the motions of the probe rather than those of the molecule, as is done in force spectroscopy measurements, adds noise to the measured extension and can alter the properties of the transition paths (24, 56, 57, 59, 111) We therefore extended the simulations to include the effects of linking the molecule to a bead, and recomputed $\langle t(x) \rangle$ and $\langle x(t|\tau) \rangle_\tau$. In contrast to $\langle x(t|\tau) \rangle_\tau$ (Fig. 9.5B, black), which was little changed from the 1D simulation, $\langle t(x) \rangle$ was distorted to become close to linear (Fig. 9.5B, red), analogous to what was observed experimentally. The simulations thus suggest that time-domain averaging can lead to an unreliable picture of the average path shape owing to the effects of the experimental geometry, whereas averaging in the extension domain is more robust.

9.5 Comparison of results to theory

We next compared the average path shapes obtained from the experimental data to theory. Since simulations showed that $\langle x(t|\tau) \rangle_\tau$ provides a reasonable

approximation of the analytical shape of the dominant transition path for harmonic barriers (Fig 9.5A), and since the transition paths in hairpin folding are generally well-described by harmonic-barrier models (14, 43, 119), we fit the average shapes to the functional form of the dominant path shape for a harmonic barrier. In 1D with constant diffusion, this path shape is given by

$$\langle x(t) \rangle_{\text{DTP}} = \frac{L \sinh(\beta\kappa D(t - \tau_{\text{TP}}/2))}{\sinh(\beta\kappa D\tau_{\text{TP}}/2)}, \quad (9.1)$$

where $2L$ is the length of the transition path, κ is the curvature of the barrier, τ_{TP} is the average transition-path time, and D is the diffusion coefficient for motion along the reaction coordinate (24). Fitting $\langle x(t|\tau) \rangle_{\tau}$ for hairpin 30R50/T4 to Eq. 1, while fixing $2L = x_2 - x_1$, using the values of κ found from energy-landscape reconstructions (43), and treating D as a free parameter, we found good agreement with the observed shape (Fig. 9.3C, yellow). The values of D returned by the fits, $6 \pm 4 \times 10^5 \text{ nm}^2/\text{s}$ for both unfolding and refolding, agreed within error with the values in the range $2\text{--}5 \times 10^5 \text{ nm}^2/\text{s}$ found previously for this hairpin by analyzing rates (42), transition-path times (43), and transition-path velocities (102). Repeating this analysis for each hairpin by fitting $\langle x(t|\tau) \rangle_{\tau}$ to Eq. 9.1 (Fig. 9.4), we found similar results in each case (Table 9.1), with values of D that were consistent within error with previous results, although as for hairpin 30R50/T4 they were all systematically somewhat higher.

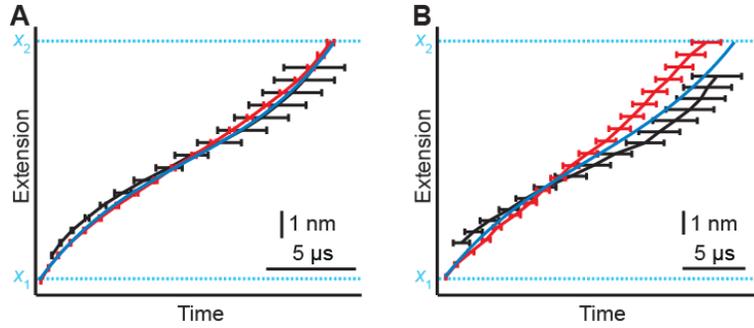


Fig 9.5: Transition-path shapes from Brownian-dynamics simulations. (A) Brownian-dynamics simulations of transitions over a 1D harmonic barrier show that both $\langle t(x) \rangle$ (red) and $\langle x(t|\tau) \rangle_\tau$ (black) match the dominant transition path shape computed analytically (blue). (B) Brownian-dynamics simulations of a bead coupled via a compliant linker to a molecule crossing the same barrier as in (A) show that the bead/linker effects cause $\langle t(x) \rangle$ to become more linear while leaving $\langle x(t|\tau) \rangle_\tau$ similar to the dominant transition path shape.

Hairpin	D ($\times 10^5$ nm ² /s)		
	unfolding	refolding	average
30R50/T4	6 ± 4	6 ± 4	6 ± 4
20R100/T4	7 ± 4	7 ± 4	7 ± 4
20R55/T4	8 ± 4	8 ± 4	8 ± 4
20TS06/T4	4 ± 3	5 ± 3	4 ± 3
20R25/T4	7 ± 3	7 ± 3	7 ± 3

Table 9.1: Diffusion coefficients from fitting dominant path shapes. Errors represent standard deviation.

9.6 Variability in transition path shapes

To obtain a measure of the variability in the transition-path shapes, we calculated the variance in $t(x)$, $\sigma^2_{t(x)}$, for each of the hairpins as a function of the progression along the transition path (Fig. 9.6). The variance was very similar in the unfolding and refolding path shapes, hence the two were averaged for each

hairpin. Not surprisingly, given the broad distribution of transit times (43), the variance in $t(x)$ increased super-linearly with increasing x . To our knowledge, there is not yet any theoretical analysis of path-shape variance to which these results can be compared. However, comparison of $\sigma^2_{t(x)}$ between different hairpins revealed some noticeable differences: after travelling a given distance x along a transition path, the variance in $t(x)$ was systematically lowest for hairpin 20R100/T4 and highest for hairpin 30R50/T4, whereas the remaining hairpins had variance between these two limits. We note that the absolute magnitudes of the observed variances are not intrinsically meaningful, because they depend systematically on the time resolution of the measurement, with many of the microscopic fluctuations being smoothed out by the instrument (102, 131). However, when comparing the variances for different measurements using the same instrument and conditions, as here, the instrumental effects are constant and hence differences in the variance reflect meaningful differences in the variability of the transition paths sampled by different molecules.

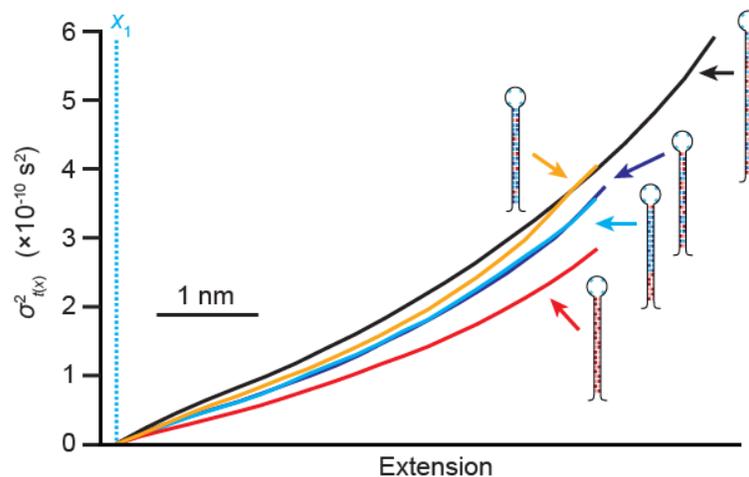


Fig 9.6: Comparison of path-shape variance for different hairpins. The variance in the average transition path shape, $\sigma^2_{t(x)}$, at each position along the reaction coordinate is similar for hairpins 20R25/T4, 20R55/T4, and 20TS06/T4, smallest for hairpin 20R100/T4, and largest for hairpin 30R50/T4. Larger variance suggests the presence of additional types of transition paths, providing evidence for multiple pathways.

9.7 Discussion of average transition path shape results

This work reveals some interesting features of transition-path shapes and how to compute them. An important practical result is that the appealingly simple definition of the average path shape in terms of $\langle t(x) \rangle$, as proposed previously (23, 24, 26), is not reliable when applied to experimental data. Although it works well for pure 1D simulations, it is quite sensitive to experimental effects like the attachment of the molecule to force probes, which induce distortions that obscure the true transition-path shape resulting from the underlying energy landscape by causing it to become more linear. This problem will likely apply to all measurement methods, not just force spectroscopy, because much of the distortion in $\langle t(x) \rangle$ appears to arise from averaging in the independent variable, t , instead of the

dependent variable x that contains the measurement noise: when we generated an ensemble of simulated transition paths simply by adding Gaussian noise to the analytical path shape defined by Eq. 9.1, we found that $\langle t(x) \rangle$ became more linear (Fig. 9.7, red), similar to the effect seen in the force spectroscopy data and simulations. In contrast, the extension-domain average $\langle x(t|\tau) \rangle$ was not affected (Fig. 9.7, black), because the noise was correctly averaged in the measured variable x and the computed shape thus remained close to the original analytical path shape (Fig. 9.7, blue).

Because $\langle t(x) \rangle$ is distorted by experimental noise, it is unreliable for reporting on the properties of the barrier, even though simulations indicate that it should be quite sensitive to the barrier shape (23). However, there is an intriguing connection between the anomalously low barrier height implied by the flatter-than-expected shape of $\langle t(x) \rangle$ and previous work finding that transition-path time distributions also implied barrier heights significantly lower than measured directly (44). Recent theoretical work has shown that distortions in the transition-path time distribution implying anomalously low barriers can be induced by memory effects in the folding dynamics (80). Such memory effects would be expected to arise naturally in force spectroscopy measurements from the compliant linker coupling the molecule to the force probe, because of the finite response time for propagation of molecular motions through the linker to the probe. Hence the distortions of $\langle t(x) \rangle$ could also be viewed, at least in part, as a consequence of linker/probe-induced memory effects.

Turning to the extension-domain average, the fact that $\langle x(t|\tau) \rangle_\tau$ matches the

functional form of the dominant transition path shape suggests that this average provides a reasonable, heuristic approximation to the dominant path. Even though there is no formal equivalence between these two quantities, it is intuitively appealing that an appropriately constructed average of the path shapes should be able to approximate the dominant path shape. Indeed, such reasoning suggests an even simpler heuristic: approximating the dominant path not as the average of all paths but only of those close to the average transition path time, which are taken in this case as representative of the average behavior of the ensemble of curves. Calculating this restricted average, $\langle x(t|\tau_{\text{TP}}) \rangle$ (Fig. 9.8, black), we found that it was indeed very similar to the more general average $\langle x(t|\tau) \rangle_{\tau}$ computed previously (Fig. 9.8, grey), and it, too, fit very well to Eq. 9.1 for the shape of the dominant transition path (Fig. 9.8, red). The values for D obtained from the fits were very similar to those in Tables 9.1 & 9.2), again being close to the results obtained by other methods but with a systematic tendency to be a bit higher. Hence $\langle x(t|\tau_{\text{TP}}) \rangle$ can be used as an approximation to the dominant path shape that is simpler to compute than $\langle x(t|\tau) \rangle_{\tau}$.

The reasonable agreement between the extension-domain averages and expectations for a 1D harmonic barrier with constant D is perhaps not surprising, given that this same approximation has worked well for describing other transition-path properties such as the average transition time, the distribution of transit times, and the average velocity at the barrier peak (14, 43, 102). However, even though the values of D implied by the fits to Eq. 9.1 are consistent within error with the values found from previous analyses of transition-path properties in the same limit,

they are systematically somewhat higher: in the range $\sim 4\text{--}8 \times 10^5 \text{ nm}^2/\text{s}$, compared to $\sim 2\text{--}5 \times 10^5 \text{ nm}^2/\text{s}$ (14, 43, 102). Analysis of the simulations indicates that most of this discrepancy likely arises from the use of $\langle x(t|\tau) \rangle_\tau$ as an imperfect approximation for the dominant path shape, since the value of D from fitting $\langle x(t|\tau) \rangle_\tau$ to Eq. 9.1 was $\sim 50\%$ higher than the actual value imposed in the simulations. Other factors involving deviations from the assumptions underlying Eq. 9.1 may also contribute to the overestimate, however, including the presence of anharmonicity in the measured barrier profiles (43, 75), the possibility of position-dependence in the diffusivity (89, 111), the presence of memory effects (80, 131), and the finite barrier size (132).

Lastly, we consider the variability in path shapes. Naively, the fact that the average $\langle x(t|\tau) \rangle_\tau$ matches the shape predicted by Eq. 9.1 for the dominant transition path suggests that the observed paths primarily involve transitions over a single, 1D barrier that is largely harmonic, *i.e.* the path variability reflects the diversity of shapes expected inherently from diffusive motion over this barrier. However, comparing the variance in average path shapes for the different hairpin sequences reveals that the situation must be more complex. Indeed, the differences in $\sigma^2_{t(x)}$ seen between the hairpins in Fig. 9.6 indicate that the diversity of transition paths occupied during the hairpin folding is sequence-dependent: the transitions in some hairpins (*e.g.* 30R50/T4) display a broader range of path shapes, whereas those in others (*e.g.* 20R100/T4) occupy a more restricted range. These differences must reflect changes in the character of the energy barrier separating the folded and unfolded states. As one possibility, if the transition paths pass through only a single

saddle-point in the landscape, then a broader saddle-point (reflecting, say, a larger transition-state ensemble) would increase the variance in path shapes. Another possibility is that increased variance could reflect the presence of multiple, parallel paths through the transition states such that there are transition paths crossing through different saddle-points in the landscape. These possibilities are depicted in Fig. 9.9. A multiplicity of microscopically distinct paths across the energy barrier is certainly expected from the statistical nature of the energy landscape picture of folding (10, 133, 134), and parallel pathways have been both observed directly in multi-state folders (135–138) and inferred from the kinetics of two-state and downhill folders (139–142), but it has not previously been possible to quantify the extent of the diversity of transition paths in a folding reaction. Regardless of the precise origins of the differences in diversity of transition-path shapes observed here, however, we note that simple 1D models of folding are not sufficient to account for such differences—despite the quantitative success of these models in explaining other properties like transition-path times (43, 44), velocities (102), and occupancies (14, 15)—underlining the notion that transition-path shapes are more sensitive to the microscopic details of the energy landscape than other such properties.

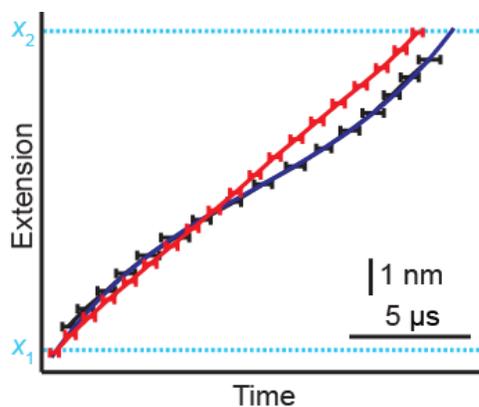


Fig 9.7: Effects of noise on time-domain and extension-domain averages. Computing the average path shapes $\langle t(x) \rangle$ (red) and $\langle x(t|\tau) \rangle_\tau$ (black) from an ensemble of simulated transition paths generated by adding noise sampled from a Gaussian distribution to each point on the dominant transition path defined by Eq.9.1 (blue), $\langle t(x) \rangle$ became more linear, as in experimental data, whereas $\langle x(t|\tau) \rangle_\tau$ remained unaffected and matched the original analytical path shape.

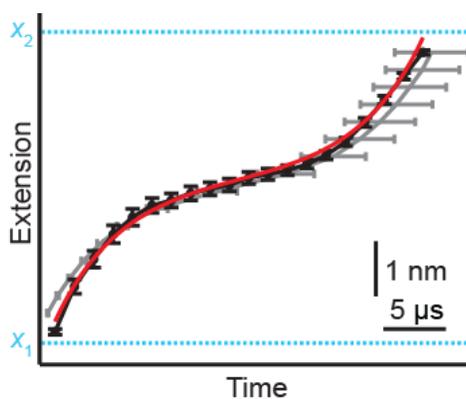


Fig 9.8: Extension-domain average path shape at τ_{TP} . The path shape averaged in the extension domain for transitions of the hairpin 30R50/T4 with duration equal to the average transition path time, $\langle x(t|\tau_{TP}) \rangle$ (black), is very close to the shape obtained from averaging over all transition times, $\langle x(t|\tau) \rangle_\tau$ (grey), and is well fit by Eq. 9.1 for the dominant path shape (red), returning a diffusion coefficient close to the result obtained from fitting $\langle x(t|\tau) \rangle_\tau$. Similar results were found for the other hairpin sequences.

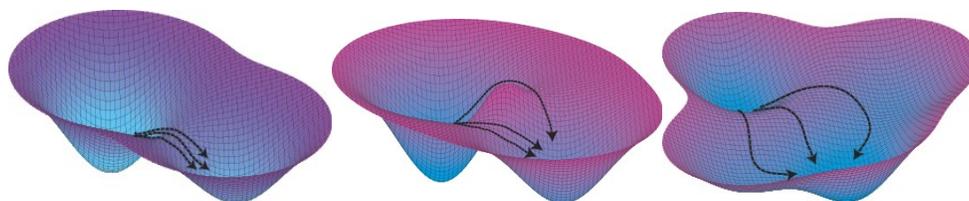


Fig 9.9: Diversity in transition-path ensembles. (Left) A single, narrow, saddle-point separates the two state basins. In such a case, the variance about the average path shape is expected to be low. (Center) The presence of more than one saddle-point can lead to a higher variance about the average path shape, reflecting the presence of multiple, parallel paths through the transition state. (Right) A single, broad, saddle-point separating the two state basins can result in a higher variance about the average path shape, reflecting a larger transition-state ensemble.

Hairpin	D ($\times 10^5$ nm ² /s)		
	unfolding	refolding	average
30R50/T4	6 ± 1	6 ± 1	6 ± 1
20R100/T4	7 ± 2	6 ± 2	6 ± 2
20R55/T4	8 ± 2	8 ± 2	8 ± 2
20TS06/T4	4 ± 1	4 ± 1	4 ± 1
20R25/T4	7 ± 1	6 ± 1	6 ± 1

Table 9.2: Diffusion coefficients from fitting extension-domain averages at τ_{TP} . Fitting the extension-domain average path shapes with transition time τ_{TP} , results in values for the diffusion coefficient which are very close to the values obtained from fitting the average over all transition times. Errors represent standard deviation.

Chapter 10

Future work and outlook

The transition-path studies discussed in this thesis have focused on simple DNA hairpins. This choice was deliberate and there are several reasons for it. First, the physical properties of the free energy landscape for the specific hairpin sequences studied have been well characterized through various methods (58, 87). Knowledge of the underlying energy landscape allows for important consistency checks when implementing newly developed analysis methods. Second, the simple zippering mechanism of DNA hairpin folding allows for relatively straightforward interpretation of results. Under applied tension, base-pairs are expected to zip/unzip sequentially and produce uniform length changes. Such hairpins likely come as close as possible to realizing an ideal 1D system in folding, and indeed the transition-path properties of hairpins studied to date have all reflected close to ideal 1D behavior. Lastly, the changes in extension when folding and the timescale over which these hairpins fold, are sufficiently long, allowing the TPs to be observed directly.

These TP studies of DNA hairpins have certainly yielded important results. They have allowed for the testing of several established theories and demonstrated the potential for TP studies to probe folding reactions at the most elementary levels. However, there remains a lot to be learned by applying these same methods to molecules with more complex structures and presumably more complex folding mechanisms. Investigations into more complicated folding processes, where TP dynamics may inform on biological function, will therefore require TP

measurements of more complex systems such as globular proteins.

However, from a more physics-based perspective, there is much that can still be learned from TP measurements of DNA hairpins. The simplicity with which DNA sequences can be designed and the accuracy of model predictions of DNA hairpin landscapes make them ideal for investigations of sequence-dependent features of TPs. The nearly ideal 1D behavior of these hairpins and the thoroughness with which their TP properties have been characterized under equilibrium conditions make them an excellent testing ground for measuring TPs out of equilibrium, where very few theoretical expectations exist.

Experimental studies of TPs are relatively young, with much of the vast information they contain about folding mechanisms still to be explored, technical capabilities still developing, and most studies prompting new questions to be answered. Theories that will complement future experiments continue to be developed, for example providing frameworks for describing different shapes of barriers (143), exploring inertial effects in TPs (144, 145), improving methods for simulating TPs (146), and developing methods for predicting TPs and their statistics (147, 148). The effects of experimental design on TP measurements are being explored to identify conditions for reliable measurements (24, 56, 57, 60). Advances in instrumentation continue to push forward experimental frontiers, such as improved AFM cantilevers enabling μs -resolution measurements of both globular (149) and membrane proteins (150). There is also much room for further integration of atomistic simulations with TP measurements, as well as comparisons of TP measurements to ultrafast kinetic measurements of downhill folding (151).

These advances provide exciting opportunities for increasingly detailed characterization of the microscopic dynamics in folding reactions.

References

1. K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, The Protein Folding Problem. *Annu. Rev. Biophys.* **37**, 289–316 (2008).
2. L. Pauling, R. B. Corey, H. R. Branson, The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci.* **37**, 205–211 (1951).
3. C. B. Anfinsen, E. Haber, M. Sela, F. H. White, The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1309–1314 (1961).
4. C. B. Anfinsen, Principles that Govern the Folding of Protein Chains. *Science* **181**, 223–230 (1973).
5. K. A. Dill, J. L. MacCallum, The Protein-Folding Problem, 50 Years On. *Science* **338**, 1042–1046 (2012).
6. D. Thirumalai, C. Hyeon, RNA and Protein Folding: Common Themes and Variations. *Biochemistry* **44**, 4957–4970 (2005).
7. J. A. Doudna, T. R. Cech, The chemical repertoire of natural ribozymes. *Nature* **418**, 222–228 (2002).
8. J. D. Watson, F. H. C. Crick, Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
9. C. Levinthal, Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45 (1968).
10. P. G. Wolynes, J. N. Onuchic, D. Thirumalai, Navigating the folding routes. *Science* **267**, 1619–1620 (1995).
11. J. Kuriyan, B. Konforti, D. Wemmer, *The Molecules of Life: Physical and Chemical Principles* (Garland Science, 2013).
12. M. T. Woodside, S. M. Block, Reconstructing Folding Energy Landscapes by Single-Molecule Force Spectroscopy. *Annu. Rev. Biophys.* **43**, 19–39 (2014).
13. R. Zwanzig, Diffusion in a rough potential. *Proc. Natl. Acad. Sci.* **85**, 2029–2030 (1988).

14. K. Neupane, A. P. Manuel, M. T. Woodside, Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nat. Phys.* **12**, 700–703 (2016).
15. K. Neupane, A. P. Manuel, J. Lambert, M. T. Woodside, Transition-Path Probability as a Test of Reaction-Coordinate Quality Reveals DNA Hairpin Folding Is a One-Dimensional Diffusive Process. *J. Phys. Chem. Lett.* **6**, 1005–1010 (2015).
16. H. Yu, *et al.*, Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *Proc. Natl. Acad. Sci.* **109**, 14452–14457 (2012).
17. H. S. Chung, Transition Path Times Measured by Single-Molecule Spectroscopy. *J. Mol. Biol.* **430**, 409–423 (2018).
18. H. A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304 (1940).
19. G. Hummer, From transition paths to transition states and rate coefficients. *J. Chem. Phys.* **120**, 516–523 (2004).
20. D. E. Makarov, *Single Molecule Science: Physical Principles and Models* (CRC Press Taylor & Francis Group, 2015).
21. N. D. Socci, J. N. Onuchic, P. G. Wolynes, Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* **104**, 5860–5868 (1996).
22. P. Faccioli, M. Sega, F. Pederiva, H. Orland, Dominant Pathways in Protein Folding. *Phys. Rev. Lett.* **97**, 108101 (2006).
23. D. E. Makarov, Shapes of dominant transition paths from single-molecule force spectroscopy. *J. Chem. Phys.* **143**, 194103 (2015).
24. P. Cossio, G. Hummer, A. Szabo, Transition paths in single-molecule force spectroscopy. *J. Chem. Phys.* **148**, 123309 (2018).
25. R. Du, V. S. Pande, A. Yu. Grosberg, T. Tanaka, E. S. Shakhnovich, On the transition coordinate for protein folding. *J. Chem. Phys.* **108**, 334–350 (1998).
26. W. K. Kim, R. R. Netz, The mean shape of transition and first-passage paths. *J. Chem. Phys.* **143**, 224108 (2015).
27. H. S. Chung, J. M. Louis, W. A. Eaton, Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci.* **106**, 11837–11844 (2009).

28. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).
29. J. D. Bryngelson, P. G. Wolynes, Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* **93**, 6902–6915 (1989).
30. J. D. Bryngelson, P. G. Wolynes, Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci.* **84**, 7524–7528 (1987).
31. A. Ashkin, Acceleration and Trapping of Particles by Radiation Pressure. *Phys. Rev. Lett.* **24**, 156–159 (1970).
32. A. Ashkin, J. M. Dziedzic, Optical Levitation by Radiation Pressure. *Appl. Phys. Lett.* **19**, 283–285 (1971).
33. A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, S. Chu, Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.* **11**, 288–290 (1986).
34. K. C. Neuman, S. M. Block, Optical trapping. *Rev. Sci. Instrum.* **75**, 2787–2809 (2004).
35. K. Neuman, E. Chadd, G. Liou, K. Bergman, S. Block, Characterization of photodamage to escherichia coli in optical traps. *Biophys. J.* **77**, 2856–2863 (1999).
36. Y. Harada, T. Asakura, Radiation forces on a dielectric sphere in the Rayleigh scattering regime. *Opt. Commun.* **124**, 529–541 (1996).
37. M. S. Rocha, Optical tweezers for undergraduates: Theoretical analysis and experiments. *Am. J. Phys.* **77**, 704–712 (2009).
38. G. Roosen, La lévitation optique de sphères. *Can. J. Phys.* (2011) <https://doi.org/10.1139/p79-175> (January 13, 2021).
39. A. Ashkin, Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophys. J.* **61**, 569–582 (1992).
40. K. Svoboda, S. M. Block, Biological Applications of Optical Forces. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 247–285 (1994).
41. A. G. Baker, C.-Y. Chuang, M. Whitmore, M. J. Comstock, Randomizing phase to remove acousto-optic device wobble errors for high-resolution optical tweezers. *Appl. Opt.* **57**, 1752–1756 (2018).

42. K. Neupane, *et al.*, Transition Path Times for Nucleic Acid Folding Determined from Energy-Landscape Analysis of Single-Molecule Trajectories. *Phys. Rev. Lett.* **109**, 068102 (2012).
43. K. Neupane, F. Wang, M. T. Woodside, Direct measurement of sequence-dependent transition path times and conformational diffusion in DNA duplex formation. *Proc. Natl. Acad. Sci.*, 201611602 (2017).
44. K. Neupane, *et al.*, Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **352**, 239–242 (2016).
45. Alan V. Oppenheim, Ronald W. Schaffer, *Discrete-Time Signal Processing*, Second Edition (Upper Saddle River, N.J. : Prentice Hall, 1999).
46. K. C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **5**, 491–505 (2008).
47. T. T. Perkins, Optical traps for single molecule biophysics: a primer. *Laser Photonics Rev.* **3**, 203–220 (2009).
48. M. T. Woodside, M. Valentine, “Single-Molecule Manipulation Using Optical Traps. In Handbook of Single-Molecule Biophysics, P. Hinterdorfer, and A. Oijen, eds.” in (Springer US, 2009), pp. 341–370.
49. W. J. Greenleaf, M. T. Woodside, E. A. Abbondanzieri, S. M. Block, Passive All-Optical Force Clamp for High-Resolution Laser Trapping. *Phys. Rev. Lett.* **95**, 208102 (2005).
50. R. Phillips, J. Kondev, J. Theriot, H. Garcia, N. Orme, *Physical Biology of the Cell* (Garland Science, 2013).
51. J. F. Marko, E. D. Siggia, Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
52. M. D. Wang, H. Yin, R. Landick, J. Gelles, S. M. Block, Stretching DNA with optical tweezers. *Biophys. J.* **72**, 1335–1346 (1997).
53. S. B. Smith, Y. Cui, C. Bustamante, Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* **271**, 795–799 (1996).
54. C. Bouchiat, *et al.*, Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements. *Biophys. J.* **76**, 409–413 (1999).
55. C. G. Baumann, *et al.*, Stretching of single collapsed DNA molecules. *Biophys. J.* **78**, 1965–1978 (2000).

56. M. T. Woodside, J. Lambert, K. S. D. Beach, Determining Intrachain Diffusion Coefficients for Biopolymer Dynamics from Single-Molecule Force Spectroscopy Measurements. *Biophys. J.* **107**, 1647–1653 (2014).
57. P. Cossio, G. Hummer, A. Szabo, On artifacts in single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14248–14253 (2015).
58. M. T. Woodside, *et al.*, Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6190–6195 (2006).
59. G.-M. Nam, D. E. Makarov, Extracting intrinsic dynamic parameters of biomolecular folding from single-molecule force spectroscopy experiments. *Protein Sci.* **25**, 123–134 (2016).
60. K. Neupane, M. T. Woodside, Quantifying Instrumental Artifacts in Folding Kinetics Measured by Single-Molecule Force Spectroscopy. *Biophys. J.* **111**, 283–286 (2016).
61. J. Gladrow, M. Ribezzi-Crivellari, F. Ritort, U. F. Keyser, Experimental evidence of symmetry breaking of transition-path times. *Nat. Commun.* **10**, 55 (2019).
62. T. Ha, Single-Molecule Fluorescence Resonance Energy Transfer. *Methods* **25**, 78–86 (2001).
63. I. V. Gopich, A. Szabo, Decoding the Pattern of Photon Colors in Single-Molecule FRET. *J. Phys. Chem. B* **113**, 10965–10973 (2009).
64. H. S. Chung, K. McHale, J. M. Louis, W. A. Eaton, Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times. *Science* **335**, 981–984 (2012).
65. H. S. Chung, W. A. Eaton, Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* **502**, 685–688 (2013).
66. K. Truex, H. Sung Chung, J. M. Louis, W. A. Eaton, Testing Landscape Theory for Biomolecular Processes with Single Molecule Fluorescence Spectroscopy. *Phys. Rev. Lett.* **115**, 018101 (2015).
67. K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, How fast-folding proteins fold. *Science* **334**, 517–520 (2011).
68. J. Kubelka, J. Hofrichter, W. A. Eaton, The protein folding ‘speed limit.’ *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004).
69. P. Hänggi, P. Talkner, M. Borkovec, Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341 (1990).

70. H. S. Chung, S. Piana-Agostinetti, D. E. Shaw, W. A. Eaton, Structural origin of slow diffusion in protein folding. *Science* **349**, 1504–1510 (2015).
71. R. Best, G. Hummer, W. A. Eaton, Native Contacts Determine Protein Folding Mechanisms in Atomistic Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **110** (2013).
72. S. Basak, *et al.*, Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed $\beta\alpha$ protein. *Proc. Natl. Acad. Sci.* **116**, 6806–6811 (2019).
73. O. K. Dudko, G. Hummer, A. Szabo, Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci.* **105**, 15755–15760 (2008).
74. H. Yu, *et al.*, Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape. *Proc. Natl. Acad. Sci.* **112**, 8308–8313 (2015).
75. M. T. Woodside, *et al.*, Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science* **314**, 1001–1004 (2006).
76. S. Chaudhury, D. E. Makarov, A harmonic transition state approximation for the duration of reactive events in complex molecular rearrangements. *J. Chem. Phys.* **133**, 034118 (2010).
77. D. E. Makarov, Reconciling transition path time and rate measurements in reactions with large entropic barriers. *J. Chem. Phys.* **146**, 071101 (2017).
78. E. Pollak, Transition Path Time Distribution, Tunneling Times, Friction, and Uncertainty. *Phys. Rev. Lett.* **118**, 070401 (2017).
79. R. Satija, A. Das, D. E. Makarov, Transition path times reveal memory effects and anomalous diffusion in the dynamics of protein folding. *J. Chem. Phys.* **147**, 152707 (2017).
80. E. Medina, R. Satija, D. E. Makarov, Transition Path Times in Non-Markovian Activated Rate Processes. *J. Phys. Chem. B* **122**, 11400–11413 (2018).
81. E. Carlon, H. Orland, T. Sakaue, C. Vanderzande, Effect of Memory and Active Forces on Transition Path Time Distributions. *J. Phys. Chem. B* **122**, 11186–11194 (2018).
82. F. Sturzenegger, *et al.*, Transition path times of coupled folding and binding reveal the formation of an encounter complex. *Nat. Commun.* **9**, 4708 (2018).

83. J.-Y. Kim, F. Meng, J. Yoo, H. S. Chung, Diffusion-limited association of disordered protein by non-native electrostatic interactions. *Nat. Commun.* **9**, 4707 (2018).
84. K. Truex, H. S. Chung, J. M. Louis, W. A. Eaton, Testing Landscape Theory for Biomolecular Processes with Single Molecule Fluorescence Spectroscopy. *Phys Rev Lett* **115**, 018101 (2015).
85. P. Hanggi, P. Talkner, M. Borkovec, Reaction-Rate Theory - 50 Years After Kramers. *Rev. Mod. Phys.* **62**, 251–341 (1990).
86. A. P. Manuel, J. Lambert, M. T. Woodside, Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories. *Proc. Natl. Acad. Sci.* **112**, 7183–7188 (2015).
87. M. T. Woodside, *et al.*, Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* **314**, 1001–1004 (2006).
88. K. Neupane, F. Wang, M. T. Woodside, Direct measurement of sequence-dependent transition path times and conformational diffusion in DNA duplex formation. *Proc. Natl. Acad. Sci.* **114**, 1329–1334 (2017).
89. R. B. Best, G. Hummer, Coordinate-dependent diffusion in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1088–1093 (2010).
90. D. A. N. Foster, *et al.*, Probing position-dependent diffusion in biomolecular folding reactions using single-molecule force spectroscopy. *Biophys. J.* **114**, 1657–1666 (2018).
91. C. Hyeon, D. Thirumalai, Multiple Probes are Required to Explore and Control the Rugged Energy Landscape of RNA Hairpins. *J. Am. Chem. Soc.* **130**, 1538–1539 (2008).
92. D. E. Makarov, Shapes of dominant transition paths from single-molecule force spectroscopy. *J. Chem. Phys.* **143**, 194103 (2015).
93. D. Chandler, Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **68**, 2959–2970 (1978).
94. A. M. Berezhkovskii, D. E. Makarov, Communication: Transition-path velocity as an experimental measure of barrier crossing dynamics. *J. Chem. Phys.* **148**, 201102 (2018).
95. I. Popa, J. M. Fernández, S. Garcia-Manyes, Direct quantification of the attempt frequency determining the mechanical unfolding of ubiquitin protein. *J. Biol. Chem.* **286**, 31072–31079 (2011).

96. D. K. Klimov, D. Thirumalai, Viscosity Dependence of the Folding Rates of Proteins. *Phys. Rev. Lett.* **79**, 317–320 (1997).
97. M. Jacob, M. Geeves, G. Holtermann, F. X. Schmid, Diffusional barrier crossing in a two-state protein folding reaction. *Nat. Struct. Mol. Biol.* **6**, 923–926 (1999).
98. T. Cellmer, E. R. Henry, J. Hofrichter, W. A. Eaton, Measuring internal friction of an ultrafast-folding protein. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18320–18325 (2008).
99. R. B. Best, G. Hummer, Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci.* **102**, 6732–6737 (2005).
100. A. N. Gupta, *et al.*, Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat. Phys.* **7**, 631–634 (2011).
101. B. W. Zhang, D. Jasnow, D. M. Zuckerman, Transition-event durations in one-dimensional activated processes. *J. Chem. Phys.* **126**, 074504 (2007).
102. K. Neupane, N. Q. Hoffer, M. T. Woodside, Measuring the Local Velocity along Transition Paths during the Folding of Single Biological Molecules. *Phys. Rev. Lett.* **121**, 018102 (2018).
103. H. Gelman, M. Gruebele, Fast protein folding kinetics. *Q. Rev. Biophys.* **47**, 95–142 (2014).
104. R. Du, V. S. Pande, A. Yu. Grosberg, T. Tanaka, E. S. Shakhnovich, On the transition coordinate for protein folding. *J. Chem. Phys.* **108**, 334–350 (1998).
105. J. Lu, J. Nolen, Reactive trajectories and the transition path process. *Probab. Theory Relat. Fields* **161**, 195–244 (2015).
106. D. J. Bicout, A. Szabo, Electron transfer reaction dynamics in non-Debye solvents. *J. Chem. Phys.* **109**, 2325–2338 (1998).
107. J. E. Shea, C. L. Brooks, From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52**, 499–535 (2001).
108. E. R. Henry, W. A. Eaton, Combinatorial modeling of protein folding kinetics: free energy profiles and rates. *Chem. Phys.* **307**, 163–185 (2004).
109. S. Piana, K. Lindorff-Larsen, D. E. Shaw, Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci.* **109**, 17845–17850 (2012).

110. J. Chahine, R. J. Oliveira, V. B. P. Leite, J. Wang, Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. *Proc. Natl. Acad. Sci.* **104**, 14646–14651 (2007).
111. D. A. N. Foster, *et al.*, Probing Position-Dependent Diffusion in Folding Reactions Using Single-Molecule Force Spectroscopy. *Biophys. J.* **114**, 1657–1666 (2018).
112. A. N. Naganathan, J. M. Sanchez-Ruiz, V. Muñoz, Direct Measurement of Barrier Heights in Protein Folding. *J. Am. Chem. Soc.* **127**, 17970–17971 (2005).
113. D. De Sancho, U. Doshi, V. Muñoz, Protein Folding Rates and Stability: How Much Is There Beyond Size? *J. Am. Chem. Soc.* **131**, 2074–2075 (2009).
114. D. Thirumalai, From Minimal Models to Real Proteins: Time Scales for Protein Folding Kinetics. *J. Phys. I* **5**, 1457–1467 (1995).
115. V. Muñoz, M. Cerminara, When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. *Biochem. J.* **473**, 2545–2559 (2016).
116. B. C. Carter, M. Vershinin, S. P. Gross, A Comparison of Step-Detection Methods: How Well Can You Do? *Biophys. J.* **94**, 306–319 (2008).
117. S. J. Hagen, Solvent Viscosity and Friction in Protein Folding Dynamics. *Curr. Protein Pept. Sci.* **11**, 385–395 (2010).
118. K. Neupane, *et al.*, Transition Path Times for Nucleic Acid Folding Determined from Energy-Landscape Analysis of Single-Molecule Trajectories. *Phys. Rev. Lett.* **109**, 068102 (2012).
119. K. Neupane, N. Q. Hoffer, M. T. Woodside, Testing Kinetic Identities Involving Transition-Path Properties Using Single-Molecule Folding Trajectories. *J. Phys. Chem. B* **122**, 11095–11099 (2018).
120. K. Neupane, *et al.*, Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **352**, 239–242 (2016).
121. N. Q. Hoffer, K. Neupane, A. G. T. Pyo, M. T. Woodside, Measuring the average shape of transition paths during the folding of a single biological molecule. *Proc. Natl. Acad. Sci.*, 201816602 (2019).
122. M. T. Woodside, *et al.*, Nanomechanical Measurements of the Sequence-Dependent Folding Landscapes of Single Nucleic Acid Hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6190–5 (2006).

123. D. Thirumalai, From Minimal Models to Real Proteins: Time Scales for Protein Folding Kinetics. *J. Phys. I* **5**, 1457–1467 (1995).
124. B. A. Shoemaker, J. Wang, P. G. Wolynes, Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. Edited by A. R. Fersht. *J. Mol. Biol.* **287**, 675–694 (1999).
125. J. Shimada, E. I. Shakhnovich, The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11175–11180 (2002).
126. L. L. Chavez, S. Gosavi, P. A. Jennings, J. N. Onuchic, Multiple routes lead to the native state in the energy landscape of the β -trefoil family. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10254–10258 (2006).
127. J. Juraszek, P. G. Bolhuis, Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci.* **103**, 15859–15864 (2006).
128. F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, T. R. Weikl, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci.* **106**, 19011–19016 (2009).
129. D. L. Ensign, V. S. Pande, The Fip35 WW Domain Folds with Structural and Mechanistic Heterogeneity in Molecular Dynamics Simulations. *Biophys. J.* **96**, L53–L55 (2009).
130. E. R. Henry, R. B. Best, W. A. Eaton, Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **110**, 17880–17885 (2013).
131. A. M. Berezhkovskii, D. E. Makarov, Single-Molecule Test for Markovianity of the Dynamics along a Reaction Coordinate. *J. Phys. Chem. Lett.* **9**, 2190–2195 (2018).
132. A. G. T. Pyo, N. Q. Hoffer, K. Neupane, M. T. Woodside, Transition-path properties for folding reactions in the limit of small barriers. *J. Chem. Phys.* **149**, 115101 (2018).
133. K. A. Dill, H. S. Chan, From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* **4**, 10–19 (1997).
134. W. A. Eaton, P. G. Wolynes, Theory, simulations, and experiments show that proteins fold by multiple pathways. *Proc. Natl. Acad. Sci.*, 201716444 (2017).
135. B. Onoa, *et al.*, Identifying Kinetic Barriers to Mechanical Unfolding of the *T. thermophila* Ribozyme. *Science* **299**, 1892–1895 (2003).

136. J. B. Udgaonkar, Multiple Routes and Structural Heterogeneity in Protein Folding. *Annu. Rev. Biophys.* **37**, 489–510 (2008).
137. S. S. Mojumdar, *et al.*, Partially native intermediates mediate misfolding of SOD1 in single-molecule folding trajectories. *Nat. Commun.* **8**, 1881 (2017).
138. J. Roca, *et al.*, Monovalent ions modulate the flux through multiple folding pathways of an RNA pseudoknot. *Proc. Natl. Acad. Sci.*, 201717582 (2018).
139. S. E. Radford, C. M. Dobson, P. A. Evans, The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302–307 (1992).
140. J. Sabelko, J. Ervin, M. Gruebele, Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6031–6036 (1999).
141. E. J. Guinn, B. Jagannathan, S. Marqusee, Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat. Commun.* **6**, 6861 (2015).
142. V. Muñoz, M. Cerminara, When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. *Biochem. J.* **473**, 2545–2559 (2016).
143. E. Pollak, Stochastic Transition State Theory. *J. Phys. Chem. Lett.* **9**, 6066–6071 (2018).
144. J. O. Daldrop, W. K. Kim, R. R. Netz, Transition paths are hot. *EPL Europhys. Lett.* **113**, 18004 (2016).
145. M. Laleman, E. Carlon, H. Orland, Transition path time distributions. *J. Chem. Phys.* **147**, 214103 (2017).
146. H. Jung, K. Okazaki, G. Hummer, Transition path sampling of rare events by shooting from the top. *J. Chem. Phys.* **147**, 152716 (2017).
147. W. M. Jacobs, E. I. Shakhnovich, Structure-Based Prediction of Protein-Folding Transition Paths. *Biophys. J.* **111**, 925–936 (2016).
148. W. M. Jacobs, E. I. Shakhnovich, Accurate Protein-Folding Transition-Path Statistics from a Simple Free-Energy Landscape. *J. Phys. Chem. B* **122**, 11126–11136 (2018).
149. D. T. Edwards, J. K. Faulk, M.-A. LeBlanc, T. T. Perkins, Force Spectroscopy with 9- μ s Resolution and Sub-pN Stability by Tailoring AFM Cantilever Geometry. *Biophys. J.* **113**, 2595–2600 (2017).

150. H. Yu, M. G. W. Siewny, D. T. Edwards, A. W. Sanders, T. T. Perkins, Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science* **355**, 945–950 (2017).
151. M. Szczepaniak, *et al.*, Ultrafast folding kinetics of WW domains reveal how the amino acid sequence determines the speed limit to protein folding. *Proc. Natl. Acad. Sci.*, 201900203 (2019).
152. C. H. Reinsch, Smoothing by spline functions. *Numer. Math.* **10**, 177–183 (1967).

Appendices

Appendix A DNA hairpin sequence

Name	Sequence
30R50/4T	gagtcaacgtactgatcacgctggatccta(t) ₄ taggatccagcgtgatcagt acgttgactc
20R0/4T	tattatatattaatataata(t) ₄ ttatatattaatataata
20R25/4T	aagttaacatctagattcta(t) ₄ tagaatctagatgtaactt
20R55/4T	gagtcaacgtctggatcctg(t) ₄ caggatccagacgttgactc
20R100/4T	cgccgcgggccggcgcgcg(t) ₄ ccgcgcgccggcccgcggcg
20TS06/T4	gccggctattattatattc(t) ₄ gaatataaataatagccggc

Table A.1: DNA hairpin sequences used in this thesis

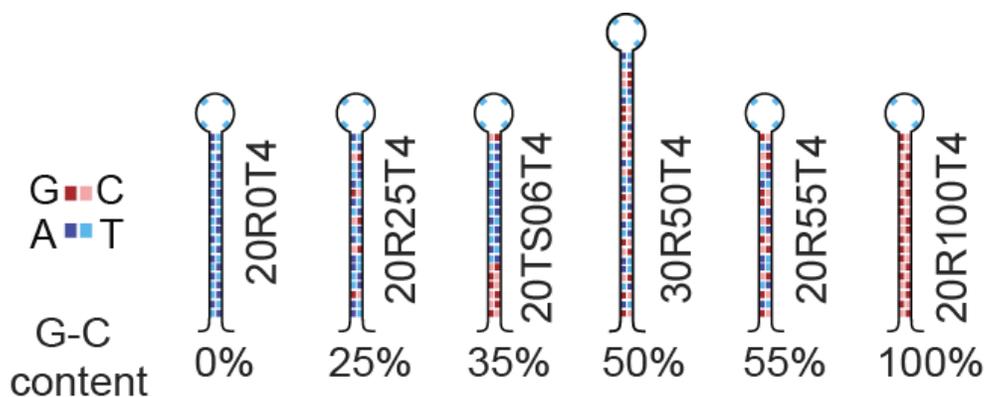


Fig A.1: DNA hairpin legend.

Appendix B Sample preparation

DNA hairpins connected to double-stranded (ds) DNA handles were prepared

as follows, an oligomer containing the hairpin sequence separated by abasic sites from a 5' ligation overhang and a 3' priming sequence was used to create a kilobase-long dsDNA handle with the hairpin on one end via autosticky PCR. The ligation overhang was used to ligate the PCR product to another kilobase dsDNA handle with a complementary overhang. The resultant construct was attached to 600- and 820-nm polystyrene beads via biotin/avidin and dioxigenin/anti-digoxigenin pairs to generate “dumbbells” for trapping. Hairpin constructs samples were incubated at ~100 pM with 250 pM polystyrene beads to form dumbbells. Dumbbells were diluted to ~500 fM in 50 mM MOPS, pH 7.0, with 200 mM KCl and oxygen scavenging system (8 mU/ μ L glucose oxidase, 20 mU/ μ L catalase, 0.01% w/v D-glucose), before insertion into a sample cell for the optical trap.

Appendix C Smoothing of transition-path trajectories

TP velocities can be calculated by differentiating the measured TP trajectory. However, numerical differentiation tends to amplify the effects of noise present in the TP trajectory and therefore straightforward differentiation of raw data is ill-advised. SMFS measurements of TPs are inherently noisy due to the coupling of the molecule to force probes (beads) that themselves experience Brownian motion during the measurement. Luckily, this noise will be normally distributed and can, therefore, in principle, be reduced by applying smoothing algorithms to the data.

Smoothing is a delicate procedure, often being described as more of an art than a science and requires the use of internal consistency checks as well as visual inspection in order to avoid distortion of the data. Most commercial data analysis software packages support the use of several smoothing algorithms, including

smoothing splines, boxcar filtering, and Savitsky-Golay filtering, each having its own set of smoothing parameters. This inconsistency in smoothing parameters can make comparing the effects of one algorithm to another difficult and necessitates the need for careful visual inspection of the smoothed data.

The degree to which the data should be smoothed is another point that warrants careful consideration. The use of internal consistency checks should be used whenever possible. For example, in the case of calculating the velocity along TPs, one can make use of the knowledge of the distance between states and the average transition-path time, to calculate the expected average velocity $\langle v \rangle = \Delta x / \tau_{tp}$

. Comparing this value to that obtained by calculating the average velocity from the smoothed, differentiated trajectories, can aid in avoiding over smoothing the data. Furthermore, because a set of TPs from any particular molecule will display a wide range of transition-path times, but all travel the same distance, the TPs will display varying amounts of curvature. For this reason, it is often best practice to implement a smoothing procedure that varies the amount of smoothing from one TP to another, based on the individual transition-path time.

For the velocity measurements presented in this thesis, we chose to smooth the extension trajectories using a smoothing spline interpolation (152). This approach finds the spline $g(t)$ that minimizes $\int g'' dt$ along the transition path,

subject to the condition $\sum_{i=0}^n \left(\frac{g(t_i) - x_i}{\sigma} \right)^2 \leq S$, where x_i is the extension at a given

point i , $g(t_i)$ is the value of the spline at that point, σ is the average standard deviation of the extension arising from the thermal fluctuations, and S is the

smoothing factor. The standard deviation was found by calculating the standard deviation in extension values while the molecule was fluctuating in the folded and unfolded states and then taking the average of the two values. When the standard deviation is properly characterized, optimal smoothing (152) is said to occur with a smoothing factor $S=1$. However, upon visual inspection and comparison of the obtained average velocity to the expected value, it was decided that the smoothing factor should be reduced and that an adaptive smoothing procedure should be implemented. The final smoothing procedure used worked as follows: First, the unsmoothed transition-path time for each TP was calculated, then the smoothing factor for each TP was chosen based on the transition-path time and varied linearly from $S=0.3$ to $S=0.7$, where all TPs with transition-path times lower than $10 \mu\text{s}$ were smoothed with $S=0.3$ and all those with transition-path times higher than $60 \mu\text{s}$ were smoothed with $S=0.7$.

In short, there is no straightforward way to determine the correct method or degree of smoothing. Therefore, comparisons of the velocity distributions and velocity profiles calculated via several different smoothing procedures should be done in order to assess the robustness of the results (Fig C1), and ultimately choices on smoothing should be based on an intimate knowledge of the data and the nature of the measurements.

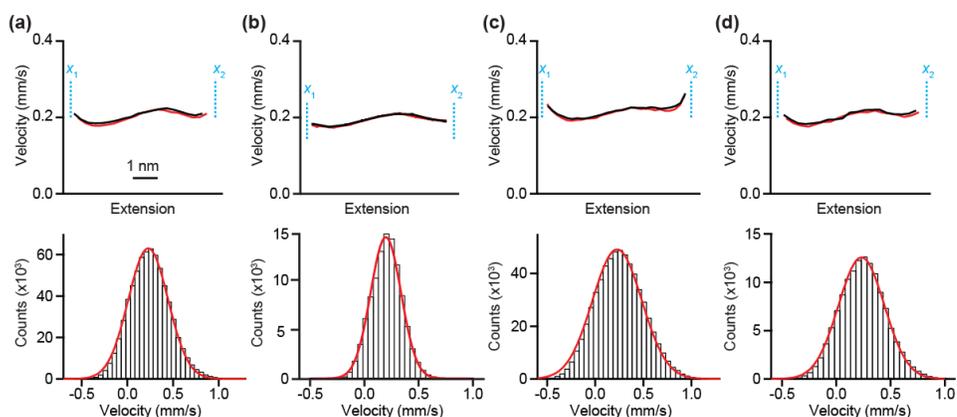


Figure C1: Effects of smoothing on velocity profiles. The velocity profiles (upper panels) were similar for different smoothing protocols: (a, b) smoothing spline with respectively $S = 0.7$ and $S = 1$, (c) Savitsky-Golay filter, (d) boxcar filter. Velocity distributions (lower panels) were qualitatively similar for similar amounts of smoothing (a, c, d).

Appendix D Boundary effects in transition path measurements

When measuring TPs the finite sampling rate of the measurements can make defining boundary regions difficult, as it is highly unlikely that data points will exist at the exact boundary locations. Due to the nature of TP measurements as continuous-time trajectories, with most of the data points corresponding to fluctuations within the stable states rather than TPs, hard boundary locations must be implemented in order to separate the TPs from the non-productive fluctuations. Moreover, the stochastic nature of TPs and the coupling of the molecule to a force probe (bead), that itself undergoes Brownian motion, leads to an appreciable uncertainty with respect to when the molecule truly crosses the boundaries. Additionally, certain analysis procedures such as smoothing or numerical differentiation carry their own boundary effects that need to be carefully considered.

There is no universal method for dealing with boundary conditions that works in every situation. Prior to performing any analysis on the TPs the possible ways that the boundary effects may manifest themselves should be carefully considered. Some examples include smoothing, numerical differentiation, occupancy statistics, transition-path times, and average path shapes.

The sigmoidal nature of TPs means that any smoothing procedure should include several points on either side of the transition region. This is because smoothing procedures generally attempt to minimize the curvature over some window of time in the trajectory. TPs typically reside in one of the stable states for relatively long periods of time before suddenly undergoing a transition event. Therefore, if data points within the stable states are not included in the smoothing, the initial departure into the transition region will appear almost instantaneous while any fast movements within the transition region will be disproportionately slowed by the smoothing. Including several points on either side of the transition region can help to reduce the effects of these sudden departures from the stable states.

Calculating transition-path times explicitly involves measuring the times where the boundary locations are crossed. This makes them highly susceptible to the boundary effects and tends to lead to underestimates of the transition-path time. In this case, the effects can be dealt with by applying a light median filtering procedure to the data. This effectively finds the average value for the points nearest to the boundary-crossing and provides a more accurate estimate of the actual crossing time than would be found with a simple interpolation.

The occupancy statistics for TPs can also be distorted near the boundary locations. This is because the boundaries are acting as absorbing boundaries and any forward bead fluctuation that might produce a data point beyond the boundary will immediately halt the TP. Meanwhile, bead fluctuations in the reverse direction will not halt the TP, leading to artificially low occupancy of the regions closest to the boundaries. This can be dealt with by including several additional data points, on both sides of the trajectory, beyond the transition region. This allows rapid fluctuations near the boundaries to be accounted for in the occupancy statistics.

Appendix E Detection system upgrades

After installing the new QPD detectors it became apparent that further upgrades to the system would be required. Power spectrum measurements revealed that the QPDs were much more sensitive to electronic noise present in the detection system, so much so that the experimental signal had been swamped out. There are two ways to deal with such an issue: increase the experimental signal by increasing the amount of laser light incident on the detector, and decrease the amount of electronic background noise present in the system. The amount of laser light reaching the detectors is primarily limited by two factors: the power of the laser itself and the efficiency of the fiber coupling.

In my first attempt to increase the signal-to-noise ratio, I replaced the 7 mW HeNe laser with a 45 mW HeNe laser made by the same company. After aligning and coupling the new laser to the fiber, it was immediately clear that the new laser was not suitable for the instrument. The laser's pointing stability was so poor that when coupled to a fiber it produced fluctuations in the power so large that they

could not be dealt with by normalizing the signal. The signal power fluctuated between values that were too high, creating a second optical trap, and values that were too low, allowing noise to overcome the signal. The new laser was then removed and the original 7 mW laser was reinstalled and the decision was made to purchase a high-quality fiber coupler in order to improve the coupling efficiency.

One critical component of a fiber coupling system is the objective lens that focuses the incoming laser light onto the end of the optical fiber. Choosing the correct focal length for this objective lens is crucial when attempting to achieve a high-efficiency coupling. The correct focal length can be determined by $f = \pi D(\text{MFD})/4\lambda$, where D is the $1/e^2$ diameter of the collimated beam incident on the lens, λ is the wavelength of input light, MFD is the mode field diameter of the optical fiber and, f is the focal length of the objective lens. Applying this equation to our system yields a focal length of 7.2 mm. For optimum coupling, the spot size of the focused beam must be less than the MFD of the single-mode fiber. As a result, if a lens or objective is not available that provides an exact match, then an optic with a focal length that is shorter than what the calculation yields, should be used. With this in mind, a Newport M-40X microscope objective lens with a focal length of 4.5 mm was chosen for the coupling. This lens, combined with a ThorLabs MBT616D three-axis stage, allowed for a coupling efficiency of >60%, with minimal power fluctuations, sufficient to overcome the noise floor of the QPD detectors.