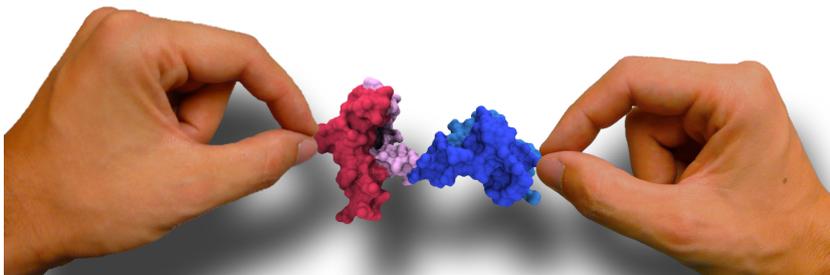




JOHANNES STIGLER

COMPLEX LIGAND-DEPENDENT FOLDING OF
SINGLE PROTEINS OBSERVED WITH OPTICAL
TWEEZERS

DISSERTATION AT THE PHYSICS DEPARTMENT E22
TECHNISCHE UNIVERSITÄT MÜNCHEN



TECHNISCHE UNIVERSITÄT MÜNCHEN
LEHRSTUHL FÜR BIOPHYSIK E22

COMPLEX LIGAND-DEPENDENT FOLDING OF SINGLE PROTEINS
OBSERVED WITH OPTICAL TWEEZERS

JOHANNES STIGLER

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur
Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Martin Zacharias

Prüfer der Dissertation: 1. Univ.-Prof. Dr. Matthias Rief
2. Univ.-Prof. Dr. Thorsten Hugel

Die Dissertation wurde am 16.10.2012 bei der Technischen Universität München eingereicht und
durch die Fakultät für Physik am 26.11.2012 angenommen.

Johannes Stigler: *Complex ligand-dependent folding of single proteins observed with optical tweezers*, © October 2012

ABSTRACT

Calcium sensing and the control of calcium-related pathways play important roles in the control of cellular processes. A key player involved in these processes is calmodulin, a small, two-domain, 148 amino acid protein that undergoes structural changes upon calcium binding. Moreover, calmodulin is able to bind target peptides in a calcium dependent manner.

This work investigated the folding/unfolding properties of single calmodulin molecules using high-resolution optical tweezers. To this end, calmodulin molecules were fused, using handles made of DNA, to glass beads that could be trapped in strongly focused laser beams. This so-called dumbbell geometry allowed the mechanical manipulation of single molecules of calmodulin.

It was found that, at high calcium, calmodulin populates a series of on- and off-pathway intermediates. The off-pathway intermediates act as kinetic traps that make calmodulin, at high calcium, an effectively slow-folding protein. Arguably, already small multi-domain proteins fold in complex networks of numerous intermediates.

To assess the biological relevance of the findings, we also studied calmodulin's folding/unfolding properties close to physiological calcium concentrations. Here, the kinetic traps found at high calcium were no longer populated. Using a folding model under ligand-binding conditions, we could reconcile the measured data and develop a model for the calcium-dependent folding and unfolding of calmodulin.

This work further presents how advanced statistical methods can help to obtain detailed insight into the folding process of complex systems from one-dimensional force-time-trajectories.

ZUSAMMENFASSUNG

Die Detektion von Calcium und die Reaktion auf Konzentrationsänderungen spielen eine wichtige Rolle in der Regelung zellulärer Prozesse. Ein wichtiger Bestandteil dieser Signalkaskade ist Calmodulin. Calmodulin ist ein zwei-Domänen Protein, bestehend aus 148 Aminosäuren, das seine Konformation Calcium-abhängig ändert. In seiner Calcium-gebundenen Konformation kann Calmodulin zelluläre Prozesse steuern indem es an Peptide seiner Interaktionspartner bindet.

In dieser Arbeit wurde die Faltung und Entfaltung einzelner Calmodulinmoleküle mit einer optischen Pinzette untersucht. Einzelne Proteine

wurden hierzu mittels DNS Strängen an mikroskopische Glaskugeln gebunden, die ihrerseits in stark fokussierten Laserstrahlen gefangen werden konnten. Diese sogenannte Hantel-Geometrie ermöglichte die Untersuchung von Calmodulin durch mechanische Manipulation.

Es konnte gezeigt werden, dass bei hohen Calcium Konzentrationen mehrere Intermediate populiert werden, die entweder auf dem oder abseits des direkten Faltungspfades liegen. Letztere sorgen dafür, dass Calmodulin bei hohen Calcium Konzentrationen oft in nicht-native Konformationen faltet und deshalb effektiv als langsam-faltendes Protein klassifiziert werden muss. Offenbar zeigen bereits kleine Mehrdomänenproteine komplexe Netzwerke von Intermediaten.

Weiterhin wurde die Faltung bei niedrigeren, physiologischen Calcium Konzentrationen untersucht. In diesem Fall konnte gezeigt werden, dass die nicht-nativen Konformationen, die die Faltung bei hohen Calcium Konzentrationen dominiert hatten, nicht mehr populiert wurden. Mit Hilfe eines Modells für die Faltung Liganden-bindender Proteine konnte ein generelles Modell für die Calcium Konzentrations-abhängige Faltung Calmodulins entwickelt werden.

Zusätzlich werden statistische Methoden präsentiert, mit deren Hilfe detaillierte Informationen über die Faltung komplexer Systeme aus eindimensionalen Kraft-Zeit-Trajektorien gewonnen werden können.

PUBLICATIONS

Some ideas, tables and figures have appeared previously in the following publications:

- STIGLER J, RIEF M (2012) Calcium-dependent folding of single calmodulin molecules. *PNAS* 109:17814–17819.
- STIGLER J, RIEF M (2012) Hidden markov analysis of trajectories in single-molecule experiments and the effects of missed events. *ChemPhysChem* 13:1079–1086.
- STIGLER J, ZIEGLER F, GIESEKE A, GEBHARDT JCM, RIEF M (2011) The complex folding network of single calmodulin molecules. *Science* 334:512–516.

ACKNOWLEDGMENTS

A number of people helped, directly or indirectly, to put together the work that is presented here.

First of all, I want to thank *Matthias Rief*, my supervisor, for all the long helpful discussions, many great ideas, the encouragement and fun during my time at E22. One could not wish for a more relaxed, and at the same time, efficient, lab atmosphere.

I also am very thankful to those who, in the beginning, and later, had to endure my naive questions about biochemistry, mainly *Anja Gieseke*, *Daniela Scheickl*, *Gabriel Zoldak* and *Daniela Bauer*.

Also everybody who got me started in the lab, especially *Thomas Bornschlögl* with the AFM as well as *Christof Gebhardt* and *Melanie Brunnbauer* with the optical tweezers deserve my acknowledgements.

Whenever I was lost in the depths between physics, biology and chemistry, a number of people always helped me out with long and fruitful discussions: *Lorenz Rognoni*, *Fabian Ziegler*, *Benjamin Pelz*, *Alex Mehlich*, *Uli Merkel* and *Ziad Ganim*.

For all of their technical help, I want to thank *Uli Kleßinger*, *Rudi Lehrhuber*, *Monika Rusp*, *Gabi Chmel* and *Karin Vogt*.

While fighting bureaucracy, *Elke Fehlsenfeld* and *Nicole Mittermüller* were always of great help.

Finally, I want to thank everybody in the bio corner of the physics building and all the members of the Rief, Bausch, Hugel and Woehlke labs for the great fun and my family for their support.

CONTENTS

I	INTRODUCTION AND THEORY	1
1	INTRODUCTION	3
2	CALMODULIN AND EF HAND PROTEINS	5
2.1	Introduction: EF hand proteins, ligand binding	5
2.1.1	The EF hand principle	5
2.1.2	Calmodulin	6
3	OPTICAL TWEEZERS	9
3.1	The principle of optical trapping	9
3.2	Optical tweezers in force spectroscopy	11
3.3	The optical tweezers instrument used in this work	12
3.4	Force application in a dumbbell assay	12
3.4.1	The problem of oxygen damage	14
4	ANALYSIS OF SINGLE MOLECULE TRAJECTORIES	17
4.1	Hidden Markov Models	17
4.1.1	Introduction	17
4.1.2	Data preparation	18
4.1.3	Initialization of the model parameters	19
4.1.4	Classification of data points – The forward-backward algorithm	19
4.1.5	Optimization of model parameters	20
4.1.6	The Viterbi algorithm	21
4.1.7	Numerical issues	22
4.1.8	Performance of the algorithms	22
4.2	Transition rates and the issue of missed events	23
4.2.1	Obtaining transition rate constants	23
4.2.2	The problem of missed events	25
4.3	A test for memory in a trajectory	28
4.4	The dumbbell configuration in thermodynamic equilibrium	29
4.4.1	Polymer models	30
4.5	Equilibrium free energies	31
4.6	Models for the force dependence of transition rate constants	32
4.6.1	The Bell Model	32
4.6.2	A model for folding	33
4.6.3	Folding and unfolding rate models and the applicability to equilibrium data	34
4.7	Non-equilibrium methods	35
4.8	Summary	35
5	FOLDING OF LIGAND BINDING PROTEINS	37
5.1	Microscopic and macroscopic binding	37
5.2	Cooperativity	38

5.3	The influence of ligand binding on the folding and unfolding	38
5.3.1	Example: Force-dependent folding and unfolding rates for a domain with two calcium binding sites	40
5.4	Stabilization by ligand binding	42
II	RESULTS	43
6	THE COMPLEX FOLDING OF CALMODULIN	45
6.1	A complex network of states	45
6.2	State identification: truncation mutants	47
6.3	Further evidence from a crosslinked mutant	48
6.4	Shifting the equilibrium	50
6.4.1	State occupancies	50
6.4.2	Dwell times and rate constants	51
6.5	The full network, allowed transitions	53
6.6	On the off-pathway nature of F ₂₃ and F ₁₂₃	55
6.7	The effect of kinetic traps on the folding time	57
6.8	How good is the assignment?	58
6.8.1	Brownian dynamics simulations	59
6.9	Controls: Interference of the oxygen scavenger?	61
6.10	Discussion	61
6.10.1	Anticooperativity between the domains	61
6.10.2	Folding rate constants	62
6.10.3	Kinetic traps	62
6.10.4	Are the transitions independent?	63
6.10.5	Structural information	65
6.10.6	Influence of the ubiquitin handles	65
6.11	Summary	68
7	CALCIUM-DEPENDENT FOLDING OF CALMODULIN	71
7.1	Full-length calmodulin at lower calcium concentrations	71
7.2	Folding/unfolding of the N-terminal domain at low calcium	73
7.3	Folding/unfolding of the C-terminal domain at low calcium	74
7.4	Stoichiometry of binding to the native state	75
7.5	Stoichiometry of binding to the transition state	77
7.6	Transient populations of the apo state	78
7.6.1	When can the transient apo-states be observed?	80
7.7	Discussion	81
7.7.1	The folding intermediates at low calcium	81
7.7.2	Accuracy of the determined values	82
7.7.3	Conformational changes	84
7.7.4	Apparent ligand-dependent shift of the transition state barrier position	85
7.7.5	Influence of the ubiquitin handles	87
7.7.6	Open questions	88

7.8	Summary	89
8	OUTLOOK	91
III	APPENDIX	93
A	RESOLVING NUMERICAL ISSUES FOR HIDDEN MARKOV MODELS	95
B	NON-EQUILIBRIUM METHODS	97
B.1	Direct measurement of rate constants in pulling cycles	97
B.2	Fluctuation theorems	97
B.2.1	The Crooks fluctuation theorem	98
B.2.2	Jarzynski's equality	99
C	SIMULATIONS	101
C.1	Brownian dynamics simulations	101
C.2	Folding time simulations	102
C.3	Simulation of folding and unfolding of ligand-binding proteins	102
D	MATERIALS AND METHODS	105
D.1	Construct production	105
D.2	Measurement procedure	106
D.3	Sequences	109
	BIBLIOGRAPHY	111

LIST OF FIGURES

Figure 1	Crystal structure of calmodulin	6
Figure 2	Simple ray optics picture of optical tweezers	9
Figure 3	Potential of an optical trap and resulting force	10
Figure 4	Single beam and dual beam	11
Figure 5	Schematic of a dual beam setup	12
Figure 6	Schematic for a dumbbell configuration	13
Figure 7	A toy example for a Hidden Markov model	18
Figure 8	Discretization of observation data.	19
Figure 9	Assignment performance of a Hidden Markov model on simulated data	23
Figure 10	Assignment probabilities for a Hidden Markov model	24
Figure 11	Effect of temporal detection limitations on an idealized three-state network	25
Figure 12	Augmented network of a three-state system with temporal cutoff	26
Figure 13	Test for non-stochastic long-range behavior on simulated data	30
Figure 14	The effect of an external force on an energy landscape	33
Figure 15	Microscopic model for two binding sites	37
Figure 16	Model for the folding of a protein with two identical binding sites	39
Figure 17	Theoretical folding and unfolding rates of a ligand binding protein	41
Figure 18	Setup and Stretch-and-Relax cycles for WT calmodulin	46
Figure 19	Equilibrium fluctuations of full-length calmodulin	47
Figure 20	Comparison of traces of full-length calmodulin with truncation mutants	48
Figure 21	Trace comparison CaM 128/144 and wildtype	49
Figure 22	Shifting the equilibrium in full-length calmodulin	50
Figure 23	Force-dependent transition rate constants of full-length calmodulin	52
Figure 24	F_{123} and F_{1234} do not exchange	54
Figure 25	Energy landscape of calmodulin	54
Figure 26	Averaged transition signal from the unfolded state	55
Figure 27	Off-pathway nature of F_{123}	56
Figure 28	Distribution of calculated folding times for full-length calmodulin at high calcium	57

Figure 29	The effect of kinetic traps on the refolding	58
Figure 30	Histogram of all states in full-length calmodulin	59
Figure 31	F_{23} and F_{12} are distinguishable by lifetime at higher forces	60
Figure 32	Results of a five-state model	61
Figure 33	Brownian dynamics simulation of the folding of full-length calmodulin	62
Figure 34	Calculated folding times at various loads	63
Figure 35	The transitions are first-order Markovian	64
Figure 36	Influence of an unfolded ubiquitin on the kinetics of full-length calmodulin	66
Figure 37	Kinetics of a full-length calmodulin construct without ubiquitins	66
Figure 38	Full-length calmodulin at different calcium concentrations	71
Figure 39	The N-terminal domain of calmodulin at different calcium concentrations	73
Figure 40	The C-terminal domain of calmodulin at different calcium concentrations	75
Figure 41	Kinetic and equilibrium parameters of the single domains of calmodulin at different calcium concentrations	76
Figure 42	Calcium-dependent stoichiometry of the isolated domains of calmodulin	78
Figure 43	Cumulative lifetime histograms of the single domain constructs at intermediate calcium concentrations	79
Figure 44	The transitions into the C-domain's apo states are purely stochastic	80
Figure 45	Calcium-dependent free energies of folding for the state F_{23}	82
Figure 46	Equilibrium free energies calculated from zero-force extrapolated rates	83
Figure 47	Calcium dependence of the total length and barrier positions	84
Figure 48	Primary folding/unfolding pathways of CaM_{34} at different calcium concentrations	85
Figure 49	Simulations explain the apparent shift of the transition state position	86
Figure 50	Influence of an unfolded ubiquitin on the kinetics of the isolated domains	88
Figure 51	“Dead” states in the isolated N-terminal domain	89
Figure 52	The Crooks fluctuation theorem	98
Figure 53	Representation of the transition rate matrix for a domain with two binding sites	103
Figure 54	Pulling protocol of a typical measurement	106

Figure 55	Experimental calibration of the bead approach	107
Figure 56	Background signal from the AOD	109
Figure 57	Sequences of the calmodulin constructs	110

LIST OF TABLES

Table 1	Missed-event correction for a two-state system	28
Table 2	Missed-event correction for a three-state system	28
Table 3	Contour length increases for CaM 128/144	49
Table 4	Equilibrium free energies of full-length calmodulin at 10 mM Ca ²⁺	51
Table 5	Transition state positions and folding/unfolding rate constants of full-length calmodulin at 10 mM Ca ²⁺	53
Table 6	Equilibrium free energies of full-length calmodulin molecule with folded, unfolded and no ubiquitins.	67
Table 7	Equilibrium free energies of full-length calmodulin at different calcium concentrations	72
Table 8	Apo energies and calcium affinities of the isolated domains of calmodulin	77
Table 9	Length differences depending on the folding state of the ubiquitin handles	88

ACRONYMS

AFM	Atomic Force Microscopy
AOD	Acousto-optic Deflector
BSA	Bovine Serum Albumin
CDF	Cumulative Density Function
DTDP	2,2'-dithiodipyridine
EDTA	Ethylenediaminetetraacetic Acid
FRET	Förster Resonance Energy Transfer
GLOXY	Glucose - glucose oxidase - catalase oxygen scavenging system

NMR	Nuclear Magnetic Resonance
PCA	Protocatechuic Acid
PCD	Protocatechuate-3,4-dioxygenase
PCR	Polymerase Chain Reaction
PSF	Point Spread Function
SAXS	Small Angle X-ray Scattering
SNR	Signal-to-Noise Ratio
TCEP	Tris(2-carboxyethyl)phosphine
WLC	Worm-like Chain
eWLC	Extensible Worm-like Chain

Part I

INTRODUCTION AND THEORY

1

INTRODUCTION

Proteins are ubiquitous molecular components in all prokaryotic and eukaryotic cells. Based on a limited “alphabet” of 20 amino acids, a plethora of different folds and conformations can be found. Although the basic aspects of folding are understood, the sheer number of different combinations of amino acid sequences make *de novo* structure prediction for proteins (unlike for the considerably simpler systems DNA and RNA) difficult. The process of folding poses even greater challenges to theoretical models. The description of a high-dimensional and rugged folding landscape where the native structure is the state with the lowest free energy is particularly successful [77].

Recent advances in computational technology have permitted the simulation of multiple folding and unfolding events for even larger proteins in all-atom simulations [61, 92]. These simulations found fast transitions between metastable states and oftentimes complex trajectories with a variety of intermediates [60].

However, these findings were until recently difficult to confirm experimentally. While ensemble techniques have been used to find intermediate states in protein folding [109], they are limited to observing average properties and cannot be used to follow the folding properties of single molecules. Single molecule techniques, however, allow to watch a single molecule over long time spans. Among the most successful techniques is Förster Resonance Energy Transfer (FRET), which has been used to study conformational changes in proteins [86, 87], as well as their folding behavior [81]. Fluorescence techniques also have been used to quantify friction in the unfolded state [93] and even the time it takes to cross the folding/unfolding barrier [26].

Another single molecule technique that has been particularly successful is force spectroscopy. Atomic Force Microscopy (AFM) as a tool to study the unfolding of single proteins has been developed in the late 1990s [88] and was further improved to study various aspects of the folding and unfolding process, such as the force-dependent kinetics [89], misfolding [76], different pulling geometries [18], the interaction with ligands [16, 17, 52] or near-equilibrium folding/unfolding studies [12, 53].

While AFM provides the general feasibility to follow single molecules along their folding/unfolding trajectory, the technique suffers from some drawbacks. Due to the necessity to couple one end of the protein to a surface, it is intrinsically susceptible to signal deterioration by drift. Also thermal noise tends to obscure the signal at low forces in the force regime below about 50 pN. The development of another mechanical manipulation technique, optical tweezers, has advanced significantly in recent years. Optical tweezers is now a feasible tool for the single molecule study

of protein folding [19, 43, 91] as well as the folding of nucleic acid hairpins [105–107]. Due to its superior resolution over AFM in the low force range, optical tweezers was chosen as measurement technique in the work presented here.

I will give a general introduction to optical tweezers in force spectroscopy in Chapter 3 and demonstrate its applicability to study the folding and unfolding behavior of the small calcium binding protein calmodulin (for a general introduction, see Chapter 2), a theory of the expected energetics and kinetics under ligand binding conditions is presented in Chapter 5 at unprecedented resolution. Further, I will demonstrate how advanced statistical methods can be used to analyze the complex trajectories obtained with this technique (Chapter 4). Finally, Chapter 6 contains the experimental findings for the folding/unfolding of calmodulin at high concentrations of calcium. Chapter 7 discusses the influence of calcium binding on the folding/unfolding properties of calmodulin's isolated domains.

2

CALMODULIN AND EF HAND PROTEINS

2.1 INTRODUCTION: EF HAND PROTEINS, LIGAND BINDING

Calcium regulated pathways play important roles in physiology. Calcium signaling is one of the major components of signal transduction in muscles and neurons. In general, cells build up a steep concentration gradient of calcium across the membrane, with the intracellular concentration being about 20,000-fold lower than the outside concentration [28]. Upon an outer stimulus, ion channels in the membrane open that allow the quick (within milliseconds) influx of calcium and cause the intracellular calcium concentration to rise from ≈ 100 nM up to ≈ 1 mM [28]. On the inside, calcium buffering and calcium sensory proteins work in accord to process the signal. The most prominent calcium sensor protein, calmodulin, works as an interface between the small-sized calcium ions to the size scale of the protein world. Structural rearrangements caused by calcium binding are thought to expose binding sites for kinases that can trigger further pathways. To understand the working principles of calcium binding proteins, first a common introduction on the structural basis that underlies most of the calcium binding proteins is necessary.

2.1.1 The EF hand principle

EF hands are common structural motifs in calcium binding proteins. Divalent ions, primarily calcium, can bind into a loop that is flanked by two alpha helices. This helix-loop-helix motif can be found in many members of the calmodulin superfamily. Most members of the superfamily have multiple copies of EF hands, predominantly an even copy number. Parvalbumin with its three EF hands (only two of which bind calcium) being an exception [54].

The pairing of EF hands into domains appears to be a common motif amongst the members of the calmodulin superfamily, with only few exceptions [74]. Close spatial packing of calcium binding sites is a reason for the effect of binding cooperativity, i. e. the enhancement of binding affinity of the second binding site, if already one ligand is bound. Effectively, binding cooperativity limits the concentration range over which only one of the binding sites is occupied. The protein is hence most likely to be either ligand-free or saturated with two ligands, but less likely to have only one ligand bound.

The name EF hand was coined by Kretsinger and Nockolds. In their paper they described the helices E and F of parvalbumin: "The over-all configuration of the EF region is remarkably similar to a right hand with thumb and forefinger extended at approximate right angle and the remaining three fingers clenched." [55]

2.1.2 Calmodulin

Calmodulin is an alpha-helical protein consisting of four EF hands. The two N-terminal EF hands form the N-terminal domain, the two C-terminal domains form the C-terminal domain. Its length is about 148 residues and it is found in all eukaryotic cells [24].

Genetic analyses suggested that the two domains of calmodulin are a result of a gene duplication of a one-domain precursor protein. The two EF hands per domain are assumed to be the result of an earlier gene duplication [54]. Based on data from the same homology analyses, EF hands can generally be classified into *odd* and *even* structures. In calmodulin, the first odd-even pair, consisting of EF hands 1 and 2 constitute the N-domain, the second odd-even pair, EF hands 3 and 4, the C-domain. Despite the fact that evolution has caused the preferred association of odd-even paired domains, certain fragments have been found to fold into even-odd paired structures, yet with much lower stability [58].

Crystallography studies have revealed a two-domain structure for the calcium-loaded holo state as well as for the calcium-free apo-state [21, 57].

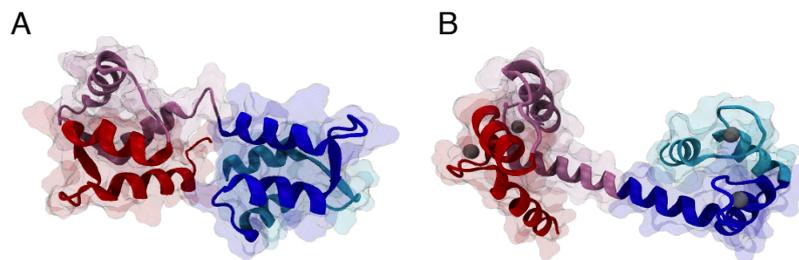


Figure 1: Crystal structure of calmodulin. (A) Crystal structure of apo-calmodulin (PDB 1cfc). (B) Structure of holo-calmodulin (PDB 1cll). The EF hands are colored from dark red to bright red (N-terminal domain) and dark blue to light blue (C-terminal domain).

The crystal structures suggest that upon calcium binding, the central linker region transforms from an unstructured and floppy conformation (see Figure 1A) into a more extended conformation (see Figure 1B) and thus exposes hydrophobic residues, that can serve as binding sites for peptide sequences. Nevertheless, the notion of the extended alpha-helical central linker structure in the calcium loaded form has been challenged by some researchers. Due to the incompatibility of the crystal structure data with small angle X-ray scattering data in solution, it has been suggested that the central linker in solution is, indeed, flexible [46]. A more recent crystal structure also shows an alternative more “closed” conformation of calcium loaded calmodulin [39].

The structural rearrangements in calmodulin upon calcium binding along with the associated change in binding affinity to peptides classify the protein as a calcium sensor. The pairing of EF hands in the individual

domains is an important part of this mechanism. The close connection between the neighboring EF hands helps coordinate calcium binding between them. This so-called binding cooperativity ensures that by modulating the affinity of the EF hands, depending on whether or not the other EF hand is calcium bound, there is only a very narrow range where only one calcium ion is bound to a domain. In most of the dynamic range, there are either zero or two calcium ions bound. The affinities for calcium are dependent on ionic strength and lie the range of tens of micromolar for the N-terminal domain and micromolar for the C-terminal domain [6, 64].

A number of studies have investigated the properties of the single domains. Historically, these were obtained by enzymatic cleavage in the central linker region by trypsin. Following the nomenclature, the N-terminal domain is also referred to as Tr₁C, the C-terminal domain as Tr₂C. It was found that the two isolated domains are both structurally and functionally very similar to the domains in full-length calmodulin, indicating that they act largely independently in full-length calmodulin.

2.1.2.1 Earlier biophysical studies on calmodulin

A number of biophysical studies in the last decades have focused on the thermodynamic stabilities and calcium binding properties of EF hand proteins, especially calmodulin [6, 64, 68, 69]. To date, the macroscopic binding constants of full-length calmodulin, as well as the macroscopic binding constants of its tryptic fragments is well known at a variety of ionic strengths. However, folding studies on calmodulin are not as abundant as calcium binding studies. Nuclear Magnetic Resonance (NMR) experiments have revealed the time scales of conformational changes within the domains upon calcium association and dissociation [66]. Even more data was obtained from rapid mixing studies of calmodulin with calcium [79] and simulations [23, 98].

Earlier single molecule studies using AFM could confirm that the two domains of calmodulin can fold independently [53] and undergo rapid transitions between the folded and unfolded states. Also at the single molecule level, the binding to ligands could be investigated [52]. Optical tweezers have also been used earlier in our group to study the folding of calmodulin [42]. However, although indications of some of the effects that will be presented in this work were already visible in this early data, experimental issues, such as the lack of a functional oxygen scavenging system or the use of polystyrene beads instead of silica beads have prevented a more detailed analysis.

For instance, in the calcium binding protein calbindin D_{9k}, the cleavage of the linker between the two EF hands result in a reconstituted dimer, but abolishes the cooperativity [65].

“Tr_NC: Trypsin fragment N of calmodulin”.

3 | OPTICAL TWEEZERS

Optical tweezers as a tool to manipulate microscale objects has been developed in the 1980s, first as a device to capture minuscule objects, such as viruses or bacteria [4]. More recently, the technique has also been used to apply forces to molecules and study their mechanical behavior. Notable are experiments studying the stepping of molecular motors or polymerases [104], as a surface-detached scaffold to study the motion of molecular motors [14], as well as to study the folding/unfolding of nucleic acids and proteins [3, 19, 37, 48, 91, 106].

3.1 THE PRINCIPLE OF OPTICAL TRAPPING

The principle of optical trapping can be rationalized in a simple ray-optical picture. When light is refracted at an object, it transfers a part of its momentum onto the refracting object [96]. Using an objective with high numerical aperture and high power it is possible to trap small dielectric objects. The efficiency of trapping is also determined by the object's refractive index and shape. Due to their symmetry, the trapping of microspheres or bead-like objects is described most easily. Most commonly used in today's optical tweezers measurements are beads made from polystyrene or silica. However, also other materials such as titania or even beads with coatings have been utilized [11, 49].

Shown in Figure 2 is a simple ray-optical schematic of a bead-like object captured in a laser focus created by a high numerical aperture objective.

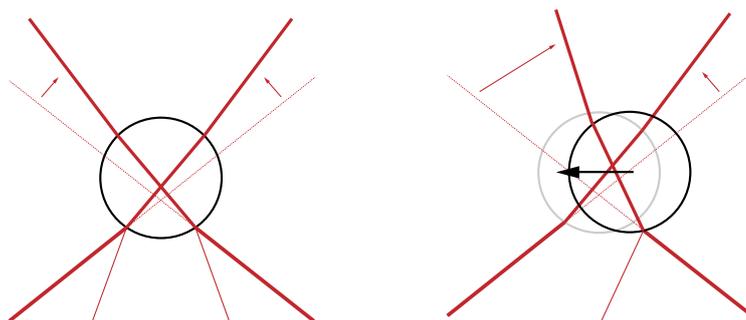


Figure 2: Simple ray optics picture of optical tweezers. Reflection and refraction of incoming light rays (red) result in a transfer of momentum onto the trapped bead. In the left case, all forces cancel out and the bead is stably trapped. In the right case, the bead displacement from the initial position results in a net restoring force back to the center.

An instructive applet for optical tweezers can be found online at <http://www.physics.gla.ac.uk/Optics/projects/tweezers/trapsimulation>

The figure only shows the outer-most rays. Upon incident with the first surface, part of the light is reflected back, resulting in an axial force on the bead (scattering force). However, the refraction of light on the first and second surfaces result in a transfer of momentum from light onto the bead which results in a force in the negative axial direction. This force precisely cancels out with the scattering force and causes the bead to be stably trapped in the focus. If the bead is deflected from the center, as illustrated in the right part of [Figure 2](#), the directions of the reflected and refracted beams change, resulting in a net force on the bead that drives it back into the stably trapped position.

The picture gets more complicated when the wave-like properties of light are taken into account [96]. Here, the ray assumption, that the size of the trapped particle d is much bigger than the light wavelength λ , is no longer true. Nevertheless, to a first-order approximation the ray description still holds. Taking into account the finite size of the trapping spot, one can assume an approximately gaussian shaped trapping potential. This is shown for a one-dimensional projection in [Figure 3](#). The force acting on the bead, which is defined by $F = -\nabla\Phi$ for a potential Φ is approximately linearly dependent on the displacement of the bead from the trap center for small displacements (indicated by the dark shaded region in [Figure 3](#)). An optical trap therefore resembles in the limit of small bead displacements a Hookean spring.

Notably, also very small objects with $d \ll \lambda$, such as quantum dots, can be efficiently trapped

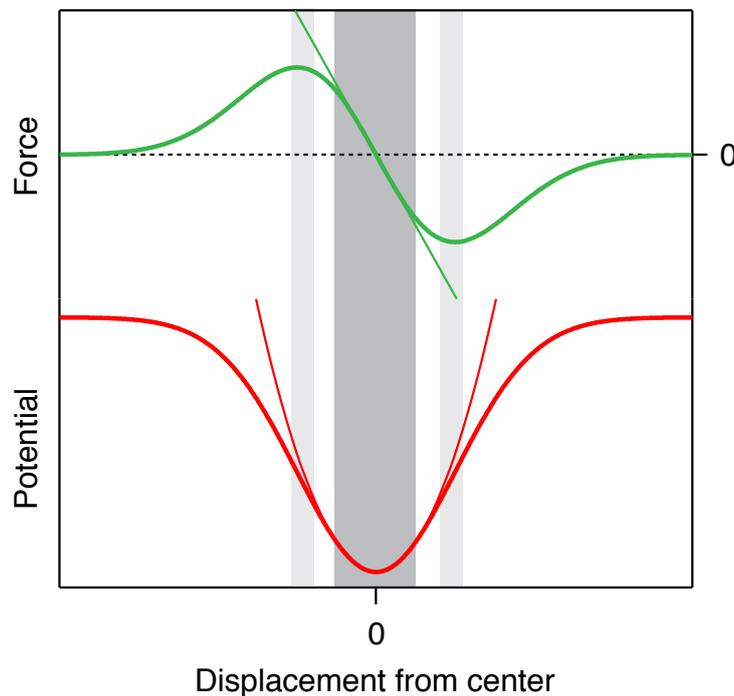


Figure 3: Potential of an optical trap and resulting force. For small bead displacements (dark shaded area), the potential is approximately harmonic, corresponding to a Hookean force response (thin lines).

3.2 OPTICAL TWEEZERS IN FORCE SPECTROSCOPY

In a typical force spectroscopy measurement, force is applied to a sample by “pulling” it at specified positions, hence the term *optical tweezers*. The increased resolution over AFM experiments [72] makes it an ideal low-force complement, where the sample molecule can be manipulated with a Hookean pulling apparatus. In optical tweezers, more sophisticated measurement modes are possible.

In addition to the Hookean spring pulling mode, it is possible to exploit the nonlinearity of the optical trapping potential. In certain regions of the potential, the force does not depend much on the displacement, indicated as lightly shaded regions in Figure 3. Here, the trap stiffness $k = -\partial/\partial x F$ is essentially zero. The force for small deviations from the maximal/minimal force value is thus constant. Measurements in this area of the potential are often called a *passive force clamp* [44].

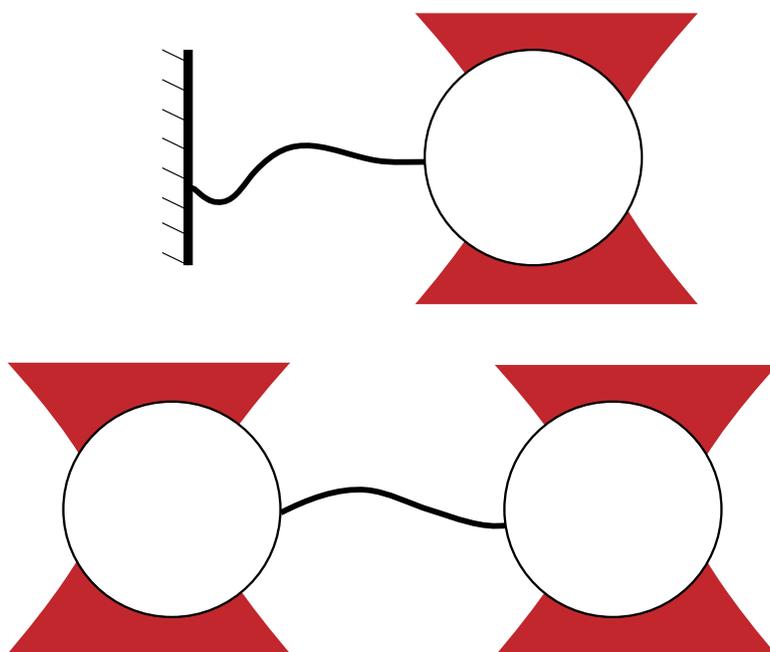


Figure 4: Comparison between single beam and dual beam optical tweezers. Single beam experiments resort to tethering one end of the molecule to a fixed surface (upper part). Dual beam do not rely on a fixed surface. Instead, both attachment points are trapped in solution.

The Signal-to-Noise Ratio (SNR) of force spectroscopy measurements is often determined by mechanical influences. Especially if one end of the sample molecule is attached to a fixed surface, i. e. to a point not in suspension, exterior noise easily couples into the measurement and deteriorate the signal. A configuration where two beads are trapped simultaneously and the molecule is tethered between those two beads can significantly decrease the influence of exterior mechanical influences, such as vibrations. Figure 4 shows the two possible configurations.

If the measurement apparatus is in equilibrium, the average forces acting on in all components are equal. The force can hence be measured everywhere in the system.

Further improvements on the SNR can be achieved, if the force in a dual beam setup is not determined by the signal from a single bead alone, but from the sum of the absolute values of both bead displacements. Moffitt et al. showed that if both traps are created from the same initial beam, noise due to pointing instabilities and beam-direction deteriorating air density fluctuations can be significantly reduced [72].

3.3 THE OPTICAL TWEEZERS INSTRUMENT USED IN THIS WORK

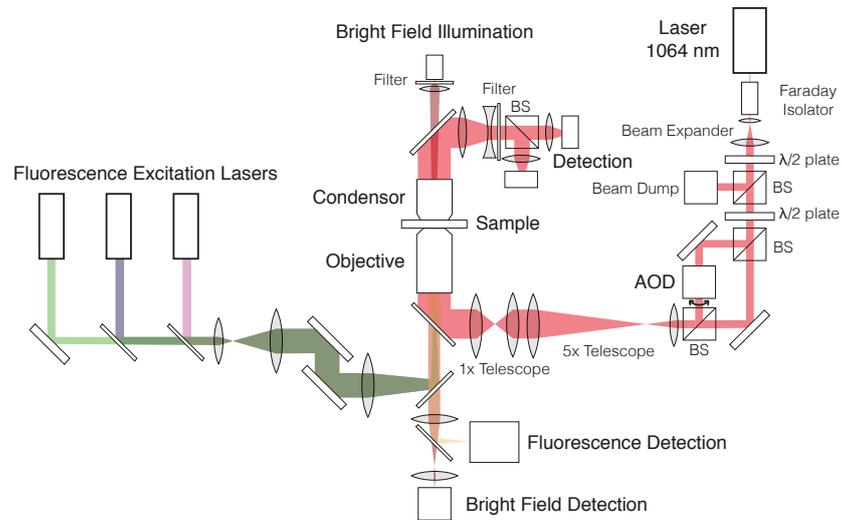


Figure 5: Schematic of the dual beam setup used in this work.

The setup used for this work was designed by Christof Gebhardt (for details, see [42]) and is shown schematically in Figure 5. Both traps are created from the same Nd:YAG laser at a wavelength of 1064 nm. The beam is split into two with orthogonal polarizations. The path length where the two beams are separated is intentionally kept small and enclosed in a special casing to reduce the influence of air fluctuations. For most of the optical path the two trap beams are aligned and pass through the same optical elements. One of the beams is steered using an Acousto-optic Deflector (AOD). The AOD also frequency-shifts one of the beams, which helps to reduce interferences in the detection stage.

The wavelength in the near-infrared minimizes the heating in the sample coming from absorption of biological molecules and water [96].

3.4 FORCE APPLICATION IN A DUMBBELL ASSAY

In force-spectroscopy, a commonly used configuration for applying force to a sample is the so-called dumbbell configuration. Figure 6 shows the

general idea. Using handle molecules of dsDNA, a single sample protein can be attached to functionalized dielectric beads.

The two different DNA handle molecules are produced by Polymerase Chain Reaction (PCR) on a lambda phage template. The sense primers are triply functionalized with biotin or digoxigenin. The anti-sense primers contain an abasic site that causes the polymerase to fall off and leave a single stranded overhang on the molecule. Small ssDNA oligos (shown in green in Figure 6) can hybridize to this region. These oligos are functionalized with a thiol on their 5' ends and can be attached via a disulfide bond to the protein prior to the handle assembly.

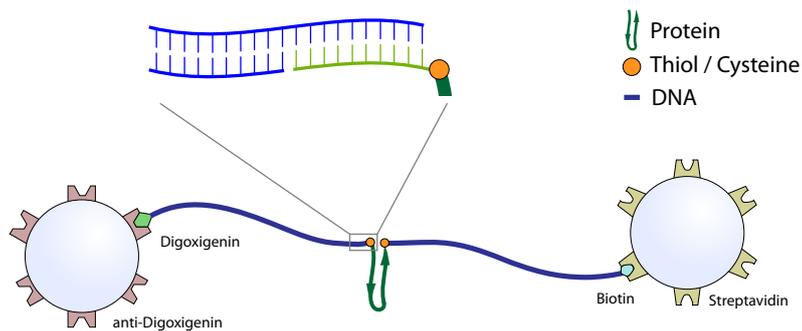


Figure 6: Schematic for a dumbbell configuration in an optical tweezers assay. Functionalized dsDNA molecules (blue) are attached using digoxigenin/anti-digoxigenin or biotin/neutravidin interactions to silica beads. By introducing a single-stranded overhang, these DNA handles can hybridize to ssDNA oligos, that were previously linked to the protein using a thiol bond.

Several strategies can be employed for creating these oligo-protein hybrids. For this work, cysteine residues were inserted at the desired attachment points (mostly at the N and C-terminus). Exposed cysteines can form disulfide bridges and be used to form multimers of the protein, or, if desired, as in this case, use as a docking site for reactive thiol linkers of the oligos. To this end, a strategy put forward by [Cecconi et al.](#) was used [20]. Directly after expression and a first purification step, the free cysteines were treated with 2,2'-dithiodipyridine (DTDP). The treatment is beneficial in two ways. First, multimerization of proteins, which is undesired for the kind of intended single molecule experiments, is inhibited. Second, the DTDP-bound cysteines on the protein allow a faster reaction with DTDP-untreated thiols on the oligos. Reaction times in this case are on the order of a few hours and were mostly performed over night (for details, see [Appendix D](#)).

Oftentimes, many different proteins are to be measured in the same N-C-terminal pulling direction. To avoid experimental difficulties at the production stage of these constructs, it is helpful to build a platform with defined reaction sites for oligos. To this end, a strategy introduced by Christof Gebhardt was used. The sample protein was inserted recombinantly into an expression vector between two ubiquitin molecules. These

ubiquitins were designed with N- and C-terminal solvent-exposed and reactive cysteines. The reactivity was increased by flanking the cysteines with lysine residues that, at neutral pH, are positively charged and facilitate the reaction with negatively charged DNA. The vector further includes a His₆-tag for easy protein purification. Further advancements include tryptophan residues outside of the region where force is applied to enhance the UV absorbance of the construct and make facilitate the handling at the production stage.

The flanking with ubiquitin molecules is advantageous in several ways. First, if the protein to be measured is flexible, there is a considerable chance that the N- and C-termini come close and form a crosslink between the cysteines. The ubiquitins in this case form a spacer that keeps the cysteines apart and solvent exposed. Second, although ubiquitins have been shown to display a much higher mechanical stability than most proteins studied here [89], ubiquitin unfoldings can be observed in the assay. They can be readily refolded and serve as markers for a successful formation of a functional dumbbell configuration. Third, the defined attachment points for oligos makes the reaction less error-prone and increases the yield of successfully attached handles. Fourth, the fusion protein approach turned out to increase solubility for some protein samples.

Further improvements on the assembly can involve oligos with reactive maleimide groups, which are unable to dimerize but instead specifically attach to reactive cysteines under certain pH conditions. The reduction of dimerized oligos in the sample significantly reduces the background fraction of tethers that do not include a protein at all.

The choice of beads was governed by the following considerations: Although a wide range of bead sizes can be trapped in the instrument, there is a trade-off between visual recognizability of the beads in the bright-field part of the instrument (the bigger the bead, the better), and their hydrodynamic drag. The dominant part of friction is governed by the Stokes term, which is proportional to the bead radius, hence, smaller beads allow the measurement of faster systems. The trapping strength is also governed by the difference in index of refraction between the bead and the liquid they are suspended in. Polystyrene outperforms here silica as a material. However, silica has the distinct advantage of reduced sample heating [80] and lowered damage to the sample due to the creation of free oxygen radicals [59]. In this study, solely silica beads with a diameter of 1 μm were used.

3.4.1 The problem of oxygen damage

Earlier measurements with optical tweezers in our group oftentimes suffered from “dying” molecules, i. e. the pattern of transitions changed irreversibly over time (see, for example, [42, Appendix B.2]). A publication by Landry *et al.* identified free oxygen radicals that were produced by the high light power irradiation of the beads as a possible reason for the observations. The study investigated a series of oxygen scavenging agents

and found that ascorbic acid, which was used in our group earlier, performs worse than other systems, such as Glucose - glucose oxidase - catalase oxygen scavenging system (GLOXY) or Protocatechuic Acid (PCA)/ Protocatechuate-3,4-dioxygenase (PCD). Further, the study suggested the use of silica instead of polystyrene as material for the beads. In this work, we followed both suggestions. All measurements were performed in the presence of an oxygen scavenging system (GLOXY, unless noted otherwise) and with silica beads. Using this approach, the “dying” of molecules that was observed earlier could be reduced drastically. While earlier the duration of measurements was limited by the time until the molecule got irreversibly damaged (up to a few minutes), in this study, the major temporal limitation was due to additional beads that fell into the traps during measurement. Irreversible switching of the molecular behavior was almost never observed.

4

ANALYSIS OF SINGLE MOLECULE TRAJECTORIES

4.1 HIDDEN MARKOV MODELS

The typical outcome of single molecule experiments are time-trajectories of some kind of observable. In fluorescence measurements, for instance, a photon count, or, in force measurements as in this study, a series of force or extension values. If the sample undergoes transitions from one state to another, this is reflected in the measured observable. In an ideal world, we could directly infer the state of the sample from the observable. For instance, in an optical tweezers protein folding/unfolding study of a single domain of calmodulin, an extension of zero would indicate a folded protein, an extension of about 30 nanometers would indicate an unfolded one. In reality, however, the signal is usually deteriorated by instrumental and/or thermal noise. In the given example, it would not be clear what a measured extension of ten nanometers would correspond to. Furthermore, in complex situations, many states might project onto the same observable. Say, a misfolded intermediate that also has an extension of zero, but a much lower stability and thus lower lifetime. As will be explained in later chapters, all of the mentioned examples occur in the folding of calmodulin. Hidden Markov models are a useful and versatile method that can be used to analyze these kind of trajectories.

4.1.1 Introduction

Hidden Markov models have been developed in the 1960s and 1970s and mainly applied to digital signal processing applications, such as speech recognition [83]. Only recently, they were also applied in the analysis of single molecule data [9, 56, 70, 81, 87].

Figure 7 illustrates the principle of a Hidden Markov model. The model is based on the assumption that the sample performs a memory-less continuous-time Markov process (i. e. a series of transitions between states), that is hidden from the experimenter.

The model works on “snapshots” of the trajectory, i. e. on the discrete time representation of the trajectory. At each time point, the sample molecule can either stay in its current state or transition into another state (Figure 7A). The probability for a transition from a state i to a state j is called the *transition probability* T_{ij} . For the analysis of constant trap separation or constant force traces, T_{ij} can be assumed time-independent.

Further, the model assumes an initial distribution for the probability to obtain a certain value of the observable, given that the molecule is in

The assumption is a result of the description of protein folding as a diffusion process in a potential. In the overdamped case, the fate of a transition is independent of what happened before, resulting in a memoryless process.

a certain (hidden) state (Figure 7B). These distributions are called the *emission probabilities* $E_i(x)$, in this case the probability to observe an extension x , given that the molecule is in state i . Again, the emission probabilities can be assumed to be time-independent for the applications presented here. The set $\mathcal{M} := \{E_i(x), T_{ij}\}$ is called the *model parameters*.

Using prior assumptions for the emission and transition probabilities, the model now uses probabilistic arguments to infer the hidden state of the molecule at each time point from the whole trajectory of data points (Figure 7C).

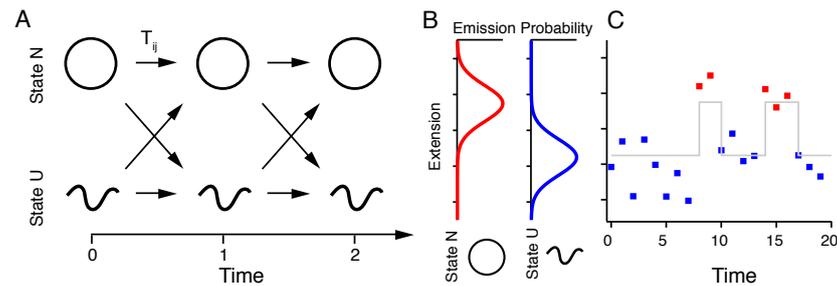


Figure 7: A toy example for a Hidden Markov model. (A) The hidden layer. On a discrete time scale, the protein can at each time point either stay in its previous state or change its state. The probabilities for transitioning are called T_{ij} . (B) The two states are assumed to have each a different average extension with overlapping distributions. These distributions are called the emission probabilities $E_i(x)$. (C) An example for a time trajectory. Simply speaking, the model compares the extension values at hand with the given emission probabilities and uses the transition probabilities to infer the hidden state of the protein at any time point.

The following discussion contains a more detailed description of the theory of Hidden Markov models.

4.1.2 Data preparation

Many implementations of Hidden Markov models assume that the emission probabilities are Gaussian. Even though this is a sufficient approximation for many purposes, a more robust method is to let the model optimize the emission probabilities itself. In fact, in assays such as force-unfolding with optical tweezers, the emission probabilities, i. e. the Point Spread Function (PSF), are expected to be non-gaussian due to the non-Hookean properties of the dsDNA linkers [56].

While it is possible to calculate (or measure) a PSF and let the Hidden Markov model only optimize the determining parameters for the emission probabilities, it is simpler to let the system optimize the emission probabilities in full. This also has the advantage that supervision of the convergence of the model is facilitated. Improper classification of data-

points is often clearly discernible by bimodal emission probability distributions.

In practice, it is straightforward to discretize the trajectory of input data $x(t)$ into a number of bins (see Figure 8 for an illustration). These data are called *observation values* o_t .

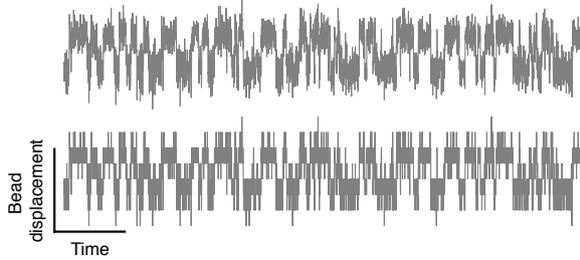


Figure 8: Discretization of observation data for Hidden Markov processing. Top: before processing, bottom: after discretization into 8 bins.

The input data $x(t)$ can be from any observable. Here, it is for most purposes the sum of the bead displacements $x = x_1 + x_2$, which is directly proportional to the trajectory of forces.

4.1.3 Initialization of the model parameters

As pointed out earlier, the PSF in a dumbbell setup with non-linear linkers is not necessarily Gaussian. Nevertheless, Gaussian approximations for the PSF are adequate initial guesses for the emission probabilities.

For the following discussion, let's assume we want to perform a Hidden Markov analysis on a (previously discretized) trace o_t of M data points. Let us further assume, that the hidden process has N states.

We can then initialize the emission probabilities $E_i(x)$, $i \in \{0, 1, \dots, N-1\}$ with the normalized Gaussian PSF-approximations. Here, the emission probabilities are assumed to be discretized into the same number of bins as the previously discretized data trace:

$$E_i(x) = \frac{\exp\left(-\frac{(x-\mu_i)^2}{\sigma_i}\right)}{\sum_{j=0}^{N-1} \exp\left(-\frac{(x-\mu_j)^2}{\sigma_j}\right)} \quad (1)$$

Here, μ_i and σ_i are the means and widths of the emission value distributions of the single states (cf. Figure 7B). It is practical to determine these values by e. g. creating a histogram of the data trace and fitting it with a sum of Gaussians.

The transition probability matrix T_{ij} are chosen as probability per time step and normalized such that $\sum_{j=0}^{N-1} T_{ij} = 1$.

4.1.4 Classification of data points – The forward-backward algorithm

As pointed out earlier, the Hidden Markov model uses probabilistic arguments to infer the hidden state trajectory from the given data trajectory

o_t based on the model parameters $\mathcal{M} = \{E_i(x), T_{ij}\}$. The algorithm to do so is called the *forward-backward algorithm* [82, 83].

The *forward probabilities* $\alpha(i, t)$ are the probabilities for a system, given that it is in state i at time point t , and given the model parameters \mathcal{M} , to produce the given series of observation values $o_{t'}, t' = 0 \dots t$. They can be calculated recursively as

$$\alpha(i, t) = \begin{cases} E_i(o_0), & t = 0 \\ \sum_{n=0}^{N-1} E_i(o_t) \cdot T_{ni} \cdot \alpha(n, t-1) & t > 0 \end{cases} \quad (2)$$

Accordingly, the *backward probabilities* $\beta(i, t)$ are the probabilities for a system, again given that it is in state i at time point t , and given the model parameters \mathcal{M} , to produce the observation values $o_{t'}, t' = t+1 \dots M-1$. The backward probabilities can be calculated recursively from the end by

$$\beta(i, t) := \begin{cases} 1, & t = M-1 \\ \sum_{n=0}^{N-1} E_n(o_{t+1}) \cdot T_{in} \cdot \beta(n, t+1) & t < M-1 \end{cases} \quad (3)$$

Simply put, $\alpha(i, t)$ calculates the probability for measuring the given trajectory until t while $\beta(i, t)$ gives the probability for measuring the given trajectory from $t+1$ on to the end.

The product $\alpha(i, t)\beta(i, t)$ is now the probability to produce the given data trajectory o_t , given the model parameters \mathcal{M} , and given that we are in state i at time point t .

Using Bayes' theorem, this can be inverted to obtain the probability that the system is in state n at time point t , given \mathcal{M} and the data trajectory o_t :

$$p(n, t) = \frac{\alpha(n, t)\beta(n, t)}{\sum_{i=0}^{N-1} \alpha(i, t)\beta(i, t)}. \quad (4)$$

We can use $p(n, t)$ to assign each data point to the state with the maximal probability $p_{\max}(t) := \max_n p(n, t)$. Using the described optimization techniques that are presented in the following section, the model as described converges to a *maximum likelihood estimator* for the model parameters. An extension to the model by [Chodera et al.](#) allows to obtain statistical information about the uncertainty of the model parameters [25].

4.1.5 Optimization of model parameters

The classification as described above is optimal if the estimates for the transition probabilities and emission probabilities i. e. \mathcal{M} are the true parameters. A measure for the performance of the model is the likelihood function $\mathcal{L} = \sum_{i=0}^{N-1} \alpha(i, t)\beta(i, t)$. \mathcal{L} is the overall likelihood to produce the trajectory o_t given \mathcal{M} . The model parameters \mathcal{M} can be "optimized" in the sense of a maximization of \mathcal{L} using a set of equation called the *Baum-Welch algorithm* [5].

Note that \mathcal{L} is independent of t .

An estimate for the emission probabilities is given by the distribution of emission values for each state. After a run of the forward-backward algorithm, these can be calculated by

$$\hat{E}_i(x) = \frac{\sum_{t|x=o_t} \alpha(i, t) \beta(i, t)}{\sum_t \alpha(i, t) \beta(i, t)}. \quad (5)$$

Likewise, an estimate for the transition probabilities T_{ij} can be generated by considering the probability that a state i at time point t transitions into a state j at time point $t + 1$:

$$\frac{\alpha(i, t) \cdot T_{ij} \cdot \beta(j, t + 1) \cdot E_j(o_{t+1})}{\mathcal{L}}. \quad (6)$$

Consequently, T_{ij} can be approximated by the fraction of time points where a transition from i to j occurs over the total number of data points that constitute most likely a state i (which is given by $\sum_t \alpha(i, t) \beta(i, t) / \mathcal{L}$)

$$\hat{T}_{ij} = \frac{\sum_t \alpha(i, t) \cdot T_{ij} \cdot \beta(j, t + 1) \cdot E_j(o_{t+1})}{\sum_t \alpha(i, t) \beta(i, t)} \quad (7)$$

Equation 5 together with Equation 7 are commonly referred to as the Baum-Welch algorithm. It can be shown that by substituting E_i with \hat{E}_i and T_{ij} with \hat{T}_{ij} , \mathcal{L} increases. Hence, the forward-backward algorithm together with the Baum-Welch algorithm constitutes a maximum likelihood method.

4.1.6 The Viterbi algorithm

The state assignment using Equation 4 only uses part of the information in the model. Instead of calculating the most likely state at each data point separately, the Viterbi algorithm calculates the most likely sequence of states that produces the data trajectory o_t , given the model parameters \mathcal{M} [41, 101]. In practice, it turns out that classification using the Viterbi algorithm is more sensitive to the accuracy of the transition probabilities T_{ij} than the classification using the maximum probability (Equation 4). For meaningful data point assignment using the Viterbi algorithm, either a good prior estimate of T_{ij} , or converging optimization using the Baum-Welch algorithm is necessary. Viterbi probabilities can be calculated as follows:

We define $v(i, t)$ as the maximal probability that a model with underlying parameters \mathcal{M} and unspecified sequence of hidden states produces the observations $o'_t, t' = 0 \dots t$ and is in state i at time t . It can be calculated recursively:

$$v(i, t) = \begin{cases} \alpha(i, 0) E_i(o_0), & t = 0 \\ \max_j (v(j, t - 1) \cdot T_{ji}) E_i(o_t), & t > 0 \end{cases} \quad (8)$$

Using this definition we can calculate the most likely state of the last data point at time $M - 1$

$$s(M - 1) = \arg \max_j v(j, M - 1). \quad (9)$$

State assignments at lower values of t can be calculated recursively:

$$s(t-1) = \arg \max_j (v(j, t-1) \cdot T_{js(t)}) . \quad (10)$$

In the interest of easy computability, it is good practice to calculate the values

$$\psi(i, t) := \begin{cases} 0, & t = 0 \\ \arg \max_j (v(j, t) \cdot T_{ji}), & t > 0 \end{cases} \quad (11)$$

while computing [Equation 8](#). The state assignment can then be performed by calculating the recursion of s for $t < M - 1$ using

$$s(t-1) = \psi(s(t), t) . \quad (12)$$

4.1.7 Numerical issues

The calculation of the equations as presented here can be numerically challenging, since the probabilities involved are often very small. The products of these small probabilities easily cause numerical underflow. A technique of renormalization helps to keep all numbers in numerically accountable ranges. Details can be found in [Appendix A](#).

4.1.8 Performance of the algorithms

Simulations can help to assess the performance of the data point classification by the algorithms introduced above. Challenging data are, for instance, very rapid transitions, or data with strongly overlapping emission profiles, i. e. force distributions. In the extreme case, two distinct levels might even share the exact same average force. Nevertheless, while simple classifiers, such as thresholding, fail for these data, the states are still separable by a Hidden Markov model, if the overlapping levels differ in their lifetimes.

[Figure 9](#) shows an example for such a simulated trace. The red and green levels share the same average force. Both exchange with the purple level. Since the average lifetime of the green and red levels are well separated, a Hidden Markov classifier with Viterbi optimization is able to correctly classify up to 96 % of the data points (see [Figure 9C](#)). While the speed of convergence depends on the quality of the initial guess for the transition probabilities T_{ij} , the overall performance does not.

The full maximum likelihood optimization of emission and transition probabilities together with the Viterbi algorithm often fails in real-life data, when residual noise with correlation times in the millisecond range deteriorates the signal. In this case, it is often better to keep the transition probabilities T_{ij} fixed at a certain (low) value and rely on the per-data point classifier ([Equation 4](#)). [Figure 10](#) shows the performance of such a

A thresholding classifier defines a threshold value and assigns all data points with forces above the threshold to one class, and all data points with forces below the threshold to the other class.

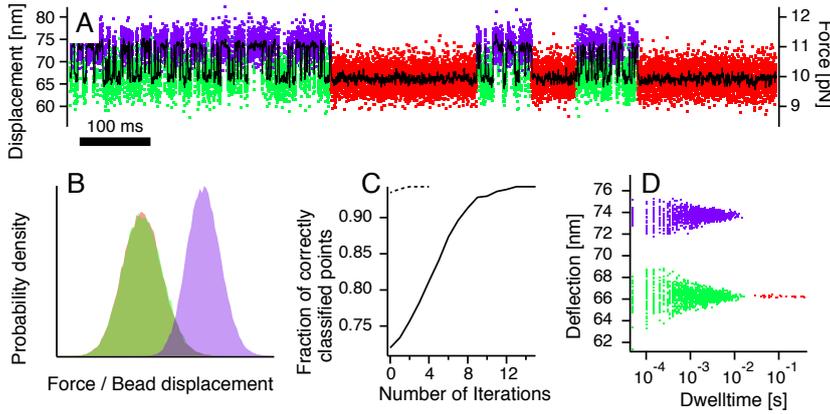


Figure 9: Assignment performance of a Hidden Markov model on simulated data with levels at identical forces. (A) Simulated trajectory with data points colored as classified by a Hidden Markov model. Note that the green and red states are indistinguishable in force but distinguishable in lifetime. (B) Emission probabilities after twelve iterations. (C) Performance of correct classification of data points using the Viterbi classifier. Dashed line: good initial estimates for T_{ij} , continuous line: bad initial parameters for T_{ij} . (D) Scatter plot of the average deflection all found dwells versus their lifetimes.

classification. Even though the transition probabilities were fixed at unrealistically low values, the system performed well and robustly in classifying the data points. Ambiguities in the assignment were low (for instance, more than 95 % of all data points were assigned with a p_{\max} value of more than 90 %, Figure 10B). The highest uncertainty was found at the transitions between states (cf. Figure 10A, upper trace). Wrong classifications of single data points in these ranges, however, only slightly shift the detected time point of the transition, and do not affect the overall state assignment.

4.2 TRANSITION RATES AND THE ISSUE OF MISSED EVENTS

4.2.1 Obtaining transition rate constants

After successful classification of the data points using a Hidden Markov model, we can use the optimized transition probabilities T_{ij} to calculate the transition rate constants k_{ij} . For a sampling rate δ , they are given by

$$k_{ij} = -\ln(1 - T_{ij}) \cdot \delta. \quad (13)$$

Naturally, these values are only meaningful if the transition probabilities were optimized and have converged. As pointed out earlier, the optimization sometimes fails when the analyzed trajectory is not fully Markovian,

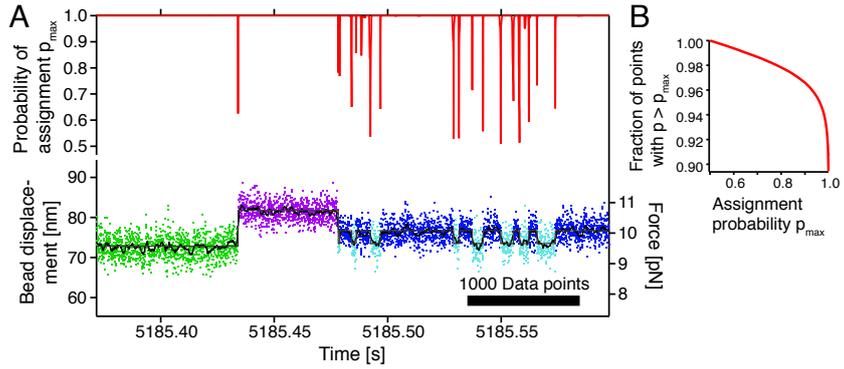


Figure 10: Assignment probabilities for a Hidden Markov model using the per-data point classifier (Equation 4). (A) Lower panel: Data trace with four states classified by a hidden Markov model into states (indicated by different colors). Upper panel: The assignment probability p_{\max} for the classification. (B) Most data points are classified with high certainty.

For instance, a memory is easily introduced if the trajectory of a dumbbell is recorded at sampling rates higher than the 3 dB frequency of the hydrodynamic damping of the beads.

i. e. it shows memory-effects, as could arise from sampled slow dynamics of the system. In reality, it shows, however, that a much more pronounced effect that prevents the success of full-optimization strategies, are high frequency oscillations that couple into the system from the outside. Moreover, sometimes the experimenter might only want to focus on slow dynamics and, for the time being, ignore faster dynamics. In these cases, the optimization of transition state probabilities is better disabled.

It should be also noted that the determination of error intervals for the transition rate constants as calculated from Equation 13 is challenging, yet feasible using a Bayesian Hidden Markov approach [25].

A more robust way of determining the transition probabilities is to first determine the off-rates from a certain state and then split it into its pathway components. The off-rate from a state i can be calculated as

$$k_i = \frac{1}{\langle \tau_i \rangle}, \quad (14)$$

where $\langle \tau_i \rangle$ is the average over all dwell times in state i .

Since after Hidden Markov classification we can track the fate of each transition, i. e. we know which state transitions into which, we can make use of this information to determine the transition rate constants. Since

$$\frac{k_{nj}}{k_{ni}} = \frac{N_{nj}}{N_{ni}} \quad (15)$$

and

$$k_n = \sum_{j \neq n} k_{nj}, \quad (16)$$

the transition rate constant from state n to state i is

$$k_{ni} = \frac{k_n}{1 + \sum_{j \neq n} \frac{N_{nj}}{N_{ni}}}. \quad (17)$$

Here, N_{ni} is the number of transitions detected from state n to state i .

However, this approach also has a few drawbacks. First, it is not clear *a priori*, that the determined distribution of dwelltimes is indeed single exponential. As practice shows, this should be verified, e. g. by fitting a single exponential distribution to it. The value for k_i is then easily obtained as a fit parameter. Furthermore, problems arise if short events are missed in the analysis. Consequences of this effect and a way to compensate for it are presented in the next section.

4.2.2 The problem of missed events

If the kinetics in the system are fast, a number of dwells and transitions will either be missed by the measurement, or, depending on the performance of the classifier, missed in the analysis. Nevertheless, following the assumption that the distributions of all dwelltimes are single exponential, the effect of missed events can be compensated for.

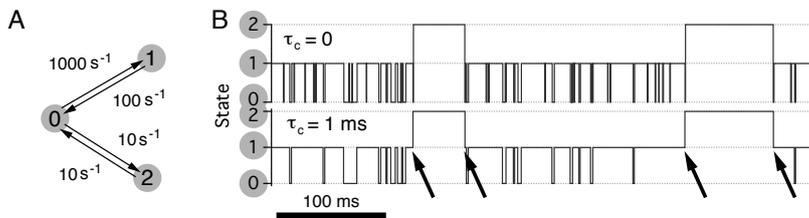


Figure 11: Effect of temporal detection limitations in an idealized three-state network. (A) Connectivity and kinetics in the simulated network. Transitions between state 1 and state 2 were not allowed. (B) Simulated trace in the given network at no temporal cutoff (top trace) and a cutoff of 1 ms (bottom trace). The cutoff leads to the appearance of transitions between the states 1 and 2 (arrows).

To illustrate the problem of the missal of short events, consider the example illustrated in Figure 11. In a three-state system with states 0, 1 and 2, there is a fast exchange between states 0 and 1, and a slow exchange between states 0 and 2. State 0 is comparatively short-lived. There are no transitions between the states 1 and 2. If a temporal cutoff τ_c is introduced, all dwells shorter than τ_c will be missed. A considerable fraction of short dwells will therefore be missed; for instance, in the given example of $\tau_c = 1 \text{ ms}$, about 64 % of all dwells in state 0. Consequently, there will be a number of falsely detected transitions between the states 1 and 2 (indicated with arrows in Figure 11B). Furthermore, due to the missal of short dwells, especially in state 0, the average dwelltime of all states that state 0 exchanges with, i. e. states 1 and 2, will be overestimated. This can also be seen in Figure 11B, where the average duration of dwells in state 1 appears considerably higher if a temporal cutoff is involved.

The following paragraphs contain a method that can account for both of the effects described above.

Integrated histograms are independent of the binning of data points and can be fitted unambiguously.

The temporal cutoff can also be included in Equation 14, e. g. using $k_i = \frac{1}{\langle \tau_i - \tau_c \rangle}$. Nevertheless, a histogram fit still has the benefit of quality control of the state assignment.

Correction for missed events

As pointed out earlier, for quality control as well as to account for the missal of short events, it is practical to, first, calculate normalized integrated histograms (also known as the Cumulative Density Function (CDF)) of the dwell times, and second use a single exponential fitting function, that assumes that there is a minimal duration T_{\min} and a maximal duration T_{\max} of dwells that can be detected:

$$f(t) = \frac{\exp(-t/\tau) - \exp(-T_{\min}/\tau)}{\exp(-T_{\max}/\tau) - \exp(-T_{\min}/\tau)}. \quad (18)$$

It is easy to introduce the temporal cutoff $\tau_c = T_{\min}$. This method should be preferred to Equation 14.

For the general case of an N-state network, we can expand on a correction method developed by [Crouzy and Sigworth](#) for traces obtained in patch clamp experiments [30]. First, we construct an augmented network, where the missed and the detected states are both represented. In the following discussion the missed version of a state i will be named i' . Figure 12 shows an example for a three-state network. The black lines are the measured rate constants \tilde{k}_{ij} . By determining all rate constants in this augmented network it is possible to approximate the “true” transition rate constants.

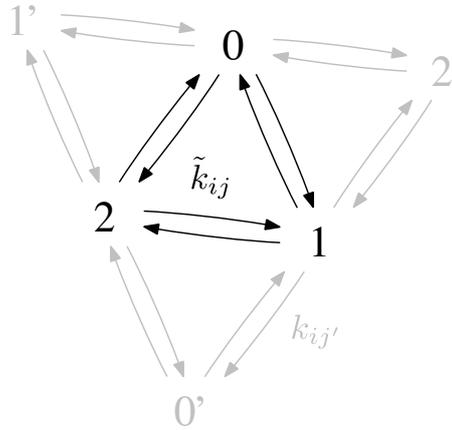


Figure 12: Augmented network of a three-state system with temporal cutoff. The states 0, 1, 2 (drawn in black) are longer than the cutoff and detected. The states 0', 1', 2' (grey) are the corresponding undetected states shorter than the cutoff. Black arrows designate detected transitions. Grey arrows are undetected transitions. By determining all rate constants in this network it is possible to approximate the real transition rate constants.

The network connectivity neglects any direct transitions between any two missed states. Although these transitions are technically possible, they are rare for typical kinetics and cutoff values and can be left out of the further considerations for simplicity. Since a detected state cannot

transition directly into its missed form, transitions of the form $k_{ii'}$ are also non-existent.

We can now estimate the fraction of dwells in a certain state i that are below the temporal cutoff τ_c and are thus missed. It is given by

$$f_i = \frac{\int_0^{\tau_c} \exp\left(-\frac{t}{\tau_i^*}\right) dt}{\int_0^{\infty} \exp\left(-\frac{t}{\tau_i^*}\right) dt} = 1 - \exp\left(-\frac{\tau_c}{\tau_i^*}\right). \quad (19)$$

Here, $\tau_i^* = \left(\sum_j k_{ij}^*\right)^{-1}$ is the true mean dwelltime in state i .

Using this fraction we can calculate the transition rate constants in the augmented network. Of all the transitions from a state i to a state j , a fraction f_j will be missed. Hence,

$$\frac{k_{ij}'}{k_{ij}^*} = f_j \Leftrightarrow k_{ij}' = f_j k_{ij}^*, \quad (20)$$

where k_{ij}' is the transition rate matrix of such transitions where the final state of the transition is missed.

As a first order approximation, the true rate constants k_{ni}^* are given by

$$k_{ni}^* = \tilde{k}_{ni} + k_{ni}' - \sum_{\substack{l \neq n \\ l \neq i}} k_{nl'i}, \quad (21)$$

where k_{ni}' is defined by Equation 20 and

$$k_{nl'i} = k_{nl}' \cdot \frac{k_{li}^*}{\sum_{j \neq l} k_{lj}^*}. \quad (22)$$

Note that Equation 22 describes only an *approximation* for the effective rate constant of transitioning from a detected state via a missed state to another detected state.

Equation 21 allows to calculate an estimate for the true rate constants k_{ij}^* based on the measured rate constants \tilde{k}_{ij} . Solutions can be computed numerically. Since the measured rates are only approximations of the rates that would be ideally measured in a system with infinite observation time, an exact solution to Equation 21 might not exist. However, by means of optimization techniques an estimate for the solution can be computed. To illustrate this, consider a two-state example, where the measured rates generally underestimate the true rates. With the true rate constants k_{01}^* and k_{10}^* and the corresponding measured rate constants \tilde{k}_{01} and \tilde{k}_{10} , Equation 21 reduces to:

$$\begin{aligned} k_{01}^* &= \tilde{k}_{01} + k_{01}' = \tilde{k}_{01} + (k_{01}^* \cdot f_1) = \\ &= \tilde{k}_{01} + k_{01}^* \cdot (1 - \exp(-\tau_c k_{10}^*)) \\ k_{10}^* &= \tilde{k}_{10} + k_{10}' = \tilde{k}_{10} + (k_{10}^* \cdot f_0) = \\ &= \tilde{k}_{10} + k_{10}^* \cdot (1 - \exp(-\tau_c k_{01}^*)) \end{aligned} \quad (23)$$

Simply put, the true rate is the measured rate plus the rate that is undetected, minus all those rates that transitioned via a missed state and were hence misclassified.

The optimal solutions $\{k_{01}^*, k_{10}^*\}$ to the nonlinear equations above can be computed numerically by minimizing $(E_0/k_{01}^*)^2 + (E_1/k_{10}^*)^2$ where

$$E_0(k_{01}^*, k_{10}^*) := k_{01}^* - \tilde{k}_{01} - k_{01}^* \cdot (1 - \exp(-\tau_c k_{10}^*)) \quad (24)$$

$$E_1(k_{01}^*, k_{10}^*) := k_{10}^* - \tilde{k}_{10} - k_{10}^* \cdot (1 - \exp(-\tau_c k_{01}^*)). \quad (25)$$

As noted earlier, this correction scheme only approximates the real rate constants. If a significant number of dwells is missed, the assumption that missed states in the augmented network cannot exchange no longer holds. These higher-order effects are particularly difficult to handle since the kinetics of missed states in the case of a sharp cutoff do not resemble a Markov process. However, simulations confirm the efficiency of the method in normal ranges of kinetics and cutoffs. As illustrating examples, [Table 1](#) shows the results of a simulation for a two-state network and [Table 2](#) shows the results for a three-state network. In each case the correction scheme was able to reverse the effects of temporal cutoffs almost completely.

Cutoff / μs	k_{01} / s^{-1} k_{10} / s^{-1}				k_{01} / s^{-1} k_{10} / s^{-1}			
	$\frac{k_{01}}{k_{01}^{\text{real}}}$	$\frac{k_{10}}{k_{10}^{\text{real}}}$	$\frac{k_{01}}{k_{01}^{\text{real}}}$	$\frac{k_{10}}{k_{10}^{\text{real}}}$	$\frac{k_{01}}{k_{01}^{\text{real}}}$	$\frac{k_{10}}{k_{10}^{\text{real}}}$	$\frac{k_{01}}{k_{01}^{\text{real}}}$	$\frac{k_{10}}{k_{10}^{\text{real}}}$
	Uncorrected rate constants				Corrected rate constants			
10	982	1997	0.98	1.00	1002	1986	1.00	0.99
50	904	1904	0.90	0.95	1000	2002	1.00	1.00
100	815	1784	0.82	0.89	993	1971	0.99	0.99
500	377	1141	0.38	0.57	963	1852	0.96	0.93
1000	223	807	0.22	0.40	919	1798	0.92	0.90

Table 1: Missed-event correction using [Equation 23](#) for a simulated two-state system. The real rate constants were $k_{01}^{\text{real}} = 1000\text{s}^{-1}$ and $k_{10}^{\text{real}} = 2000\text{s}^{-1}$.

Cutoff / μs	k_{01} k_{02} k_{10} k_{12} k_{20} k_{21}						k_{01} k_{02} k_{10} k_{12} k_{20} k_{21}					
	Uncorrected rate constants / s^{-1}						Corrected rate constants / s^{-1}					
10	1002.1	9.4	99.4	0.0	10.6	0.1	1003.2	9.4	100.4	0.0	10.7	0.0
50	990.8	9.5	95.3	0.1	10.0	0.5	995.8	9.5	100.2	0.0	10.5	0.0
100	988.5	10.0	90.6	0.1	9.8	1.1	998.5	10.1	100.2	0.0	10.9	0.1
500	945.1	9.6	59.9	0.4	5.8	3.5	993.2	9.7	98.8	0.0	9.6	0.0
1000	861.5	9.6	36.7	0.6	4.1	6.7	948.6	9.7	95.7	0.0	10.7	0.2

Table 2: Missed-event correction using [Equation 21](#) for a simulated three-state system. The real rate constants were $k_{01}^{\text{real}} = 1000\text{s}^{-1}$, $k_{02}^{\text{real}} = k_{20}^{\text{real}} = 10\text{s}^{-1}$, $k_{10}^{\text{real}} = 100\text{s}^{-1}$ and $k_{12}^{\text{real}} = k_{21}^{\text{real}} = 0$.

4.3 A TEST FOR MEMORY IN A TRAJECTORY

Several measures for quality control for the results of a Hidden Markov analysis have been introduced: first, ensure that the distributions of all

data points classified into a certain state are reasonably close to the expected PSF and, second, ensure that the distributions of lifetimes are single exponential. These two tests are necessary but not sufficient for a memoryless Markov process.

To illustrate a process involving memory, consider the example network of Figure 11A. State 0 can exchange with states 1 and 2, while the states 1 and 2 do not exchange. If the transition from 0 to 1 has a higher likelihood to occur, under the condition that a 0 to 1 transition has happened shortly before, memory is at hand. Memory effects are unexpected in a thermodynamic system at equilibrium.

For the analysis of equilibrium trajectories, we can therefore add an additional test to the two tests recapitulated above: a test for memory. If no memory is at hand, the likelihood of a transition from a state i at time $t - 1$ into another state j at time t is *only* dependent on the identity of the state at $t - 1$. As a consequence, we can make a prediction for the distribution of *consecutive transitions* from i to j .

Let p_{ij} be the probability to transition from state i into state j . The distribution of the number x of consecutive transitions from i to j is

$$f(x) = p_{ij}^{x-1} (1 - p_{ij}), \quad (26)$$

a geometric distribution. Deviations from a geometric distribution can be a consequence of memory.

The applicability of this test can be illustrated using simulations. Figure 13A shows a trajectory of a memory-less three-state network (see inset). The lifetimes of the states are chosen equal for all states and are shown in Figure 13B. Figure 13C, D show the distributions of the consecutive transitions from state 0 to 1 and 0 to 2, respectively. Both follow geometric distributions (shaded areas, Equation 26).

The same network can be simulated with a positive bias for the transition $0 \rightarrow 1$ to occur, if it has occurred previously (Figure 13E–H). While the lifetimes are still exponentially distributed (Figure 13F), the distributions of consecutive transitions (Figure 13G, H) are not geometric.

This test for geometric distributions can hence serve as a test for memory in observed trajectories and is complementary to the two other tests recapitulated above.

Note that this consideration only makes sense, if there are at least two other states that i can transition into.

$f(x)$ is the distribution of exactly x consecutive transitions, i. e. x transitions from i into j followed by a transition into another state.

4.4 THE DUMBBELL CONFIGURATION IN THERMODYNAMIC EQUILIBRIUM

A dumbbell construct in an optical tweezers instrument is in thermodynamic equilibrium on the sub-millisecond time scale, where the slowest component of the system is the relaxation of the beads [67]. The system can therefore equally be treated instead of a bead-DNA-protein-DNA-bead picture in a reduced bead-DNA-protein picture, where only one trap with stiffness $k = \left(\frac{1}{k_1} + \frac{1}{k_2}\right)^{-1}$ and only one DNA linker with length $L = L_1 + L_2$ is involved [42].

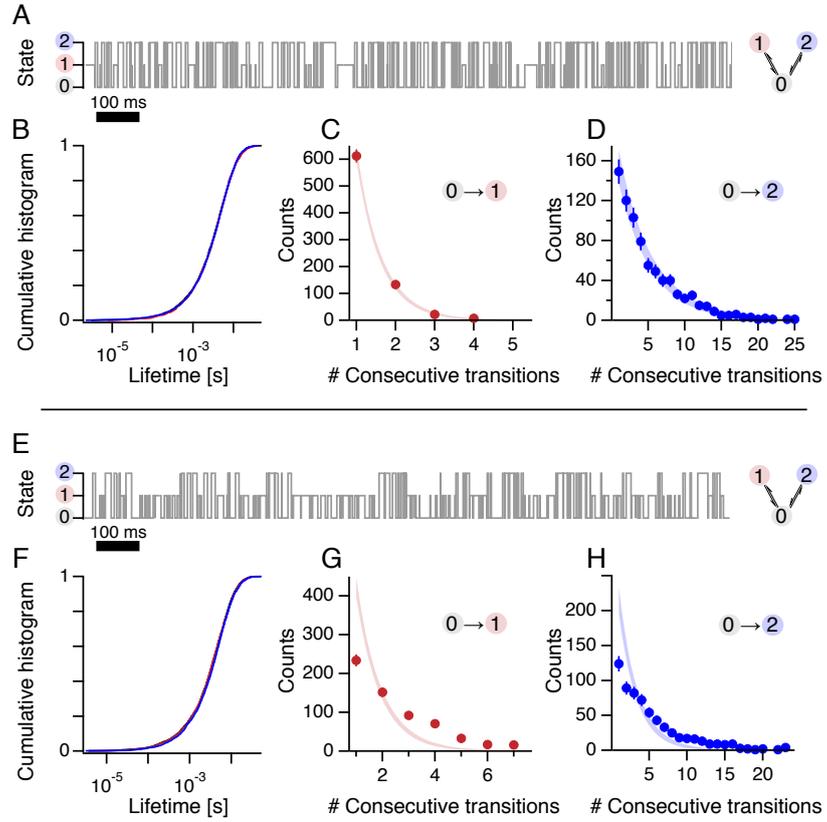


Figure 13: Test for the non-stochastic long-range behavior on simulated data. (A)–(D) Without memory. (E)–(H) With memory. (A), (E) show the simulated trajectories for the network shown in the inset. (B), (F) Integrated histograms of the lifetimes. All lifetimes are single exponentially distributed. (C), (D), (G), (H) Distributions of the number of consecutive transitions (full circles) and the prediction for a memory-less process (Equation 26, shaded area).

4.4.1 Polymer models

The force-extension behavior of polymers is mostly governed by entropy and can be described in various models. All models have their limitations and advantages. A short introduction on the various models is given here.

The Worm-like Chain (WLC) force-extension relation is derived from the entropy involved in stretching a polymer. It is hereby assumed, that the orientation of the polymer is self-correlated along its contour for short stretches. Exact solutions of the model difficult to obtain. However, the elasticity of a WLC modeled polymer is sufficiently described by the following interpolation formula [15]:

$$F_{\text{WLC}}(x) = \frac{k_{\text{B}}T}{p} \left(\frac{1}{4 \left(1 - \frac{x}{L}\right)^2} - \frac{1}{4} + \frac{x}{L} \right) \quad (27)$$

Here, L is the chain's contour length and p is its persistence length.

For DNA molecules, it was found that the extensibility does deviate slightly from a purely entropic WLC model. At high forces, contributions to the free energy of stretching due to enthalpic effects can be accounted for by including an additional stretch modulus K . This extended interpolation formula is here referred to as the Extensible Worm-like Chain (eWLC) model [103]:

$$F_{\text{eWLC}}(x) = \frac{k_B T}{p} \left(\frac{1}{4 \left(1 - \frac{x}{L} + \frac{F}{K}\right)^2} - \frac{1}{4} + \frac{x}{L} - \frac{F}{K} \right) \quad (28)$$

The interpolation formulae shown above are approximations to an exact solution, that perform reasonably well (with a relative error in force smaller than 5%). Even though corrections in terms of residual expansions have been proposed [13], in this study, Equation 27 and Equation 28 are used exclusively.

4.5 EQUILIBRIUM FREE ENERGIES

There are several possibilities to obtain equilibrium free energies in single molecule traces. At equilibrium, the determination is particularly easy. Here, the ratio of state probabilities P_i, P_j in a multi-state system is given by the Boltzmann equation

$$\Delta G_{ij} = -k_B T \ln \frac{P_j}{P_i}. \quad (29)$$

The free energy of the full dumbbell system consists of the energies of the trapped beads, the DNA linker molecule, the folded protein and the unfolded polypeptide. At a certain force F_i , when the protein is in state i , it is given by

$$G_i(F_i) = G_i^0 + G_i^{\text{device}}(F_i) = G_i^0 + G_i^{\text{beads}}(F_i) + G_i^{\text{linker}}(F_i) + G_i^{\text{p}}(F_i). \quad (30)$$

In this description, G_i^0 is the free energy of the protein in state i . Further, $G^{\text{beads}}(F) = \frac{1}{2}x(F) \cdot F$ is the (Hookean) energy stored in the displacement x of the beads from their centers, G^{linker} is the energy stored in stretching the DNA linker and G^{p} is the energy stored in stretching the unfolded polypeptide. G^{linker} and G^{p} can be calculated as integrals over eWLC (Equation 28) or WLC (Equation 27) curves, respectively:

$$G^{\text{linker}}(F) = \int_0^{x_{\text{eWLC}}(F)} F_{\text{eWLC}}(x') dx' \quad (31)$$

Whenever in an experiment the protein transitions from a state i to another state j , the force changes from F_i to F_j . The resulting change in free energy is then

$$\Delta G_{ij}(F_i, F_j) = G_j(F_j) - G_i(F_i) = \Delta G_{ij}^0 + \Delta G_{ij}^{\text{device}}(F_i, F_j). \quad (32)$$

Note that in the case of a dumbbell

$$x = x_1 + x_2 = F \cdot \left(\frac{1}{k_1} + \frac{1}{k_2} \right).$$

Consequently, by using the Boltzmann relation (Equation 29), the probability to find the system in state i is given by

$$P_i(F) = \frac{1}{1 + \sum_{j \neq i} \exp\left(-\frac{\Delta G_{ij}^0 + \Delta G_{ij}^{\text{device}}(F_i, F_j)}{k_B T}\right)}. \quad (33)$$

The probabilities are solely determined by the parameters of the linker and the differences in free energy of the various states of the protein. Equation 33 can be used as a fit function that allows the determination of the values of $-\Delta G_{ij}^0$ from the equilibrium state occupancies.

4.6 MODELS FOR THE FORCE DEPENDENCE OF TRANSITION RATE CONSTANTS

Since force in the described assay functions as a denaturant, the transition rate constants crucially depend on the applied force. In order to extract information about the folding mechanism and determine force-free folding and unfolding rate constants, models are needed. Here I will shortly describe the two models used in this study.

4.6.1 The Bell Model

The Bell model was introduced to phenomenologically describe the dissociation kinetics of ligand-receptor systems [7, 38]. In this simple model, it is assumed that the one dimensional energy landscape along the coordinate x is tilted by a potential $-F \cdot \Delta x$, when an external force F is applied (Figure 14).

In the case of unfolding, the rate constants are given by

$$k_u(F) = k_0 \exp\left(\frac{F \cdot \Delta x}{k_B T}\right), \quad (34)$$

where Δx is the distance to the transition state of unfolding and k_0 is the unfolding rate at zero force. For folding, the model can be equally applied, now with a negative Δx .

This model neglects all nonlinear contributions of the linkers and the unfolded polypeptide and also neglects a possible force dependence of Δx and k_0 . However, in the measurable range, this simple model describes the data well. Nevertheless, the negligence of nonlinearities make the Bell model unsuitable for the extrapolation of rate constants outside of the experimentally accessible range.

It should be noted that the value Δx describes the stretching that is applied to the *whole system*, not just the protein alone. In other words, if an analysis results in a certain Δx value for a protein (which sometimes is bigger than the protein dimensions themselves), this does not mean, that the protein alone is stretched this far, but the entire system. The protein can (and probably will) be stretched much less.

Note that Δx in this formula is the direct measured distance the system needs to be stretched to reach the transition state, instead of an unfolding length, that is usually given as a contour length.

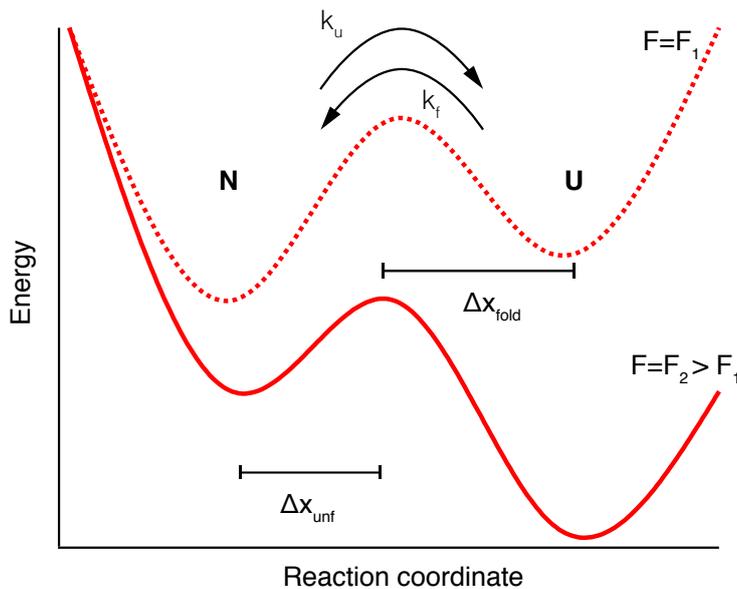


Figure 14: The effect of an external force on an energy landscape. Shown is a hypothetical two-state energy landscape at a low force (dotted) and at a high force (continuous). Upon increasing the force, the basin for the unfolded state U lies lower than the folded state N. The response of the folding and unfolding rate constants k_f and k_u as well as the distances to the transition state Δx_{fold} and Δx_{unf} is model dependent.

A more elaborate model that is based on a one-dimensional Kramers theory was developed by [Dudko et al.](#). This model includes assumptions about the specific shape of the potential well and has been solved for a harmonic barrier and a cusp-like barrier [33, 34]. Since in our experiments the unfolding rate plots do not show any pronounced curvature, the potential shape parameter ν in Dudko's formula is hard to determine. To keep the number of free parameters low, we fixed $\nu = 1$, where Dudko's model reduces to [Equation 34](#).

4.6.2 A model for folding

While the Bell model can in principle also be used to describe folding, better approaches to describe folding exist. While the elasticity of a folded protein is elusive, the elasticity of an unfolded polypeptide can be well described with a [WLC](#) model. [Gebhardt et al.](#) therefore adapted a model for folding based on an earlier model by [Schlierf et al.](#) for applications in optical tweezers [43, 90].

The model uses the information about the energies that is stored in the unfolded polypeptide chains and assumes that the unfolded polypeptide chain needs to collapse to a certain effective length in order to reach the transition state and trigger folding. During this contraction, all compo-

nents are assumed to be in thermal equilibrium. Essentially all energetic contributions during the contraction from the initial state i and the transition state T are considered:

$$k_f(F) = k_0 \exp\left(-\frac{\Delta G_{iT}(F_i = F, F_T)}{k_B T}\right). \quad (35)$$

Here, the difference in energy between the initial state and the transition state ΔG_{iT} is given by Equation 32. Since most parameters, such as the contour lengths of the folded and unfolded states, are known, so are the forces involved. The only free parameters are the zero-force rate constant k_0 and $\Delta L = L_i - L_T$, the length the unfolded polypeptide has to contract in order to reach the transition state. This length is usually measured in contour length.

4.6.3 Folding and unfolding rate models and the applicability to equilibrium data

Equilibrium processes necessarily follow the principle of *detailed balance*. Generally, the equilibrium between states A and B can be described by

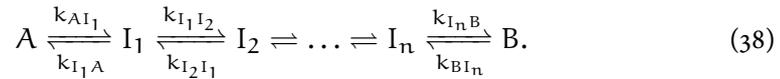


The process is in equilibrium, when

$$P_A k_{AB} = P_B k_{BA}, \quad (37)$$

where P_A and P_B are the equilibrium probabilities to find the system in state A or B , respectively.

A notable consequence of this principle is the following: If we decompose the two-state equilibrium network $A \rightleftharpoons B$ into several substeps by inserting the intermediates I_n , each of the subreactions also has to be in equilibrium:



A further important corollary of this principle is the fact that there is no net “flux” at equilibrium. A circular process where a series of states is visited one after another is therefore impossible under equilibrium conditions.

The principle of detailed balance can be directly applied to single molecule optical tweezers data. If the states i and j are in equilibrium with a free energy difference ΔG_{ij} and only the transition rate k_{ij} is known, Equation 37 can be used to calculate the reverse transition rate k_{ji} :

$$k_{ji} = \exp(-\Delta G_{ij}/k_B T) k_{ij}. \quad (39)$$

Hence, technically, it is redundant to model both the folding and unfolding rate constants for a transition when the equilibrium free energy

between the two states is known. Even more, whenever an equilibrium process is modeled, one is technically not free to independently choose the models for folding and unfolding at will. Following the principle of detailed balance (especially its corollary [Equation 39](#)), the model for the reverse process of a transition is already predetermined, if the model for the forward process has been chosen.

Nevertheless, in this work, the folding model of [Equation 35](#) was used exclusively for folding, while the Bell model [Equation 34](#) was used for unfolding. I will show in [Section 7.7.2](#) that the extrapolations for the zero-force energies obtained by fitting both models independently yields the same zero-force energies as were obtained from the equilibrium probabilities. Hence, the errors made by independently choosing two different models that are technically incompatible with detailed balance, are small in this case.

4.7 NON-EQUILIBRIUM METHODS

Under certain conditions, for example, when the kinetics of folding or unfolding are slow, the measurement of equilibrium traces can be experimentally challenging. Nevertheless, a wealth of information can be obtained also from force-ramp experiments, where the force is periodically increased and decreased and the system is intentionally steered away from equilibrium. A few of the methods are mentioned in [Appendix B](#).

4.8 SUMMARY

This chapter introduced how Hidden Markov models can be used to analyze trajectories in single molecule force spectroscopy. The method is versatile and robust and allows to easily extract kinetic and energetic information about the system.

Further, this chapter presented a correction method for missed events, that primarily appears in proteins with fast kinetics. Corrections of this sort were important for the analysis of some of the constructs in this work (see, e. g. [Chapter 7](#)).

Third, since the sample protein in typical optical tweezers force spectroscopy measurements is probed using a system of beads and linker molecules, this chapter also illustrated how to separate the kinetic and energetic information of the protein from the pulling apparatus.

5

FOLDING OF LIGAND BINDING PROTEINS

As discussed in [Chapter 2](#), calmodulin is able to bind two calcium ions in each of its domains. Following simple thermodynamic arguments it can be easily understood that proteins that can bind ligands also behave differently dependent on their ligation state. This chapter gives a short introduction into the models of ligand-binding proteins that can bind two identical ligands, and the consequences on the energetics and kinetics of folding and unfolding.

5.1 MICROSCOPIC AND MACROSCOPIC BINDING

Consider a molecule with two binding sites (here called site I and site II) that can bind identical ligands. Since each binding site can exist in its ligated and unligated form, there are four different possible configurations for ligation ([Figure 15](#)).

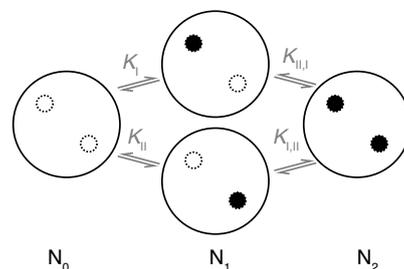


Figure 15: Microscopic model for two binding sites. The protein can exist in four different configurations: Unligated (N_0), fully ligated (N_2) and either of the two binding sites bound, the other one unbound (N_1).

The equilibrium dissociation constants for the binding to site I and site II, if the respective other binding site is unoccupied, are called K_I and K_{II} , respectively. The corresponding constants, if the other site is occupied, are named $K_{I,II}$ and $K_{II,I}$ [62, 108]. These constants are called the *microscopic* dissociation constants. However, if K_I is very similar to K_{II} , it is experimentally challenging to determine for a state N_1 with only one binding site occupied, which site is bound and which is unbound.

Therefore, a simpler description of the binding is given in terms of the *macroscopic* dissociation constants K_1 , K_2 . The microscopic and macro-

Here, the macroscopic constants carry arabic indices, microscopic constants carry roman indices.

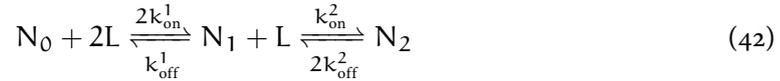
scopic constants are related according to

$$K_1^{-1} = K_I^{-1} + K_{II}^{-1} \quad (40)$$

$$K_1 K_2 = K_I K_{II,I} = K_{II} K_{I,II} \quad (41)$$

Note that the description using macroscopic binding constants does not contain information about the possible intrinsic differences between the binding sites.

The network of states can be written using the on- and off-rates k_{on} and k_{off} :



Here, again, N_i indicates that the protein has i ligands bound. The factors of two appear since there are two possible configurations for N_1 . The macroscopic equilibrium constants in this case are

$$K_1 = \frac{k_{off}^1}{2k_{on}^1} \quad (43)$$

$$K_2 = \frac{2k_{off}^2}{k_{on}^2} \quad (44)$$

5.2 COOPERATIVITY

An intriguing feature of many ligand binding proteins with multiple binding sites is called cooperativity, i. e. an increase of the affinity of one binding site once the other site is occupied. In terms of dissociation constants, the binding is cooperative whenever $K_{I,II} < K_I$ and $K_{II,I} < K_{II}$. In a description using macroscopic constants, cooperativity is at hand, if $K_2 < 4K_1$. This equation is assuming that the microscopic apo dissociation constants K_I and K_{II} are equal. If they differ, cooperativity can even occur if $K_2 > 4K_1$.

An interesting consequence of binding cooperativity can be found in the number of ions bound. As discussed, the binding event of one ligand increases the affinity for the second. Consequently, the protein is unlikely to be found in a state with only one ion bound and predominantly found either unligated or fully ligated.

5.3 THE INFLUENCE OF LIGAND BINDING ON THE FOLDING AND UNFOLDING

The folding and unfolding kinetics of ligand binding proteins can be vastly different between their bound (holo) and free (apo) conformations. In the following, I will introduce a model for the equilibrium folding and

unfolding mechanism of a single domain with two identical binding sites [10, 40, 51].

Figure 16 shows the model of a single domain with two binding sites, that can exist in its native conformation with zero, one or two ligands bound (N_0, N_1, N_2 , respectively). Further, also the unfolded state is assumed to exist in these three conformations (U_0, U_1, U_2). In addition, also the transition state between the folded and unfolded states can bind up to two ligands (TS_0, TS_1, TS_2). A transition across any of the transition states TS_n constitutes a folding or unfolding event.

Note that in the optical tweezers assay there is no macroscopic information about which binding site is occupied and which is not. Hence, the two conformations for U_1 and N_1 are not distinguishable.

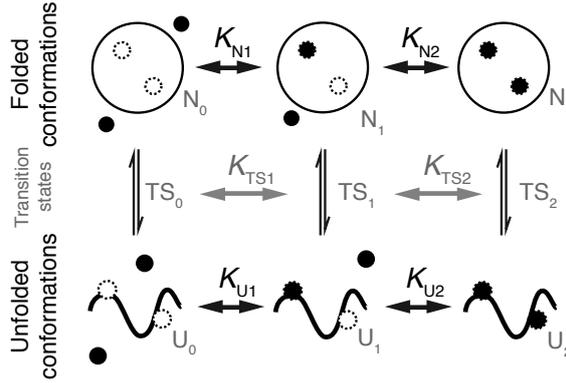


Figure 16: Model for the folding of a protein with two identical binding sites. Note that even though there are two possible conformations for N_1 and U_1 , only one is drawn.

The differences in energy between the ligand bound and ligand free conformations of the native state N are purely determined by their dissociation constants:

$$\begin{aligned} K_{N1} &= \frac{[N_0][L]}{[N_1]} = \exp\left(\frac{\Delta G_{N1-0}}{k_B T}\right) \\ K_{N2} &= \frac{[N_1][L]}{[N_2]} = \exp\left(\frac{\Delta G_{N2-1}}{k_B T}\right) \end{aligned} \quad (45)$$

where $\Delta G_{N_{i-j}}$ is the difference in energy between the conformations with i and j ligands bound. The equilibrium constants for the transition state and the unfolded state K_{TS1} , K_{TS2} , K_{U1} and K_{U2} are defined accordingly.

In equilibrium terms, the probability to find the folded state with i ligands bound is

$$P(N_i) = \frac{[N_i]}{\sum_j [N_j]} \quad (46)$$

Using Equation 45, this yields

$$\begin{aligned} P(N_0) &= \frac{1}{1 + \frac{[L]}{K_{N1}} + \frac{[L]^2}{K_{N1}K_{N2}}} \\ P(N_1) &= \frac{1}{1 + \frac{K_{N1}}{[L]} + \frac{[L]}{K_{N2}}} \\ P(N_2) &= \frac{1}{1 + \frac{K_{N2}}{[L]} + \frac{K_{N1}K_{N2}}{[L]^2}} \end{aligned} \quad (47)$$

Under the assumption that the barrier crossing time is short compared to the exchange of ligands, the unfolding rate is given by the total of the rates reaching from the top to the bottom part in Figure 16:

$$k_u = P(N_0)k_u^0 + P(N_1)k_u^1 + P(N_2)k_u^2 \quad (48)$$

Note that k_u^0 is the apo unfolding rate. Due to definition, all other rates are given at a ligand concentration of 1M.

Here, k_u^i is the intrinsic unfolding rate from a state with i ligands bound. These rates are given by Arrhenius equations:

$$\begin{aligned} k_u^1 &= k_u^0 \exp\left(-\frac{\Delta G_{N1-0} - \Delta G_{TS1-0}}{k_B T}\right) = k_u^0 \frac{K_{N1}}{K_{TS1}} \\ k_u^2 &= k_u^1 \exp\left(-\frac{\Delta G_{N2-1} - \Delta G_{TS2-1}}{k_B T}\right) = k_u^0 \frac{K_{N1}K_{N2}}{K_{TS1}K_{TS2}} \end{aligned} \quad (49)$$

where the last equation is a result of Equation 45.

Combining Equation 47, Equation 48 and Equation 49, we obtain for the effective unfolding rate

$$k_u = k_u^0 \frac{1 + \frac{[L]}{K_{TS1}} + \frac{[L]^2}{K_{TS1}K_{TS2}}}{1 + \frac{[L]}{K_{N1}} + \frac{[L]^2}{K_{N1}K_{N2}}} \quad (50)$$

Similarly, the effective folding rate is

$$k_f = k_f^0 \frac{1 + \frac{[L]}{K_{TS1}} + \frac{[L]^2}{K_{TS1}K_{TS2}}}{1 + \frac{[L]}{K_{U1}} + \frac{[L]^2}{K_{U1}K_{U2}}} \quad (51)$$

Notably, the folding rate is solely dependent on the macroscopic dissociation constants of the unfolded state and the transition state as well as the unligated (apo-) unfolding rate. The unfolding rate only depends on the apo folding rate and the dissociation constants of the folded and transition state.

5.3.1 Example: Force-dependent folding and unfolding rates for a domain with two calcium binding sites

Figure 17 shows folding and unfolding rates according to Equation 50 and Equation 51 for a domain with two binding sites (A, B) and a domain

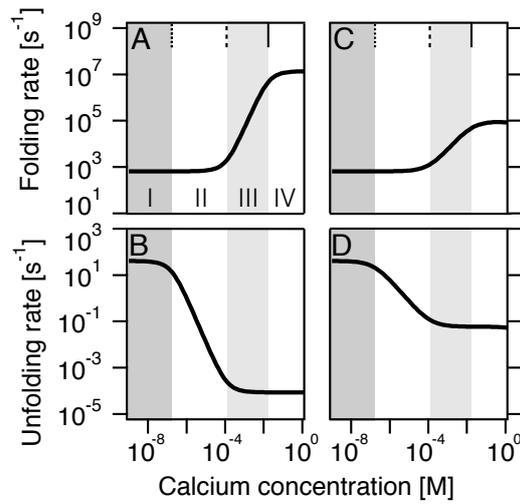


Figure 17: Theoretical folding and unfolding rates of a ligand binding protein. (A), (B) Ligand dependent transition rates for a domain with two binding sites and $K_1 = K_2$. (C), (D) Ligand dependent transition rates for a domain with one binding site. The value of K_N is indicated with a dotted line, K_{TS} with a dashed line, K_U with a continuous line.

with one binding site (C, D). For the domain with two binding sites, it was assumed that $K_1 = K_2$.

Four different concentration regions are apparent. In region I, the calcium concentration is smaller than all dissociation constants. Hence, the folding rate is equal to the apo folding rate and the unfolding rate is equal to the apo unfolding rate. In the picture of Figure 16, the transition occurs between the states $N_0 \rightleftharpoons U_0$ and the rates are the intrinsic apo unfolding and folding rates k_u^0 and k_f^0 .

In region II, $K_N < [Ca^{2+}] < K_{TS}$. The folding rates depend on the difference in free energy between the unfolded state U and the transition state TS. Since $[Ca^{2+}] < K_{TS} < K_U$, the folding rates are not influenced by the calcium concentration in region II. However, the unfolding rates, which depend on the difference in free energy between the native state N and the transition state TS, do change in this region. In the case of a domain with two binding sites (Figure 17A, B), the slope of the change is twice the slope of a domain with only one binding site (Figure 17C, D).

Region III, where $K_N < K_{TS} < [Ca^{2+}] < K_U$, shows the complementary effect. Since the transition state is now saturated with calcium, the folding rates, which depend on the difference in free energy between TS and U change. The unfolding rates become independent of calcium concentration. Again, the effect is more pronounced for the domain with two binding sites compared to the domain with one binding site.

In region IV, all states are saturated with calcium and neither the folding nor the unfolding rates depend on $[Ca^{2+}]$ any more.

Taken together, the stoichiometry of binding is directly reflected in the slopes of the folding and unfolding rates. This will help explain the folding and unfolding pathways of the single domains of calmodulin presented in [Chapter 7](#).

5.4 STABILIZATION BY LIGAND BINDING

The reason for changing transition rates is a selective stabilization of the folded, unfolded and transition states. The overall free energy difference between the unbound and bound state under ligand binding conditions is given by (see [Equation 47](#))

$$-\frac{\Delta G_0}{k_B T} = -\frac{\Delta G_0^{[L]=0}}{k_B T} - \ln \frac{1 + \frac{[L]}{K_{N1}} + \frac{[L]^2}{K_{N1}K_{N2}}}{1 + \frac{[L]}{K_{U1}} + \frac{[L]^2}{K_{U1}K_{U2}}}. \quad (52)$$

Proteins with cooperative binding, such as calmodulin, are predominantly calcium-free or ligated with two calcium ions. Under normal conditions, the unfolded state does not bind calcium. For such systems, [Equation 52](#) can be approximated by

$$-\frac{\Delta G_0}{k_B T} \approx -\frac{\Delta G_0^{[L]=0}}{k_B T} - \ln \frac{[L]^2}{K_{N1}K_{N2}}. \quad (53)$$

At calcium concentrations less than $\sqrt{K_{N1}K_{N2}}$, the free energy is independent of the concentration and equal to the free energy of the apo state, $-\Delta G_0^{[L]=0}$. For higher calcium concentrations, the free energy increases by a factor of $2 \cdot \ln(10)$ for every ten-fold increase in concentration. This “slope of two” behavior is a direct consequence of the presence of two binding sites.

Part II
RESULTS

6

THE COMPLEX FOLDING OF CALMODULIN

This chapter describes the results of N-C-terminal optical tweezers pulling experiments on single calmodulin molecules at high non-physiological calcium concentrations of 10 mM. We found that due to interactions between the domains, multiple productive folding pathways compete with pathways that lead into kinetic traps and slow down folding.

6.1 A COMPLEX NETWORK OF STATES

Figure 18A shows a cartoon representation of a construct where full-length calmodulin was inserted between two terminal ubiquitin molecules that served as spacers. Using dsDNA handles, this protein was attached to micron-sized glass beads that could be trapped in a dual beam optical tweezers setup.

Several examples for the N-C-terminal pulling cycles of full-length calmodulin are shown in Figure 18A. Black lines show stretching, blue lines relaxation. The first part of the curves is dominated by the elastic properties of the dsDNA linkers. The protein is still fully folded in this case. Upon reaching a certain force level, a part of the structure unfolds, followed by a second unfolding. The refolding occurs in the same manner. By measuring the increase in contour length, an unfolding of two EF hands can be assigned to the first event and the unfolding of two more EF hands to the second event. Under the assumption that the two domains of calmodulin fold independently, these events can thus be assigned to the independent folding and unfolding of the N- and C-terminal domains.

However, the upper trace, shown in Figure 18B, shows deviations from the previously described behavior. Here, rapid fluctuations between a level roughly corresponding to three folded EF hands and an intermediate with two EF hands folded can be observed. In typical stretch-and-relax measurements, this pattern reoccurs frequently.

More detailed information about the folding/unfolding behavior can be obtained by setting the trap centers to a fixed position and thus applying a force bias to the molecule. Given the right settings, the molecule can unfold and refold frequently. In this configuration, the forces fluctuate upon each un- and refolding event. The high stability of the instrument allows the measurement of single molecules for time scales on the order of tens of minutes. Figure 19 shows a five minute excerpt of a 45 minute experiment of the fluctuations of full-length calmodulin at an intermediate biasing force.

The biasing force is defined as the mean force of the green (F_{34}) level.

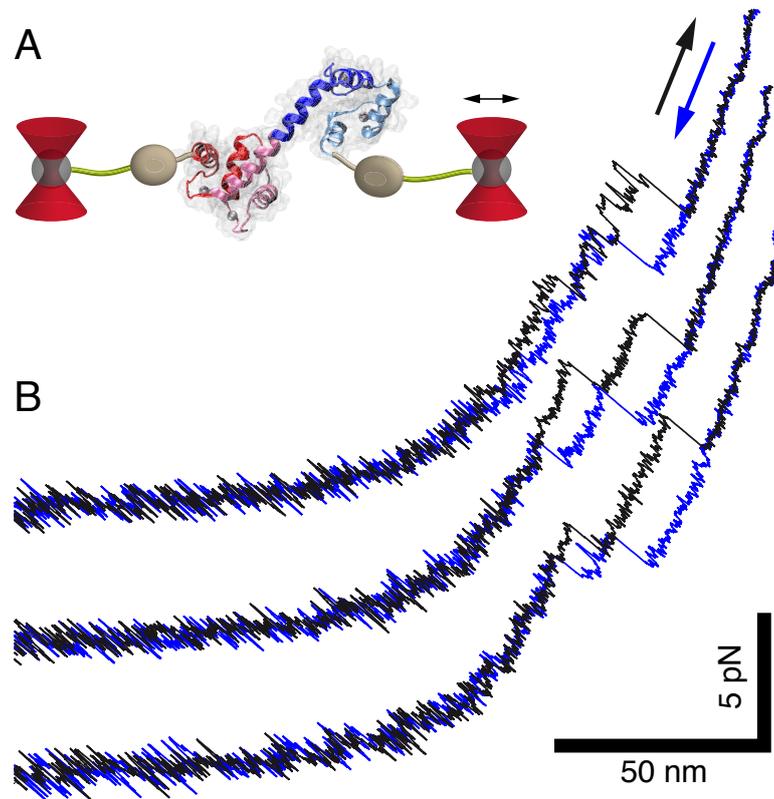


Figure 18: (A) Full-length calmodulin in the dumbbell pulling geometry. The protein is bound with ubiquitin (brown) and DNA (green) spacers to functionalized silica beads. (B) Repeated stretch-and-relax cycles for full-length calmodulin. Plotted is the force vs extension curve for the repeated unfolding and refolding of a single calmodulin molecule.

The expansion in [Figure 19](#) reveals a complex pattern of states that is clearly different from an apparent three-state behavior that would be expected from the independent folding model of the N- and C-terminal domains. We applied a Hidden Markov model (see [Section 4.1](#)) to the data, which classifies data points into states, optimizing for single exponentiality of life-time distributions. The analysis revealed the presence of at least six different states. In addition to the folded state (purple, called henceforth F_{1234}) and unfolded state (red, called U), three states cluster at a length corresponding to two folded EF hands (green, light blue, orange). Another state, corresponding to the length of three folded EF hands (dark blue) exchanges quickly with the light blue state.

We devised a series of mutants to identify the states. The following section will describe the detailed reasoning.

For independent and indistinguishable N- and C-terminal domains, the states would be folded, unfolded and one domain folded.

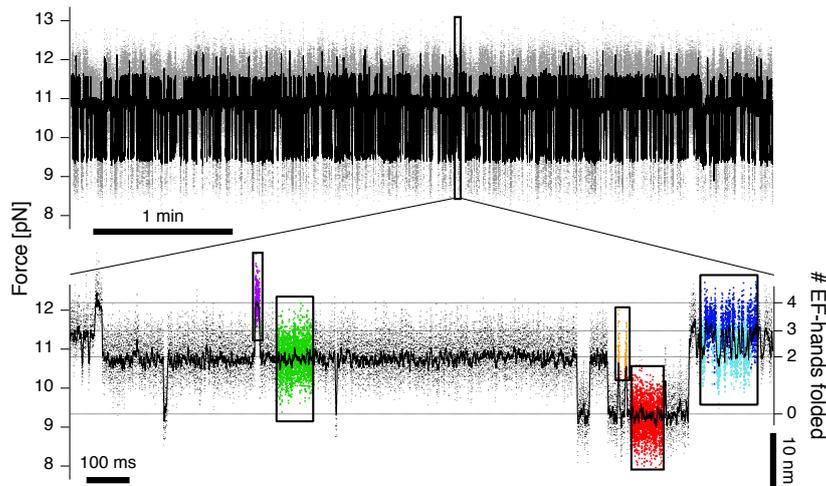


Figure 19: Equilibrium fluctuations of full-length calmodulin. Shown is a time-trace of the folding and unfolding behavior of full-length calmodulin at an intermediate force bias, where all states are populated (upper graph). Grey: Full bandwidth data, black: smoothed data. The expansion reveals the complex behavior of transitions. Boxes in the lower graph indicate regions of interest with colors assigned by a Hidden Markov classifier. Red: Unfolded (U), Purple: Folded (F_{1234}). Green: C-domain folded (F_{34}), Light blue: N-domain folded (F_{12}), Dark blue: EF hand 4 unfolded (F_{123}), Orange: non-native pairing of EF hands 2 and 3 (F_{23}). The identification of states is explained in [Section 6.2](#).

6.2 STATE IDENTIFICATION: TRUNCATION MUTANTS

To identify the states involved in the folding of full-length calmodulin, we created a series of truncation mutants where we successively removed single EF hands. A comparison between traces is shown in [Figure 20](#).

The second trace in [Figure 20](#) shows a representative trace for CaM_{234} , a mutant where the first EF hand of full-length calmodulin was deleted. In contrast to full-length calmodulin, only three states can be observed. At a length of two folded EF hands, the green state with comparable lifetimes to the full-length construct, clusters together with the short-lived orange state. As expected, at a length of four folded EF hands, the truncation mutant does not show any signal. Following the assumption that the terminal domains of calmodulin can attain very stable folds, we assign the green state to a structure, where the C-terminal domain, containing EF hands three and four, is folded, while the N-terminal domain is unfolded. We call this state F_{34} .

Additional evidence for the assignment comes from a truncation mutant where the fourth EF hand of full-length calmodulin was deleted (CaM_{123} , shown as the third trace in [Figure 20](#)). Corroborating our pre-

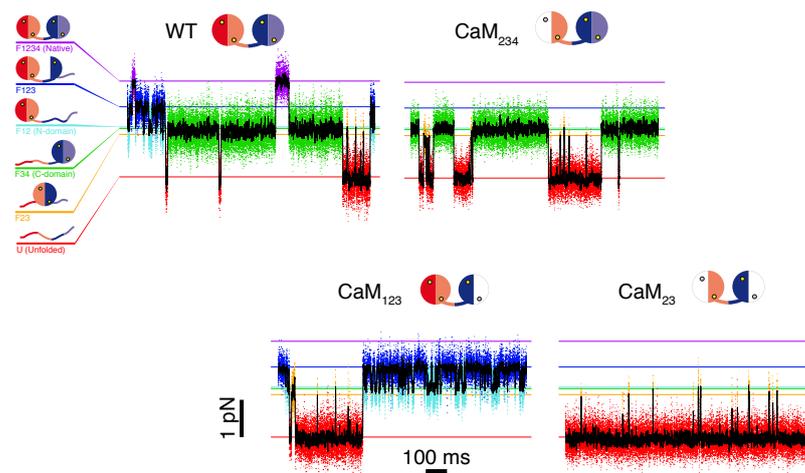


Figure 20: Comparison of traces of full-length calmodulin with truncation mutants at comparable biasing forces, where all states are populated.

vious assignment of the green state to F_{34} , this construct lacks a state at two folded EF hands compatible with a F_{34} lifetime. However, at an extension corresponding to two folded EF hands, we find the light blue state, which we therefore assign to the folded N-terminal domain and call F_{12} . This state quickly exchanges with a state with an extension of three folded EF hands (dark blue). We therefore assign the dark blue state to a structure where the first, second and third EF hands fold into a metastable intermediate (F_{123}).

The identity of the orange state becomes clear in the light of a truncation mutant that only consist of the second and third EF hand, CaM_{23} . Previous studies have indicated that such an even-odd EF hand pair can fold into a domain-like structure, albeit with low thermodynamic stability [58]. We find this state at a slightly lower extension than two folded EF hands in the trace of CaM_{23} (Figure 20, right trace), as well as in all constructs containing both the EF hands two and three, namely the traces for full-length, calmodulin, CaM_{234} and CaM_{123} . In contrast, we do not find this state in the isolated N-terminal domain CaM_{12} or the isolated C-terminal domain CaM_{34} (For traces of the isolated domains, cf. Chapter 7). The assignment of states is summarized on the left side of Figure 20.

6.3 FURTHER EVIDENCE FROM A CROSSLINKED MUTANT

A mutant where an internal crosslink was introduced by replacing residues 128 and 144 with cysteines ($\text{CaM}_{128/144}$, cf. Section D.3) was the main scope of earlier single molecule AFM studies in our group [51–53].

In the optical tweezers assay, this construct exhibits several levels that are shifted in length (and hence, force), as compared to full-length calmodulin.

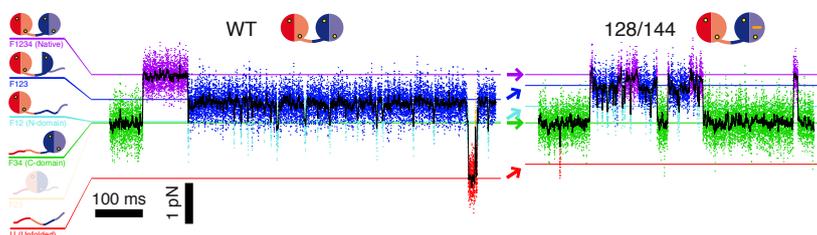


Figure 21: Comparison of traces of wildtype calmodulin and the crosslinked mutant CaM_{128/144}. Several levels (F₁₂₃, dark blue, F₁₂, light blue and U, red) appear shifted. F₂₃ was only rarely populated in CaM_{128/144} and is not labeled.

Figure 21 shows a comparison of traces for full-length calmodulin (left) and CaM_{128/144} (right), aligned at the folded extension, at comparable biasing force. The extension of the green state F₃₄, which corresponds to the unfolding of the N-terminal domain, is unchanged. In contrast, all states that involve the unfolded conformation of the crosslink-containing EF hand four (F₁₂₃, F₁₂ and U) are shifted in the crosslinked mutant, compared to the wildtype. Since the lifetimes of the unfolded state are comparably low in CaM_{128/144}, the state F₂₃ is only very rarely populated. The shift in length corresponds to the value that is expected for the crosslink (see Table 3).

Note that for a non-negligible number of measured molecules of CaM_{128/144}, the crosslink was not formed. These molecules showed wildtype behavior in the assay.

STATE	CaM _{128/144}		CaM WT	
	ΔL [nm]	ΔL_{calc} [nm]	ΔL [nm]	ΔL_{calc} [nm]
U	44.6 ± 0.5	44.8	52.2 ± 0.6	50.6
F ₂₃	—	21.6	27.4 ± 0.7	27.4
F ₃₄	23.3 ± 0.7	25.7	23.8 ± 0.5	25.7
F ₁₂	16.5 ± 0.7	19.5	23.3 ± 0.4	25.3
F ₁₂₃	5.5 ± 0.5	7.4	13.0 ± 0.3	13.2
F ₁₂₃₄	0	0	0	0

Table 3: Contour length increases for CaM_{128/144} compared with the wildtype. The expected difference in contour length between F₁₂₃₄ and the states U, F₂₃, F₁₂ and F₁₂₃ is 5.8 nm. The calculated contour length increase for the F₁₂ and F₃₄ states of full-length calmodulin are for an assumed symmetric splitting between the two domains. State F₂₃ was only rarely observed in CaM_{128/144} and was not included in the analysis.

6.4 SHIFTING THE EQUILIBRIUM

Using force as a control parameter, we can shift the equilibrium from the predominantly folded to the predominantly unfolded regime. [Figure 22A](#) shows traces of full-length calmodulin at various biasing forces. At low bias, the system spends most of its time in the folded (purple) state, while with increasing tension more and more intermediates get populated. At high bias, the system spends most of its time in the unfolded (red) state. Also note the histograms on the left of [Figure 22A](#).

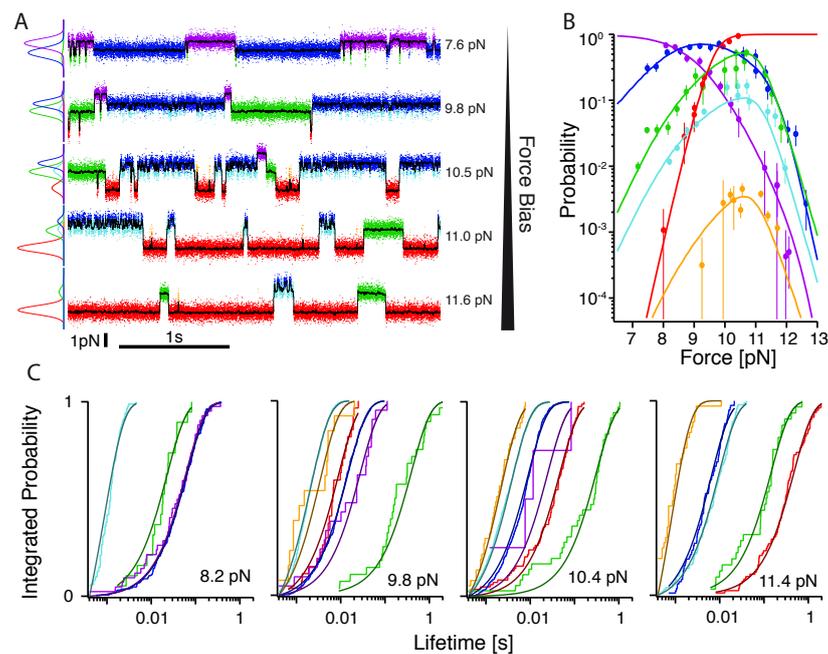


Figure 22: Shifting the equilibrium in full-length calmodulin. **(A)** sample traces at increasing force bias on the protein. The equilibrium shifts from predominantly folded to predominantly unfolded. **(B)** State occupancy probabilities at various forces. Shown is the measured equilibrium probability to be in a certain state at a given force. Fits are according to [Equation 33](#). **(C)** Integrated lifetime histograms at various biasing forces. Fits (darker shades) are single exponentials.

Using the information contained in such traces, we can extract equilibrium and kinetic information.

6.4.1 State occupancies

Note that in this representation, the probabilities do not add up to one, since the force at a certain force bias is not constant.

[Figure 22B](#) shows the measured values for the probability for the system to be in a certain state versus the mean force of the state. The data again illustrates the effects of changing the force bias. At low forces, the system is predominantly folded (purple). With increasing force, the probability to find the system in the folded state decreases, while the probabilities

for the intermediates increase. At high forces, the intermediates again die out and the system is most likely found in the unfolded (red) state.

Since at equilibrium, the population of the states is solely determined by thermodynamic parameters, we can use the probability data at hand to determine the relative equilibrium free energies of the states (Equation 33). The data for full-length calmodulin and the truncation mutants are summarized in Table 4.

STATE	CAM WT	CaM ₁₂	CaM ₃₄	CaM ₂₃	CaM ₁₂₃	CaM ₂₃₄
F ₂₃	13	–	–	12	13	13
F ₃₄	21	–	18	–	–	20
F ₁₂	20	19	–	–	20	–
F ₁₂₃	30	–	–	–	28	–
F ₁₂₃₄	36	–	–	–	–	–

Table 4: Equilibrium free energies $-\Delta G_0$ of full-length calmodulin and the truncation mutants at 10 mM Ca^{2+} . Energies are given in units of $k_B T$ as difference from state U. If a state is not present in a mutant, this is indicated with a dash.

6.4.2 Dwell times and rate constants

We can use the information contained in these traces to also extract kinetic information. After level assignment, we can track the dwell times in certain levels in the traces as shown in Figure 22A. If the system performs a memoryless process of hopping between the states, they are expected to be single exponentially distributed. Figure 22C shows the integrated and normalized lifetime distributions at various biasing forces along with single exponential fits. The data are well described by exponential lifetime distributions. As expected, the mean lifetimes shift as a function of force.

The exit-rate, i. e. the rate constant with which a state decays into another state, is given by the state’s inverse mean dwelltime. Since we are able to track the fate of decay, we can extract the particular transition rates between each of the states. Details can be found in Section 4.2. Figure 23 shows the force-dependent transition rate constants (“Chevron plot”) for each of the six possible transitions of full-length calmodulin at 10 mM Ca^{2+} .

The data in Figure 23 can be described with established models. For unfolding, a phenomenological and sufficiently simple model is the so-called Bell model (Equation 34), which assumes that the rate constants vary exponentially with the applied force. In a logarithmic plot, this relationship appears as a straight line. It should be noted that more elaborate models are available for unfolding that allow to capture details of the involved energy landscape (see Section 4.6). However, since we do not

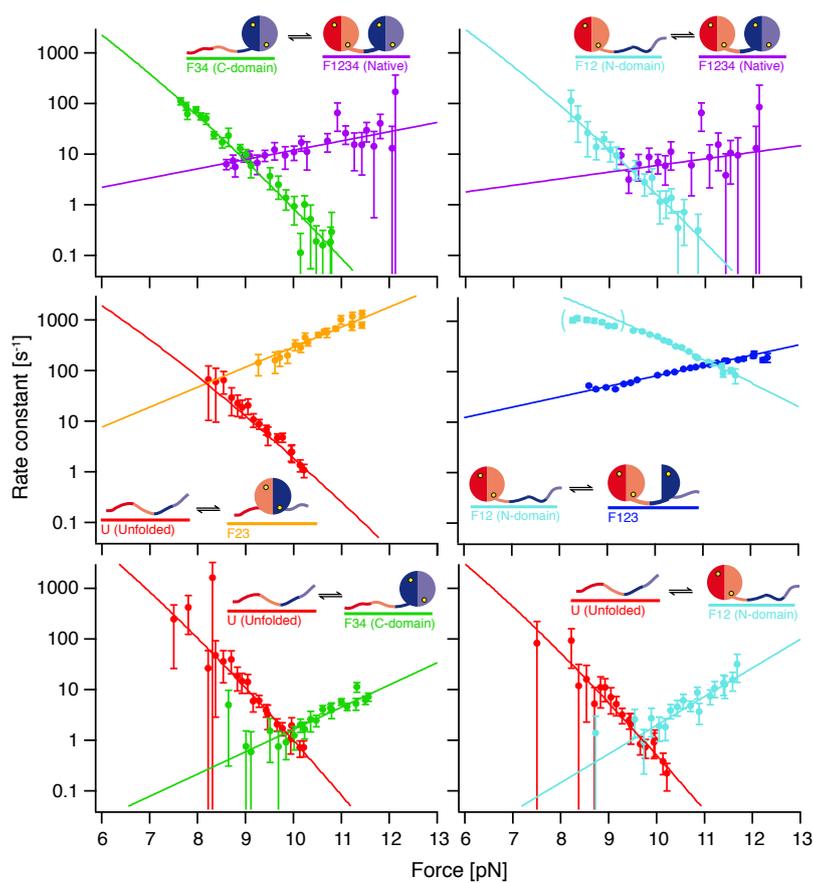


Figure 23: Force-dependent transition rate constants of full-length calmodulin at 10 mM Ca^{2+} . Continuous lines are fits to folding/unfolding models (see Section 4.6).

For a more detailed discussion concerning the choice of models, see Section 4.6.

observe a significant curvature in our plots, we cannot determine any of the additional free parameters those models involve.

For folding, we employed a model that includes the energetic contributions for stretching the linker and displacing the beads from their positions upon a folding transition [43, 90]. In their logarithmic representation, the modeled force-dependent rates show a curved dependence on the force in this model. The curvature in this case is pre-determined by the mechanical properties of the linkers. Table 5 shows the obtained parameters for the distances to the transition states of folding and unfolding, as well as the extrapolated folding/unfolding rate constants at zero force (see Section 4.6).

TRANSITION	$\log_{10} k_0^{\text{unf}} [\text{s}^{-1}]$	$\log_{10} k_0^{\text{fold}} [\text{s}^{-1}]$	$\Delta x_{\text{unf}} [\text{nm}]$	$\Delta L_{\text{fold}} [\text{nm}]$
$F_{1234} \rightleftharpoons F_{12}$	-0.7 ± 0.9	5.0 ± 0.4	1.3 ± 0.8	14.4 ± 1.3
$F_{1234} \rightleftharpoons F_{34}$	-0.8 ± 0.3	5.6 ± 0.2	1.7 ± 0.3	16.9 ± 0.8
$F_{123} \rightleftharpoons F_{12}$	-0.13 ± 0.04	5.0 ± 0.2	1.92 ± 0.04	7.2 ± 0.6
$F_{12} \rightleftharpoons U$	-5.0 ± 0.7	5.8 ± 0.5	5.0 ± 0.6	18.2 ± 1.5
$F_{34} \rightleftharpoons U$	-4.1 ± 0.5	5.8 ± 0.5	4.1 ± 0.4	17.3 ± 1.5
$F_{23} \rightleftharpoons U$	-1.4 ± 0.3	5.4 ± 0.3	3.7 ± 0.3	15.6 ± 1.0

Table 5: Folding/unfolding rate constants and transition state positions for full-length calmodulin at 10 mM Ca^{2+} and zero force bias. The rate constants are given as logarithms. The distance to the transition state for unfolding Δx_{unf} is determined by fitting a Bell model (Equation 34). The distance to the folding transition state ΔL_{fold} is determined by fitting a model including all elasticities in the system (Equation 35).

6.5 THE FULL NETWORK, ALLOWED TRANSITIONS

The full kinetic network of wildtype calmodulin at high calcium concentrations consists of six distinct detectable states. In addition to the canonical states unfolded (U), N-terminal domain folded (F_{12}), C-terminal domain folded (F_{34}) and fully folded (F_{1234}), we find two additional states that exhibit non-native interactions across the inter-domain boundary (states F_{23} and F_{123}). We find a total number of six transitions that are significantly populated. The transitions with their force-dependent rate constants are shown in Figure 23.

From the unfolded state, the protein can fold to the native state via two direct pathways. First, via a folded N-terminal domain (F_{12}) and then additional folding of the C-terminal part into the native state. Second, via a folded C-terminal domain (F_{34}) and subsequent folding of the missing N-terminal part. However, from the unfolded state, a significant fraction of transitions (about 16 %) ends up in the non-native intermediate F_{23} . This state involves non-native even-odd pairing of EF hands. An NMR structure revealed that this structure still resembles the structure of a “normal” domain [58]. In contrast to the two latter states, F_{23} was never found to proceed further to the native state. All detected transitions lead directly back into the unfolded state. This classifies F_{23} as an off-pathway intermediate.

In state F_{12} , the protein can proceed via three pathways. The first two are the canonical pathways of unfolding into state U or folding into the native state F_{1234} . In addition to these, F_{12} can quickly exchange with F_{123} , as can be seen in the light blue-dark blue stretches in Figure 19 or Figure 20. From state F_{123} , there is no significant fraction of transitions into the native state. The vast majority of transition from F_{123} decays back into F_{12} . Likewise, from the native state F_{1234} no signifi-

The general folding motif for EF hands is “odd-even”, i.e. EF hands 1 and 2 or 3 and 4.

The term “off-pathway” in this context means off-pathway for folding from the unfolded to the native state.

cant fraction of transitions can be found for transitioning into F_{123} . [Figure 24](#) shows representative stretches of the states involving F_{12} , F_{123} and F_{1234} . In these examples no direct transitions between F_{123} and F_{1234} can be found. I will argue later in [Section 6.6](#) that the lack of this transition is a statistically significant feature of the folding of calmodulin.

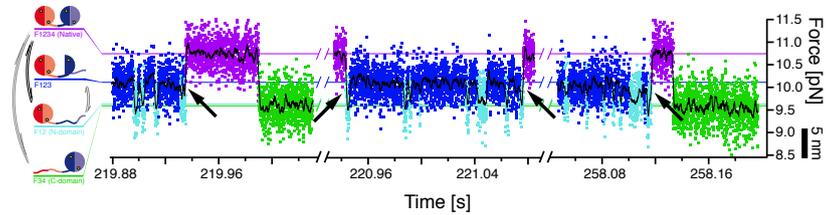


Figure 24: F_{123} and F_{1234} do not exchange. Representative stretches of data with transitions between the fast-exchanging states $F_{12} \rightleftharpoons F_{123}$ and F_{1234} . The transition always occurs via F_{12} (arrows).

Note that for the energy landscape representation, only the depths of the basins and the positions and (assuming a pre-exponential factor) heights of the barriers are known.

Taking the zero-force values for the transition rate constants ([Table 5](#)), we can construct a network of “transition fluxes” between the states. [Figure 25](#) shows this network superimposed on a representation of the energy landscape at an intermediate force.

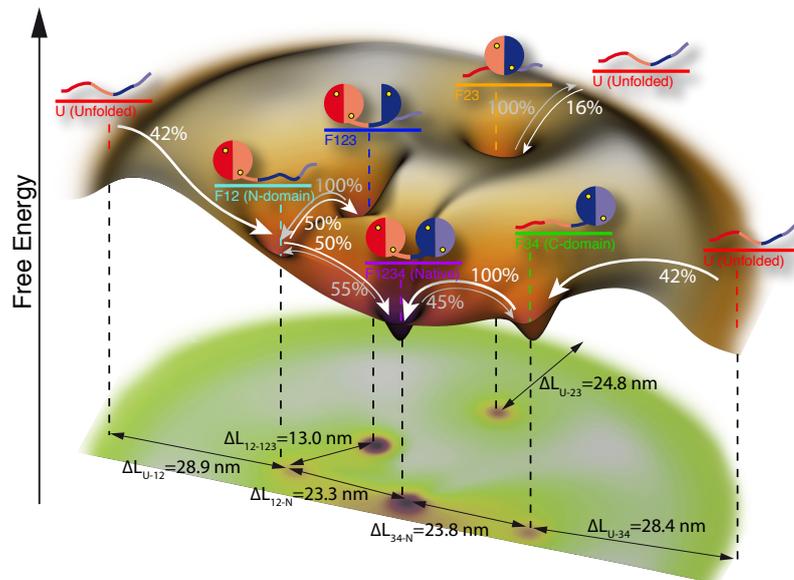


Figure 25: Interpretation for an energy landscape of the folding of calmodulin at high calcium. The flux of transitions between the states is indicated as percentages. Lengths in the lower part are distances in contour length between the states. The horizontal plane represents the reaction coordinate.

6.6 ON THE OFF-PATHWAY NATURE OF F_{23} AND F_{123}

There is a number of arguments why the states F_{23} and F_{123} are indeed off-pathway for folding to the native state.

If F_{23} were an on-pathway intermediate to F_{12} or F_{34} , short dwells at extensions corresponding to F_{23} should be visible in all transitions from U to F_{12} or F_{34} . A method of time-averaging all such transitions should reveal such an intermediate [100]. Figure 26 shows the time-averaged transitions from U to F_{23} , F_{34} and F_{12} . An obligatory intermediate F_{23} would cause an increase of the traces with a time-constant compatible to the average lifetime of F_{23} at this force (≈ 2.5 ms). However, the traces for F_{34} and F_{12} reach their steady levels within < 0.5 ms. We therefore can exclude that F_{23} can be an on-pathway intermediate for folding into F_{12} or F_{34} . This corroborates the interpretation that F_{23} always decays back into U.

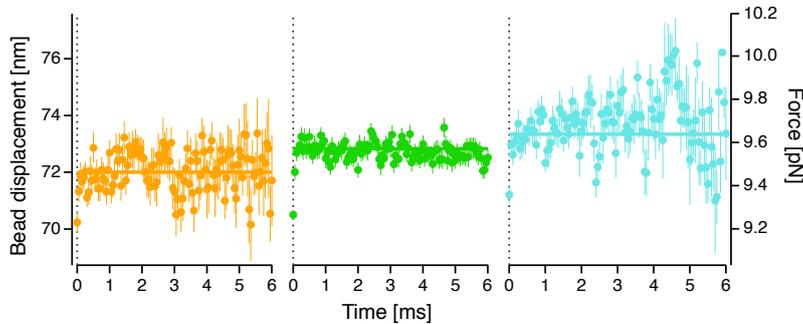


Figure 26: Averaged transitions starting from the unfolded state U into F_{23} (left, $N = 35$), F_{34} (middle, $N = 69$) and F_{12} (right, $N = 36$). An on-pathway intermediate could be seen in an exponential “smearing”. Error bars represent the standard error of the mean.

Likewise, as pointed out earlier, the transitions between the fast exchanging complex $F_{12} \rightleftharpoons F_{123}$ and F_{1234} occur via F_{12} (cf. Figure 24), rendering F_{123} an off-pathway intermediate. Nevertheless, especially at low forces, transitions occasionally can be found between F_{123} and F_{1234} . I will argue here that the appearance of these transitions is an expected artifact of very short-lived dwells in F_{12} and that, in fact, the transitions always occur via F_{12} .

Figure 27 shows the fraction of transitions between the fast-exchanging $F_{12} \rightleftharpoons F_{123}$ complex and F_{1234} that appear to transition via F_{123} . The fraction appears to be higher at low forces and lower at high forces. To test the significance of the (non-zero!) values of these fractions, simulations were performed on a system where explicitly no transitions between F_{123} and F_{1234} were allowed. Due to the limited temporal resolution, short-lived dwells in F_{12} are likely to be missed and transitions between the $F_{12} \rightleftharpoons F_{123}$ and F_{1234} complex appear to come from F_{123} , although in reality they come from F_{12} . The red lines in Figure 27, bot-

Even if (at high biasing forces) there was indeed a small fraction of transitions $F_{123} \rightleftharpoons F_{1234}$, the resulting rate constants would be much smaller than the other rate constants involved and essentially negligible.

tom are the expected fraction of these events for realistic cutoff values of $400 \mu\text{s}$ and $600 \mu\text{s}$. The measured fraction of $F_{123} \rightleftharpoons F_{1234}$ transitions is within the error intervals compatible with the value that is expected due to the missal of short F_{12} -dwells. Hence, there is no statistical evidence for the existence of such a transition. Chiefly, the data of the longest measured trace (shown as a green circle in Figure 27) also corroborate this claim. Here, at a cutoff of $400 \mu\text{s}$, the expected fraction of detected $F_{123} \rightleftharpoons F_{1234}$ transitions is $\approx 28\%$. The measured value is 30% .

Taken together, within statistical errors, there is no evidence that F_{23} or F_{123} are on-pathway.

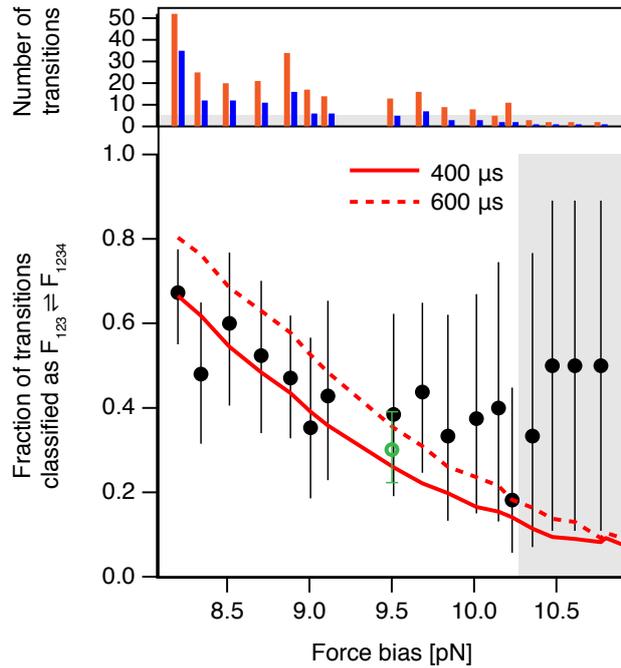


Figure 27: Off-pathway nature of F_{123} . Shown in the bottom graph is the fraction of transitions between F_{1234} and either F_{12} or F_{123} that were found directly between F_{123} and F_{1234} (black circles). Error bars are 1σ intervals. The red continuous line is the result of a simulation for the expected amount of such classified events for a temporal cutoff of $400 \mu\text{s}$, the red dashed line for $600 \mu\text{s}$. The top graph shows the total number of transitions between F_{1234} and either F_{12} or F_{123} (red) and those identified as $F_{1234} \rightleftharpoons F_{123}$ (blue). The green circle is data from the longest analyzed experiment. In the grey shaded areas the number of events are low and data unreliable.

6.7 THE EFFECT OF KINETIC TRAPS ON THE FOLDING TIME

The presence of the two kinetic traps F_{23} and F_{123} has interesting consequences on the predicted folding time of calmodulin at high calcium concentrations. Simulations of the folding behavior at zero force allow to calculate the expected distribution of folding times from U to F_{1234} , i. e. in the sense of a first passage time, the time to reach the native state from state U (Section C.2). Figure 28 shows these times.

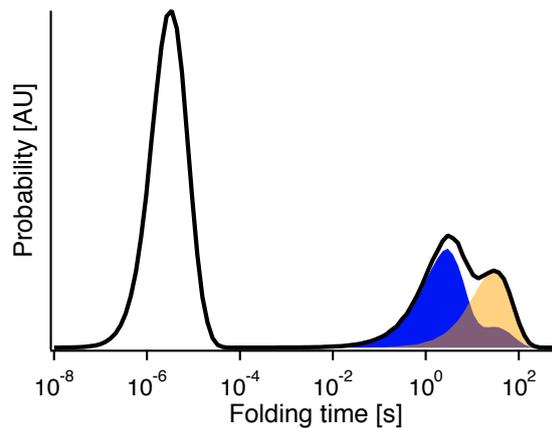


Figure 28: Distribution of calculated folding times for full-length calmodulin at high calcium and zero force (first-passage times from U to F_{1234}). The blue distribution shows only folding times of simulated trajectories that visited F_{123} at least once. The orange distribution shows those that visited F_{23} at least once.

About 40 % of all trajectories fold via F_{34} . The folding on this pathway is very rapid on the microsecond scale (see Figure 25 and Table 5). Another fast pathway is the direct folding via F_{12} , with folding times on a similar time scale. Both show up in the left peak in Figure 28. However, the presence of kinetic traps have a retarding contribution on the overall folding time. When the protein visits the kinetic traps F_{23} or F_{123} , already folded structures will have to unfold (a part of the fold in the case of F_{123} or completely in the case of F_{23}) to proceed to the native state. Due to the relatively high thermodynamic stabilities of these structures ($-\Delta G_{U,F_{23}}^0 = 13 \text{ k}_B\text{T}$, $-\Delta G_{F_{12},F_{123}}^0 = 10 \text{ k}_B\text{T}$), the unfolding on average is a very slow process. As a consequence, even though the folding times of the individual components of calmodulin are very rapid (see Table 5), the gross average folding time for full-length calmodulin is on the order of seconds.

Experimental evidence for this effect can be seen in relax cycles (Figure 29). If the folding occurs via the fast intermediate F_{34} (left traces), it occurs rapidly and hence already at high forces (arrows). However, if the folding occurs via F_{12} (right traces), it will at high forces rapidly fold into F_{123} , from where it will have to unfold again to be able to proceed fold-

ing. As a result, refolding occurs much later in the trajectory and hence at lower forces (arrows).

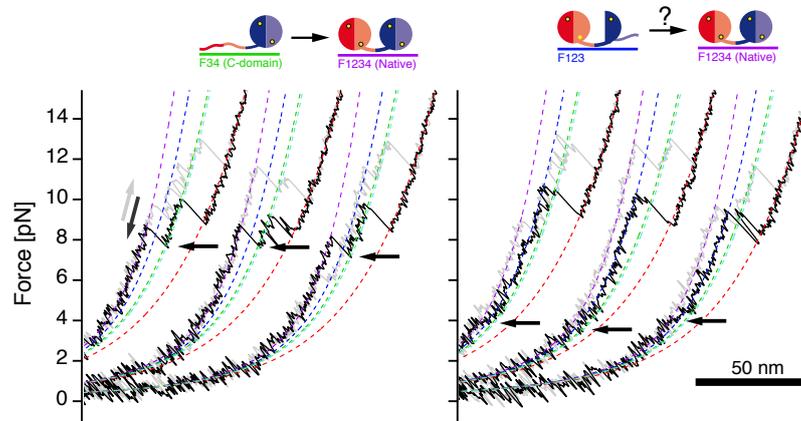


Figure 29: The effect of kinetic traps on the refolding. Shown are relax cycles at 500 nm/s of full-length calmodulin sorted according to their folding pathway. Left traces: folding occurs via F_{34} , the folding occurs early and forces are high (arrows). Right traces: folding occurs via F_{123} , the folding times are longer and forces are lower (arrows).

6.8 HOW GOOD IS THE ASSIGNMENT?

The results presented above crucially depend on the proper recognition and reliable assignment of states. This section will discuss this issue.

Figure 30 shows the distributions of the data points classified into each of the six states in full-length calmodulin at an intermediate force. The red, dark blue and purple states (U, F_{123} and F_{1234}) are easily distinguishable by their average force. The orange, green and light blue states (F_{23} , F_{34} and F_{12}), however, cluster at almost the same average force. The fitting residuals of fitting a skewed Gaussian PSF to the data does not show any detectable false classifications of states. The classification of states now depends on the distinction between F_{23} , F_{34} and F_{12} .

As discussed in Section 4.1, Hidden Markov models assign the data points not only based on the center position of the state distributions alone, but also based on the connectivity of the network. Since F_{34} has routinely much longer lifetimes than F_{12} , which interchanges quickly with F_{123} (see e. g. Figure 22A), the two states are easily distinguishable.

The idea that the states F_{12} and F_{23} might be mixed up also can be rejected. Even though they exhibit similar lengths and, at certain forces, even similar lifetimes, the network connectivity clearly distinguishes between the two. While F_{23} always decays back to U, F_{12} quickly exchanges with F_{123} . False classifications of the type where $U \rightarrow F_{12} \rightarrow U$ events might be falsely identified as an F_{23} event, are only possible, if at a certain force bias the rate constant $F_{12} \rightarrow F_{123}$ is comparable to (or smaller

For a system where the fitting residuals to a PSF were used to identify states, see a recent study by Yu et al. [110].

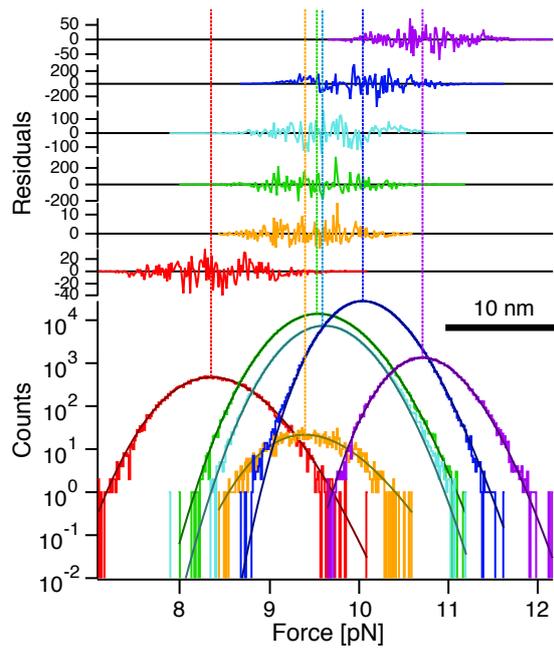


Figure 30: Histogram of all states in full-length calmodulin at a biasing force of 9.6 pN. The histograms shown with fits to a skewed gaussian PSF along with their fitting residuals.

than) $F_{12} \rightarrow U$, i. e. F_{12} unfolds before it goes into quick transitions with F_{123} . However, experimentally we find that the former rate constant is orders of magnitude higher than the latter in most of the accessible force range (see [Figure 23](#)). False classifications of this sort are hence unlikely.

Additionally, at higher force bias, F_{12} and F_{23} are separable not only by their slight differences in length, but also by their average lifetimes. [Figure 31B](#) shows the lifetime assignment at high forces. Hypothetical lifetime histograms that would be measured, if F_{12} and F_{23} were the same state, do not follow a single exponential distribution ([Figure 31A](#)).

Another necessary test is whether or not it is really justified to invoke a six-state model, or if a five-state model would be sufficient to describe the data. [Figure 32](#) shows the outcome of such an approach. A five-state Hidden Markov classifier is able to correctly classify most of the states. However, it fails to distinguish F_{34} and F_{23} and pools the two together, resulting in clearly double exponential lifetime distributions ([Figure 32A](#)). The underlying two states are easily seen in a bead displacement vs lifetime scatter plot ([Figure 32B](#)). These inconsistencies clearly rule out a five-state model and strongly suggest that at least six states are required to describe the data.

6.8.1 Brownian dynamics simulations

Brownian dynamics simulations of the folding properties can also help to assess the performance of the Hidden Markov classifier. The simulations

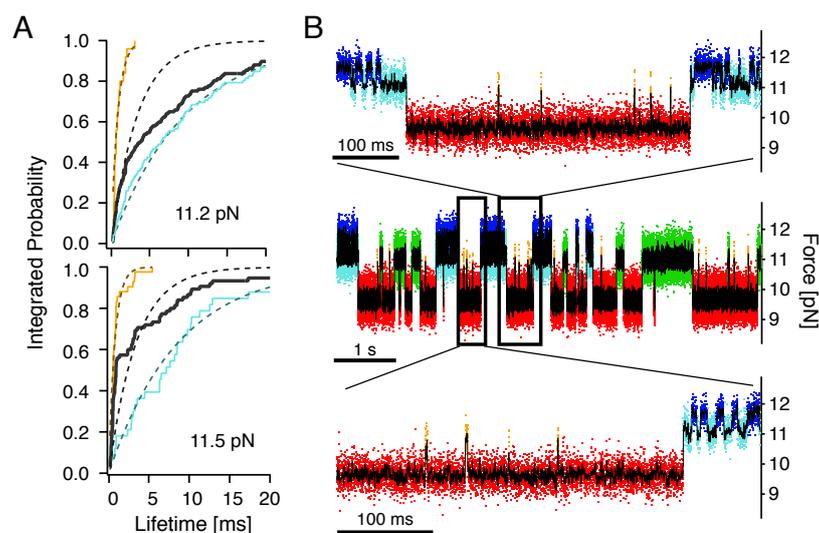


Figure 31: F_{23} and F_{12} are distinguishable by their lifetimes at higher forces. (A) Integrated lifetime histograms at high biasing forces. At these forces, F_{23} (orange) and F_{12} (light blue) exhibit clearly different lifetimes. A hypothetical lifetime histogram for the case that F_{23} and F_{12} were the same state is shown in black. It does not follow a single exponential distribution (dashed lines). (B) Sample trace at a force bias of ≈ 11 pN with temporal zooms.

performed here were based on a Markov chain Monte Carlo system that decided on the folding-state of the protein at each time point. The transition rate constants were calculated based on the data in Table 5. The response of the bead was then simulated using a Brownian dynamics system. A more detailed description can be found in Section C.1.

A comparison of measured and simulated traces is shown in Figure 33. The similarity of the traces confirms the applicability of the simulation system. Since the hidden state of the protein is known at each time of the simulation, such simulations can be used to assess the performance of the classification algorithm. The arrows in Figure 33B show stretches of data that were not correctly identified by the Hidden Markov classifier. Generally, the correct classification performance of the Hidden Markov algorithm is 96%. It should be noted that, even though not every dwell is correctly identified, the majority certainly is. Further, the main conclusions drawn here for the folding of calmodulin do not crucially depend on the exact identification of every single data point, and are statistically significant.

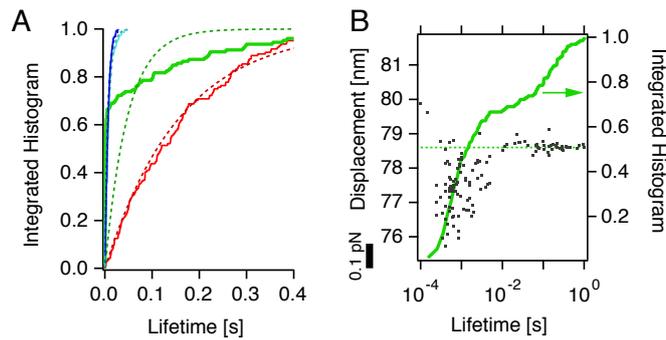


Figure 32: Results of a five-state model applied to the data. (A) While the model adequately classifies the states U (red), F_{12} (light blue) and F_{123} (dark blue), it pools all occurrences of F_{23} into the green state F_{34} , resulting in clearly double exponential lifetime distributions. Fits are single exponentials. Note that at the shown force bias F_{1234} is not populated. (B) Scatter-plot of bead displacement versus lifetime of the green state in (A). The points scattered around the dashed green line belong to F_{34} . The falsely classified occurrences of F_{23} cluster around a lower displacement value. Note that due to do statistical effects (\sqrt{N} error), the shorter lived states scatter more in the vertical dimension.

6.9 CONTROLS: INTERFERENCE OF THE OXYGEN SCAVENGER?

In addition to the controls already discussed above (truncation mutants, crosslinked mutant), additional control measurements were performed to test whether unwanted interactions of the present oxygen scavenger could interfere with the measurements. Switching the oxygen scavenger from a GLOXY system to a PCA/PCD-system [1, 59] gave identical results, indicating that no component of the GLOXY system specifically binds to calmodulin and affects its folding properties.

Moreover, measurements with highly reduced concentrations of the individual components of GLOXY, led to earlier irreversible damage to the molecule during the measurement. During its undamaged lifetime, however, its properties were indistinguishable from the normal behavior reported here.

6.10 DISCUSSION

6.10.1 Anticooperativity between the domains

The equilibrium free energy data obtained from relative state populations and summarized in Table 4 suggests anti-cooperative effects between the two domains. From the unfolded state, folding of either the N-terminal (F_{12}) or the C-terminal domain (F_{34}) results in an energetic

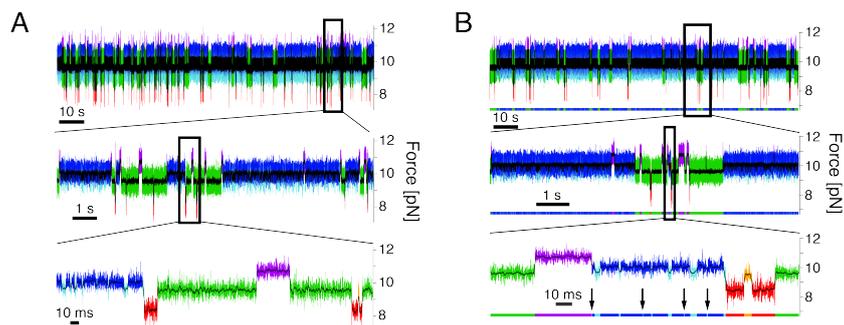


Figure 33: Brownian dynamics simulation of traces of the folding of full-length calmodulin at 10 mM Ca^{2+} . (A) Measured trace. (B) Simulated trace. The bars below the trace are show the real state the system is in at each time point. Arrows show dwells not recognized by the Hidden Markov classifier.

gain of $20 k_B T$. However, folding of the second domain only provides $15 k_B T$. This indicates that the presence of one folded domain destabilizes the other domain. An energetic coupling between the two domains has been reported earlier in bulk measurements [69]. In the cited study the data suggested that the N-terminal domain is destabilized by a folded C-terminal domain. Notably, a similar effect could also be observed for full-length calmodulin under apo conditions (see Chapter 7).

An interesting feature corroborating the notion of anti-cooperativity between the domains in the folding process is contained in the contour lengths of the states (Table 3). From the unfolded state, the lengths of the states for the folded N-terminal domain (F_{12}) as well as for the folded C-terminal domain (F_{34}) are significantly bigger than the lengths that are calculated based on an assumed symmetric splitting between the domains. This indicates that whichever domain folds first incorporates *more* than half of the available unfolded polypeptide into its structure. This is in accord with the asymmetric gains in energy.

6.10.2 Folding rate constants

Interestingly, the folding rate constants summarized in Table 5 are all in the range of $10^5 - 10^6 \text{ s}^{-1}$. A possible explanation for this finding is that calcium binding is rate-limiting for the folding. A more detailed discussion of this effect based on the folding properties at a range of different calcium concentrations will be presented in Chapter 7.

6.10.3 Kinetic traps

As described earlier, the presence of kinetic traps significantly slows down the folding of calmodulin at high calcium. One might ask, why are the measured kinetics under load much faster than the reported average fold-

ing time of about 5 seconds? Shouldn't force slow down the folding times even more? The answer lies in the properties of the kinetic traps. Once the system gets trapped in a misfolded state, this misfolded structure first needs to unfold in order to proceed with folding toward the native state. It is this unfolding, that takes very long time at zero load and effectively causes the long folding times. Force actually increases these unfolding rate constants, and thus increases the average speed of folding under certain load conditions. [Figure 34A](#) shows the calculated distribution of folding times at zero load (red trace) and increasing loads (rainbow colors). [Figure 34B](#) shows the average folding time as a function of trap separation. As discussed, there is indeed a narrow range at non-zero load, where the average folding time is faster by a factor of about 10, compared to zero-load rates. Most measurements were incidentally performed in this range, where the kinetics of a slow-folding protein are faster than at zero load.

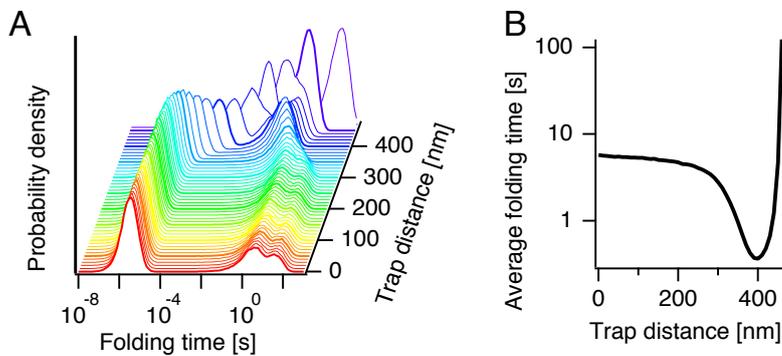


Figure 34: Calculated folding times at various trap separations. (A) Distribution of folding times at varying trap distances. (B) Average folding time as a function of trap distance.

6.10.4 Are the transitions independent?

As discussed in [Section 4.1](#), the transitions between states are expected to follow (first order-) Markovian statistics, i. e. the probability to transition into a certain state depends only on the present state. This “memorylessness” is a fundamental prerequisite for the applicability of a Markov model to the data. Deviations from a memoryless behavior could, for instance, indicate non-stochasticity of the process. Non-stochastic behavior is expected in energy-driven systems, where the flux of transitions between states is not balanced [47]. Also long-range memory (e. g. when the protein does not “forget” in which state it was two transitions ago) causes deviations from a purely stochastic behavior.

It is important to note that even in not purely stochastic processes the lifetimes in the states can still be distributed single exponentially (see [Section 4.3](#) for an example). The test, whether or not a transition is only

In a Markov chain of first order, transitions only depend on the current state's predecessor. In second order Markov chains, they also depend on the state before that, and so on.

dependent on the present state, is thus not covered by the test for lifetime exponentiality.

It is hence important to ensure that the network between states indeed resembles a stochastic Markov chain. While there are theoretical tests that can infer the order of the involved Markov chain (e. g. [85]), they only perform well on diffusion-type processes and turned out to not work robustly for the analysis of hopping trajectories. A simpler approach is to investigate the distributions of *consecutive* transitions from a certain state into another state. Memorylessness requires that the events be statistically independent, hence the distributions be geometric. Deviations from geometric distributions would occur, for example, in the presence of long-range memory.

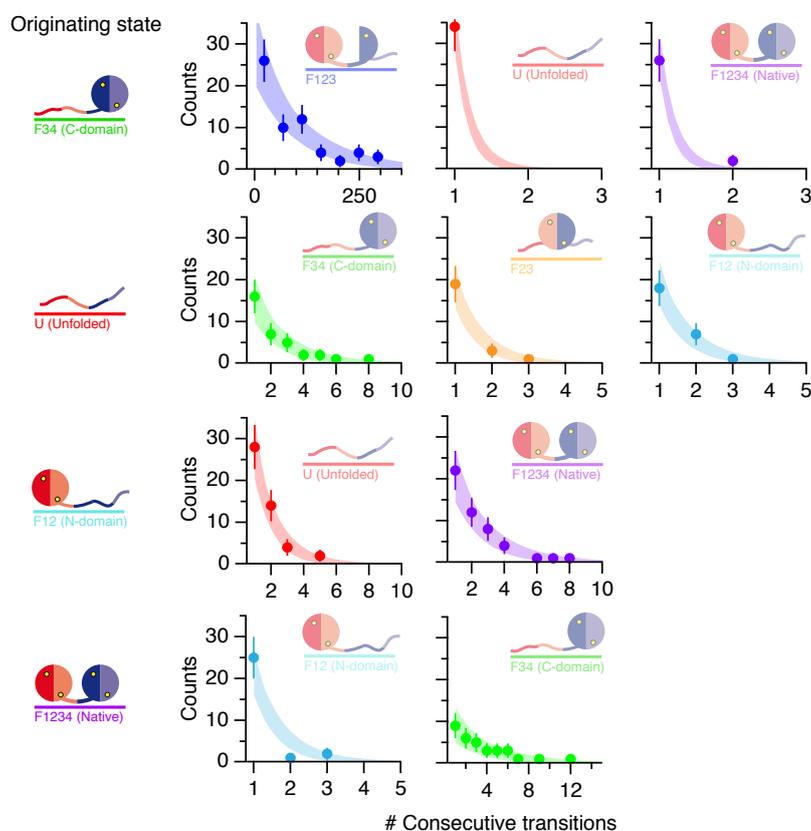


Figure 35: The transitions are first-order Markovian. Shown are the distributions of *consecutive* transitions from a state (shown on the left) into another state (indicated with cartoons in the graphs). Shaded areas are 1σ prediction intervals for the expected values of a Markov chain.

Figure 35 shows the distributions of consecutive transitions from a certain state i (shown on the left) to another state n for full-length calmodulin at 10 mM Ca^{2+} at a biasing force where all states are populated. The shaded areas are 1σ prediction intervals for the expected distributions for a memoryless Markov chain (cf. Section 4.3). The data points follow

the prediction. This justifies *a posteriori* the assumption of a stochastic Markov chain model for the transition network.

6.10.5 Structural information

From the measured data using the dumbbell assay, it is not possible to obtain detailed structural information of the folding intermediates of calmodulin. The structural information presented here in the cartoon representations of the states (e. g. [Figure 23](#)) are purely based on the crystal structure of the native state ([Figure 1B](#)) and the information obtained from the truncation mutants.

Deviations from the simple structural model where each of the domains folds individually could already be observed in the folding lengths of the intermediate states ([Table 3](#), [Section 6.10.1](#)). Further complications arise when we consider that the dumbbell-like crystal structure of full-length calmodulin as shown in [Figure 1B](#) is not the only available structure of calcium-loaded calmodulin. The extended central linker helix was found to be incompatible with Small Angle X-ray Scattering (SAXS) measurements [46] and it has been suggested that the linker is quite flexible in solution. In 2003, [Fallon and Quiocho](#) published a structure of calcium-bound calmodulin, where the overall structure is globular and the central linker helix virtually missing [39].

Since the variabilities in the linkers are big between different molecules, it is difficult to reliably compare length information between different molecules. Presumably due to low energetic differences between the conformations, it was not possible to identify different conformational states in this assay.

6.10.6 Influence of the ubiquitin handles

For the interpretation of the measured data it is necessary that the results are not influenced by contributions of the measurement procedure. As discussed earlier, specific influences of the oxygen scavenger system could be ruled out. It turned out, however, that the ubiquitin domains, that serve as spacers in the dumbbell construct, have measurable influence on the outcome of the experiment in some cases.

While ubiquitin domains have relatively low thermodynamic stability that is comparable to a single domain of calmodulin [27], they exhibit a high kinetic stability [89], such that in a typical experiment, an unfolding event of ubiquitin is rare. An unfolding event of ubiquitin appears as a sudden increase in linker length and an associated drop in force. If there is no interaction between ubiquitin and calmodulin, the energetics, kinetics and lengths of calmodulin are expected to be unaffected whether ubiquitin is folded or not.

[Figure 36](#) shows the force-dependent rate constants of a molecule, where during the experiment one of the ubiquitin handles had unfolded (empty circles). The continuous lines in the figure are the rate constants as deter-

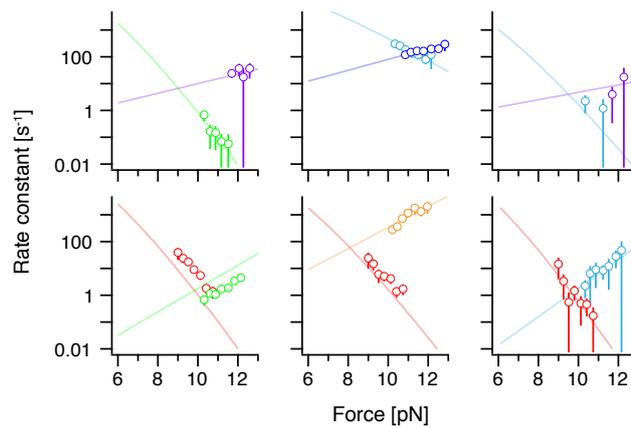


Figure 36: Influence of an unfolded ubiquitin on the kinetics of full-length calmodulin. Empty circles are measured rates after an ubiquitin domain had unfolded. Continuous lines are best fits for *folded* ubiquitins as shown in Figure 23. Coloring as in Figure 19.

mined for the intact construct (cf. Figure 23). Generally, the entire network of states and transitions persists. Some transition rates, however, appear to be influenced by the presence of an unfolded ubiquitin (see, e.g. the transition between states U and F_{34} in the lower left panel of Figure 36).

As a control, we created a construct entirely without ubiquitin residues. Even though the network of states is unaltered from the construct where ubiquitins are present, the force-dependent rate constants show slight deviations. Figure 37 shows the force-dependent rate constants for a construct with no ubiquitins (empty squares) compared with the average results of full-length calmodulin with folded ubiquitins (continuous lines).

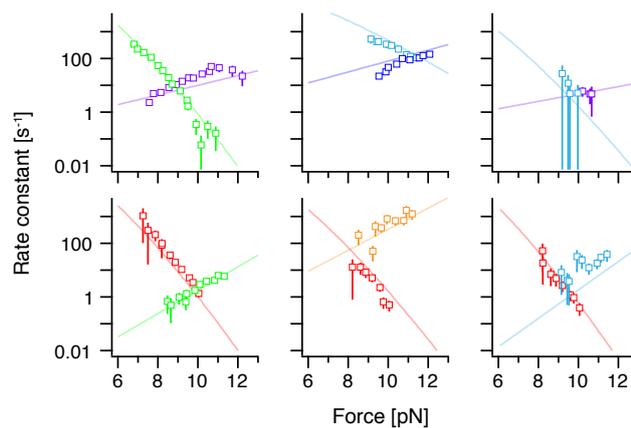


Figure 37: Kinetics of a full-length calmodulin construct at 10 mM Ca^{2+} without ubiquitin handles. Continuous lines are best fits for *folded* ubiquitins as shown in Figure 23. Coloring as in Figure 19.

For comparison, [Table 6](#) lists the equilibrium energies in all three cases (all ubiquitins folded: 2F, one ubiquitin unfolded: 1U, and no ubiquitins present: NU). The energies are given as difference from state U.

STATE	$-\Delta G_0$ ($k_B T$)			LENGTH (nm)		
	2F	1U	NU	2F	1U	NU
F ₂₃	12.7	13.9	12.4	25.4±1.2	25.3±0.6	26.1±1.6
F ₃₄	20.4	23.1	21.8	28.7±0.8	28.6±0.4	29.0±1.7
F ₁₂	20.3	20.0	19.0	29.6±0.7	29.0±0.5	30.5±1.7
F ₁₂₃	29.4	29.3	28.4	39.7±0.6	39.4±0.5	41.0±1.5
F ₁₂₃₄	34.9	41.0	34.4	51.7±0.6	54.4±0.7	52.9±1.6

Table 6: Equilibrium free energies $-\Delta G_0$ of a single full-length calmodulin molecule with folded ubiquitins (2F), one ubiquitin unfolded (1U) and a construct with no ubiquitins (NU) 10 mM Ca^{2+} . The equilibrium free energies are given in units of $k_B T$ as difference from the unfolded state U. Lengths are given contour of the unfolded polypeptide as difference from state U. Note that (2F) and (1U) are measured for the same molecule.

In direct comparison between the case of all ubiquitins folded (2F) and the construct where no ubiquitins are present (NU), the equilibrium free energies are identical within experimental uncertainty. Also the differences in contour length between the states (measured as difference from state U) show no significant discrepancies. Hence, an influence of *folded* ubiquitin residues on calmodulin, apart from slight shifts in the force-dependent rate constants ([Figure 37](#)), is unlikely.

In contrast, an influence is evident when one of the ubiquitins is unfolded (1U). In addition to slight shifts in the transition rates ([Figure 36](#)), also the equilibrium free energies of the states F₃₄ and F₁₂₃₄ are shifted. Even though the microscopic details of this interaction are unclear, one may speculate on the mechanism. One possible explanation would be a scenario where some residues of the unfolded polypeptide of ubiquitin is incorporated in the folded C-terminal domain. The net stabilization of this can be seen in the equilibrium free energies (23.1 vs 20.4 $k_B T$, [Table 6](#)), but are less apparent in length (28.6 vs 28.7 nm). The subsequent folding of the N-terminal domain provides an additional 3 $k_B T$ of free energy to a net energy difference between the 1U and 2F cases of ≈ 6 $k_B T$. Note that, since the folded C-terminal domain is different from F₃₄ in this case (because it has part of the unfolded ubiquitin incorporated), the destabilizing effect between the domains (see [Section 6.10.1](#)) might also be altered. Even though most of the length information are identical for the 1U and 2F case, there is a small difference for the native state F₁₂₃₄ (54.4 vs 51.7 nm), which can account for the additional difference in free energy. The length difference would suggest that the F₁₂₃₄ states of the cases 1U and 2F are structurally different. It should be stressed, however,

Note that, since the cases 2F and 1U are measured in the same molecule, the lengths and energies are directly comparable.

that this scenario is purely speculative. From the data shown here it is not even possible to determine which of the two ubiquitins has unfolded.

Taken together, there is a measurable influence of *unfolded* ubiquitin domains on the stability, kinetics and folding of length full-length calmodulin. The detailed mechanism of this effect is unclear. However, a possible scenario is the incorporation of part of the unfolded polypeptide of ubiquitin into the C-terminal domain. In contrast, a comparison between a construct with two folded ubiquitin domains and a construct with no ubiquitin domains at all showed that the influence of *folded* ubiquitin domains is, if present, only measurable in slightly altered force-dependent transition rate constants for few of the transitions. The lengths and energetics are unaffected. Measurements with two *folded* ubiquitin domains are therefore comparable to measurements with no ubiquitins at all.

6.11 SUMMARY

In this chapter, I presented the results of the N-C-terminal stretching of full-length calmodulin at high calcium concentrations of 10 mM. Constant velocity and constant distance data revealed a network of six distinct states, which could be assigned to general structures along the folding pathway. In addition to the canonical unfolded and native states, the previously known states where only the N-terminal domain or only the C-terminal domains are folded, could be identified. Furthermore, two additional states were found: One involving the non-native pairing of the EF hands two and three, another one, where EF hand three collapses onto the folded N-terminal domain in a non-native conformation. Using a Hidden Markov model, relative populations of states down to 10^{-4} (see [Figure 22B](#)) could be identified and the kinetics and energetics in the network could be determined. Using this information it was possible to follow the folding trajectory of a single calmodulin molecule over time.

While the two domains of calmodulin are able to fold individually, in full-length calmodulin, inter-domain interactions lead to interesting effects in its folding behavior. The two non-native states both represent kinetic traps in the energy landscape that governs the folding of calmodulin. Once in a dead-end kinetic trap, parts of the structure need to unfold and folding can start anew. As a result, calmodulin is, at high calcium concentrations, an effectively slow folding protein. [Chapter 7](#) will show that this is likely an effect induced by the artificially high calcium concentration and that similar effects at physiological calcium concentrations are unlikely.

In addition to cooperative effects between the domain, such as the presence of intermediates that involve parts of both domains (F₂₃ and F₁₂₃), also anti-cooperative effects between the domains could be observed: The folding of a domain in full-length calmodulin is driven by a lower gain in free energy, if the other domain is already folded.

In general, it could be shown that by using truncation mutants, also an energy landscape with more than one dimension (compare e. g. linear systems as DNA hairpins [106] or coiled coils [43]), can be effectively unraveled, even though experimental data is only measured as a projection onto a one-dimensional (pulling-) coordinate.

Using statistical arguments, it could be shown that the process of hopping between states in this complex network of six states indeed follows first-order Markovian statistics. The assumption to describe the hopping trajectory in the well-developed framework of Markov chains is hence justified.

The folding trajectories obtained are closely reminiscent of trajectories that were simulated in coarse-grained and all-atom molecular dynamics simulations [45, 92], suggesting that the underlying processes are generally well-described in modern simulations. The description of the folding process as a Markov chain hopping between states in an energy landscape [60, 77] is hence a successful and accurate model.

7

CALCIUM-DEPENDENT FOLDING
OF CALMODULIN

All measurements of [Chapter 6](#) were performed at high calcium concentrations of 10 mM. The high calcium gave rise to increased stability. However, *in vivo* calcium concentrations are considerably lower. Generally, they hover around the micromolar regime and increase rapidly by a factor of 10 upon exterior stimuli [28]. The calcium dissociation constants of calmodulin lie in the low micromolar range and are presumably evolutionally tuned to the low intracellular concentrations, where binding events can switch the conformation from the apo to the structurally distinct holo conformation (see [Chapter 2](#)). The study presented in this chapter investigates the folding and unfolding properties of calmodulin at calcium concentrations that fall closer to the biologically relevant range.

7.1 FULL-LENGTH CALMODULIN AT LOWER CALCIUM CONCENTRATIONS

To investigate the folding properties of calmodulin, we studied the full-length constructs of [Chapter 6](#) in an optical tweezers assay at different calcium concentrations. [Figure 38](#) shows stretch and relax cycles and constant trap separation traces at calcium concentrations of 10 mM, 100 μ M and zero (by adding 10 mM Ethylenediaminetetraacetic Acid (EDTA)).

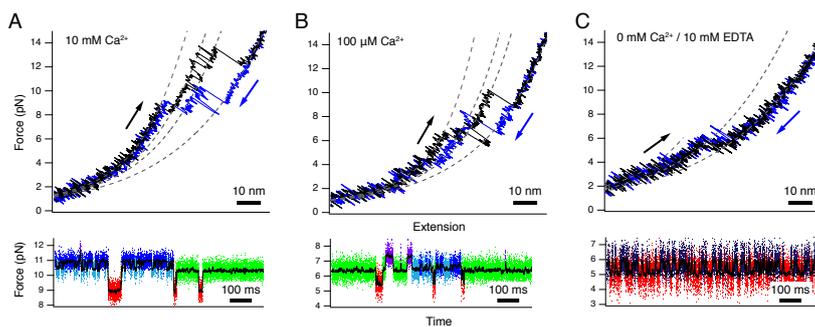


Figure 38: Full-length calmodulin at different calcium concentrations. Top traces: stretch (black) and relax (blue) cycles at 500 nm/s at (A) 10 mM Ca^{2+} , (B) 100 μM Ca^{2+} , (C) 0 mM Ca^{2+} / 10 mM EDTA. Bottom traces: Constant trap separation fluctuation traces at the respective concentrations of ligands. Dashed lines are fits to recognizable states.

The stabilizing effect of ligand binding (compare [Section 5.4](#)) can be readily seen in the top stretch and relax traces. While at 10 mM Ca^{2+} , calmodulin unfolds at the shown pulling velocity of 500 nm/s at forces around 12 pN, this figure drops at 100 μM to below 10 pN. At zero calcium conditions, the unfolding occurs at about 5 pN.

The network of six states as was determined at high calcium is fully present also at 100 μM Ca^{2+} (compare bottom traces in [Figure 38A](#) and [B](#)). However, this picture changes when no ligand is present. Only a two-state behavior can be clearly observed in [Figure 38C](#) around 5 pN, exhibiting the length of one folded domain. The presence of the second folded domain, as indicated by fits in the upper trace of [Figure 38C](#) can only be inferred. The folded state of transition at zero calcium is arguably the folded state of the N-terminal domain and therefore colored in dark blue.

The stabilizing effect of calcium can also be directly seen in the equilibrium free energies of all states shown in [Table 7](#). Notably all energies are affected by calcium, corroborating the claim that all states can bind calcium.

STATE	10 mM Ca^{2+}	100 μM Ca^{2+}
F ₂₃	13	4
F ₃₄	21	11
F ₁₂	20	11
F ₁₂₃	30	14
F ₁₂₃₄	36	17

Table 7: Equilibrium free energies $-\Delta G_0$ of full-length calmodulin at different calcium concentrations. Energies are given in units of $k_B T$ as difference from state U.

The free energies of the isolated calcium-free calmodulin domains and their binding constants were determined earlier by [Masino et al.](#) and can here serve as controls. At 10 mM Ca^{2+} , the stability for the N-terminal domain can be calculated to $19.7 \pm 0.8 k_B T$ and $21.8 \pm 0.6 k_B T$ for the C-terminal domain [[69](#)]. This corresponds well to the energies of the states F₁₂ and F₃₄ in full-length calmodulin ([Table 7](#)). At 100 μM Ca^{2+} , the calculated free energies of the N and C-terminal domains are $10.5 \pm 0.8 k_B T$ and $12.6 \pm 0.6 k_B T$, respectively. The corresponding value determined by our assay, for both the states F₁₂ and F₃₄, is 11 $k_B T$. The good agreement between energies determined based on bulk data and our own measurements corroborates the accuracy of our assay.

Unfortunately, the dependence of stability on the ligand concentration causes the unfolding and refolding forces at low Ca^{2+} to drop down to levels where the unambiguous separation between the states is not straightforward any more. We therefore chose to investigate the folding/unfolding behavior of the isolated N and C-terminal domains at lower calcium concentrations. These domains have been shown to exhibit two-state folding/unfolding behavior at high calcium concentrations [[53](#), [94](#)].

7.2 FOLDING/UNFOLDING OF THE N-TERMINAL DOMAIN AT LOW CALCIUM

Similar to full-length calmodulin, the folding/unfolding forces of the N-terminal domain depend to a great extent on the calcium concentration, as can be seen in [Figure 39A](#). While at 10 mM, the unfolding/refolding occurs around 11 pN, this value drops down to about 6 pN for 100 μ M and about 5 pN for zero calcium.

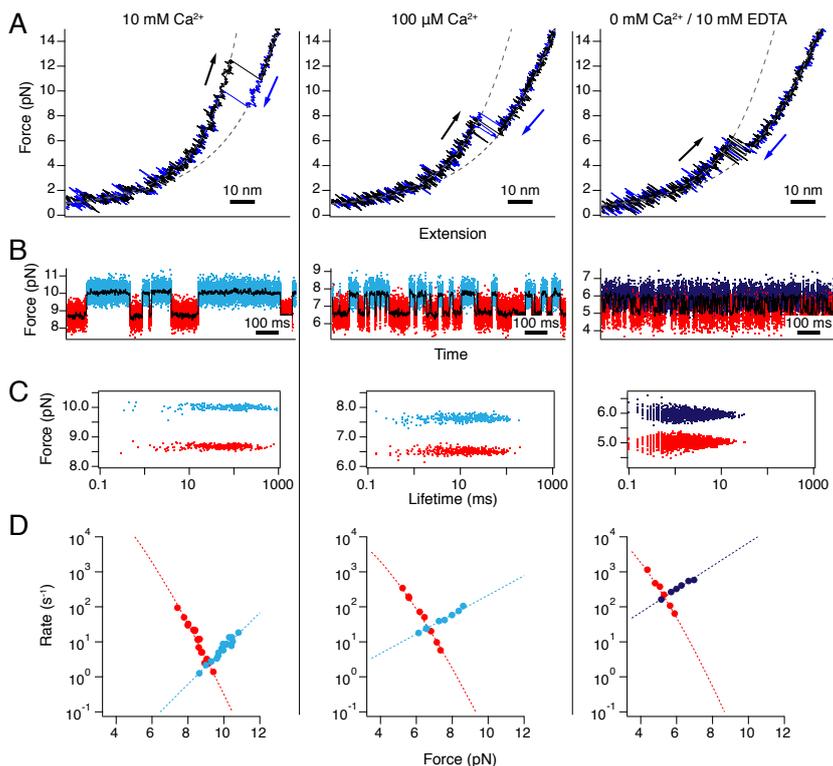


Figure 39: The N-terminal domain of calmodulin at different calcium concentrations. (A) Stretch (black) and relax (blue) cycles at 500 nm/s. (B) Constant trap separation traces. (C) Scatter plot of all detected dwell average forces vs their lifetimes. (D) Chevron plot for unfolding (blue colors) and refolding (red) and corresponding fits. The folded state in its calcium-free apo configuration is shown in a darker shade of blue.

In addition to forces, also the transition kinetics depend on the calcium concentration. This can be seen already in the force-extension cycles of [Figure 39](#), where the hysteresis between the stretch cycle (black) and the relax cycle (blue) becomes smaller for lower calcium concentrations. A reason for this is the faster kinetics at lower calcium, which effectively places the transition closer to equilibrium. [Figure 39B](#) shows constant trap separation traces at biasing forces where the folded and unfolded states are approximately equally populated. Hidden Markov analysis (see [Section 4.1](#)) helped assign data points to classes in order to deter-

mine dwells. A clear increase of the transition kinetics with decreasing calcium concentration can be seen.

A scatter plot where all dwells in a certain states are represented by their force plotted vs their lifetime (Figure 39C) shows two distinct clusters, one for the folded, one for the unfolded state, confirming the two-state behavior of the N-terminal domain at all shown calcium concentrations.

For a discussion concerning the choice of fits, see Section 4.6.

The force dependent rate constants of these two-state transitions are shown in Figure 39D along with fits that allow for extrapolation. Since the unfolding branch (blue) did not show significant curvature, a Bell fit was chosen (Equation 34). For the folding branch (red), the folding model with all energetic contributions was applied (Equation 35).

7.3 FOLDING/UNFOLDING OF THE C-TERMINAL DOMAIN AT LOW CALCIUM

Similar to the N-terminal domain, we also investigated the isolated C-terminal domain at a series of different calcium concentrations. Figure 40 shows representative results at 10 mM, 100 μM and 0 mM Ca^{2+} .

In accordance with the N-terminal domain, also for the C-terminal domain the unfolding/refolding forces depend inversely on the calcium concentration (Figure 40A). The same is true for the kinetics. Lower concentrations of calcium cause faster transitions between the folded and unfolded state (Figure 40B).

However, the middle trace in Figure 40B at 100 μM Ca^{2+} shows deviations from a simple two-state folding behavior. Arrows point to dwells in the apparently folded state that are incompatible with the lifetimes of the folded state. The long-lived folded state is colored in light green, the short-lived one in dark green. This finding is also apparent in the force vs lifetime scatter plot shown in Figure 40C. In the middle panel at 100 μM Ca^{2+} , the (upper) folded state clearly consists of two clusters, i. e. two distinct states. In comparison, the traces at 10 mM Ca^{2+} and at 0 mM Ca^{2+} do not show this behavior. Neither does this phenomenon appear for the N-terminal domain (compare Figure 39). I will argue in Section 7.6 that the stable (light green) state is the domain's holo-conformation, while the less stable (dark green) state is a transient apo conformation.

The clouds in the scatter plot are slightly curved at very short lifetimes due to limited accuracy of the Hidden Markov classifier in that range.

The Chevron plots shown in Figure 40D also support the previous notions that the average forces of folding and unfolding drop and that the average transition rate constants increase with decreasing calcium concentration. Again, as for the N-terminal domain, the dashed lines are fits to a model including all linker elasticities for the folding branch (red, Equation 35) and a Bell model for the unfolding branch (green, Equation 34).

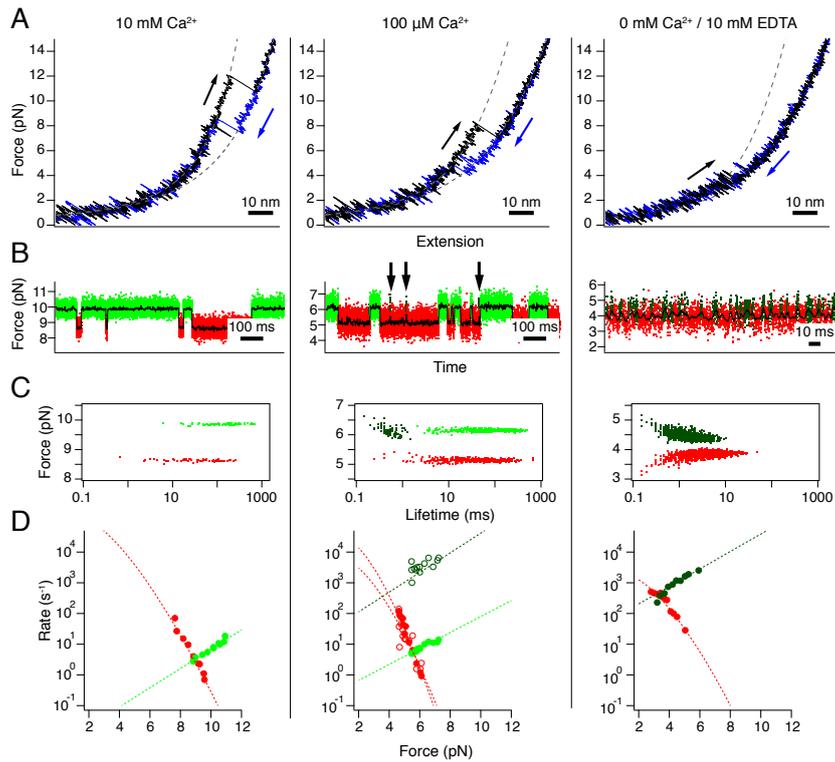


Figure 40: The C-terminal domain of calmodulin at different calcium concentrations. (A) Stretch (black) and relax (blue) cycles at 500 nm/s. (B) Constant trap separation traces. (C) Scatter plot of all detected dwell average forces vs their lifetimes. (D) Chevron plot for unfolding (green colors) and refolding (red) and corresponding fits. The folded state in its calcium-free apo configuration is shown in a darker shade of green. Arrows in the middle panel of (B) show short-lived events incompatible with the lifetime distribution of the light green folded state.

7.4 STOICHIOMETRY OF BINDING TO THE NATIVE STATE

The equilibrium free energy information obtained from the isolated single domain constructs allow for the direct estimation of the number of calcium ions binding to the native conformation. The free energies are shown in Figure 41E, F.

The slope of $-\Delta G_0$ at high calcium concentrations is two, i. e. for every ten-fold increase in calcium concentration, $-\Delta G_0$ increases by $2 \cdot \ln(10) k_B T$, the expected gain in free energy from binding two ions (see Section 5.4). The fit of the energy data to a simplified version of Equation 52

$$-\frac{\Delta G_0}{k_B T} \simeq -\frac{\Delta G_0^{[L]=0}}{k_B T} - \ln \frac{[L]^2}{K_{N1} K_{N2}} \quad (54)$$

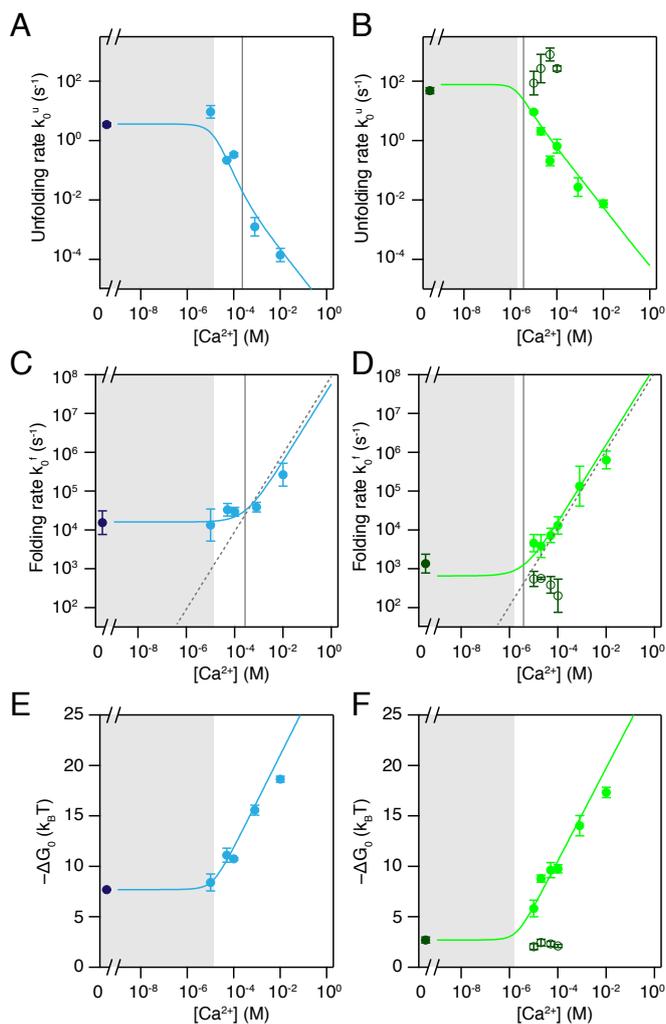


Figure 41: Kinetic and equilibrium parameters of the single domains of calmodulin at different calcium concentrations. Blue colors are values for the N-terminal domain, green colors for the C-terminal domain. Darker shades are for apo conformations. (A), (B) Unfolding rate constants. (C), (D) Folding rate constants. (E), (F) Equilibrium free energy of folding. Continuous lines are fits. The grey shaded area is the calcium range below $\sqrt{K_{N1}K_{N2}}$, where the domains are predominantly calcium free. The vertical grey line is K_{TS1} as determined from the fit. Dashed grey lines in (C), (D) represent the diffusion limited binding rate expected for the binding of a single calcium ion.

not only confirms the number of ions bound to the native state but also allows to determine the stability of the apo state $-\Delta G_0^{[L]=0}$ and the product of the dissociation constants $K_{N1}K_{N2}$. The results are summarized in Table 8. The good agreement between literature values and the experimental results confirm the validity of the assay.

DOMAIN	$-\Delta G_0^{[L]=0}$ [k _B T]		$\log_{10}(K_{N_1}K_{N_2})$	
	EXP.	LIT. [6]	EXP.	LIT. [69]
N-domain	7.7	4.5–6.4	-9.2	-9.8
C-domain	2.5	3.5–4.5	-11.5	-11.5

Table 8: Apo energies and calcium affinities of the isolated domains of calmodulin. The literature values are obtained from bulk chemical denaturation and calcium titration [6, 69].

7.5 STOICHIOMETRY OF BINDING TO THE TRANSITION STATE

Since the force range in which the transitions occur for the single domains is strongly dependent on calcium concentration, the comparison of the folding/unfolding rate constants is much facilitated if zero-force extrapolated rate constants are used. The Chevron plots in [Figure 39D](#) and [Figure 40D](#) show the extrapolations of a Bell model for unfolding and the folding model of [Equation 35](#) as dashed lines. The extrapolated rate constants of this procedure are shown as circles [Figure 41A–D](#).

The unfolding rate constants of both domains remain largely unchanged below the average dissociation constant to the native state $\sqrt{K_{N_1}K_{N_2}}$, where the domains are largely in their calcium free apo-conformation (grey areas in [Figure 41A, B](#)). At higher concentrations, the unfolding rate constants drop for both domains, i. e. a net stabilization due to slowed unfolding.

The folding rate constants show a similar effect. Up to about 10^{-3} M Ca^{2+} , the folding rate constant for the N-terminal domain is unchanged from its apo-value. For higher concentrations of calcium, it increases ([Figure 41C](#)). The folding rate constant of the C-terminal domain is stable up to about $10^{-6} - 10^{-5}$ M Ca^{2+} and increases at higher concentrations ([Figure 41D](#)).

Interestingly, the folding rate constants at high calcium closely follow the trend of a diffusion limited binding reaction of a single calcium ion (dashed grey lines in [Figure 41C, D](#)), indicating that the folding rate constant is limited by calcium binding. Further, the slope of the dependency indicates that the binding of only one ion is sufficient to trigger folding.

The calcium dependent transition rate constants also allow for a more detailed analysis of the path between the folded and unfolded state. With the assumption that the barrier crossing times are fast compared to the binding and unbinding of calcium, we can apply a model based on earlier work by [Fersht](#) and [Bodenreider and Kiefhaber](#) [10, 40]. Following the model of [Section 5.3](#), the folding and unfolding rate constants of the isolated domains can be fitted globally ([Equation 50](#) and [Equation 51](#)). For the calcium affinity to the folded state, values from the literature were used ($\log_{10} K_{N_1} = -4.62$ and $\log_{10} K_{N_2} = -5.17$ for the N-terminal domain; $\log_{10} K_{N_1} = -5.32$ and $\log_{10} K_{N_2} = -6.21$ for the C-terminal

domain [6]). The unfolded state was assumed to be largely calcium-free by assuming a dissociation constant for the unfolded state of 1 M, much higher than the concentrations used here.

The resulting fits to the data are shown in Figure 41A–D. We obtained as fit parameters $\log_{10} K_{TS1} = -3.6 \pm 0.2$ for the N-terminal domain and -5.4 ± 0.1 for the C-terminal domain. The transition state affinity for the second calcium ion was found to be greater than 1 M, corroborating the earlier claim that the transition state has at most one calcium ion bound in the range of concentrations investigated.

Equation 47 allows to calculate the precise probabilities to find the constructs ligand-free, with one calcium ion bound, or with two ions bound, based on the values of K_D as determined here. A plot for the ligand binding stoichiometries of the native state and the transition state of both the N and C-terminal constructs is shown in Figure 42. While the native state is fully saturated at high calcium concentrations, the transition state only binds one calcium ion in the relevant concentration range.

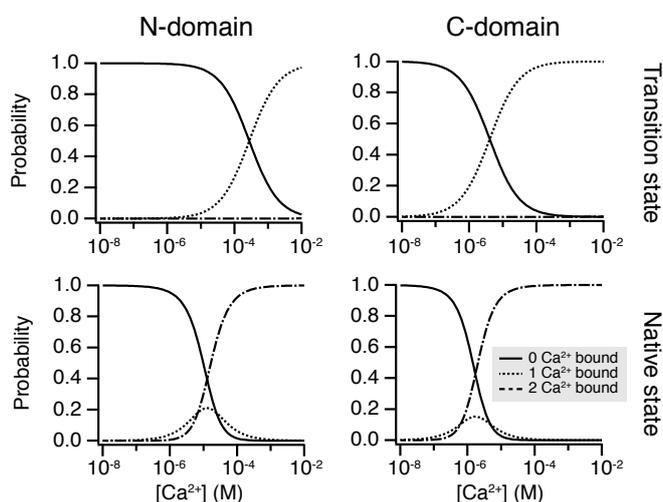


Figure 42: Calcium-dependent stoichiometry of the isolated domains of calmodulin. The probabilities to find the native and the transition state in the unbound state, with one calcium ion bound, or with two calcium ions bound, based on the calcium dissociation constant as determined by experiments.

7.6 TRANSIENT POPULATIONS OF THE APO STATE

The high calcium concentrations used in the earlier single molecule studies of calmodulin [51, 53, 94] have so far prevented the direct observation of a mixture of apo and holo conformations. Already early studies have determined the off-rates of calcium from the single domains of calmodulin and found values in the range of tens to a few hundreds per sec-

ond [68]. With dissociation constants in the micromolar range, the calcium on-rate constants as estimated by $k_{\text{on}} = k_{\text{off}}/K_D$ is approximately $10^8 \text{ M}^{-1} \text{ s}^{-1}$. At a calcium concentration of 10 mM, the resulting on-rate is 10^6 s^{-1} , much faster than the bandwidth of the optical tweezers instrument.

However, at lower concentrations of calcium, the on-rate constants are in the range of the transition rate constants and transient populations of the apo-state can be observed. This becomes apparent for the C-terminal domain at concentrations up to 100 μM (see Figure 40B, middle). Short-lived dwells appear at the extension of the folded state that are incompatible with the lifetime of the calcium-bound folded state. Even though the Hidden Markov classifier (Section 4.1) did not perform well in separating the two green populations (it instead pooled them into the same state with double-exponential lifetime distributions), the two classes could be separated by selecting a threshold based on the force vs lifetime scatter plots such as in Figure 40C, middle. Dwells shorter than the threshold were assigned to the dark green class, dwells longer than the threshold to the light green class. Figure 43B shows the lifetime CDF of the trace shown in Figure 40 middle. The green state is clearly double exponentially distributed. Figure 43C shows the CDFs after separation.

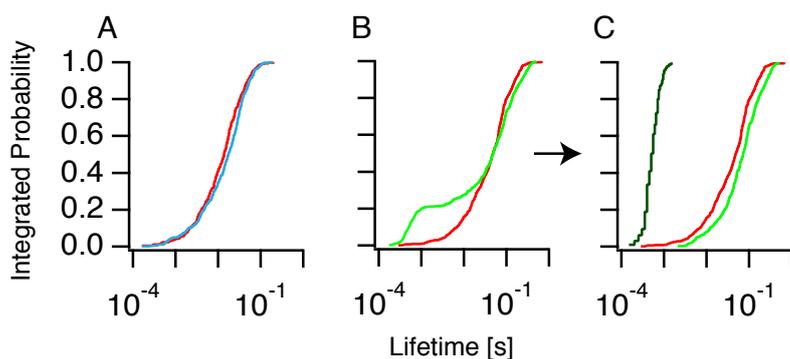


Figure 43: Cumulative lifetime histograms of the single domain constructs at 100 μM Ca^{2+} . Shown are the histograms for the traces of Figure 39, middle (N-terminal domain) in (A) and of Figure 40, middle (C-terminal domain) in (B). While the lifetimes of the unfolded states (red) follow single exponential distributions, the lifetime of the folded state in (B) is double exponentially distributed. The two populations can be separated using a threshold (C).

The two distinct populations can clearly be seen in the force-vs-lifetime scatter plots of Figure 40C, middle (light green cloud and dark green cloud). Several observations support the claim that the dwells colored in dark-green indeed are calcium-free apo conformations of the domain. First, the force-dependent lifetimes of the dark green state closely follow the lifetime of the domain at zero calcium/EDTA conditions. Also the corresponding folding rate constants into the apo population at 100 μM Ca^{2+} corresponds to the folding rate constant at 10 mM EDTA (compare

Figure 40D middle and right). Second, the extrapolated folding and unfolding rate constants at zero load (dark green empty circles in Figure 41B and D) are compatible with the corresponding rate constants of the domain at zero calcium (dark green full circles). Third, also the free energies of folding (Figure 41F) of the domain at zero-calcium and the transient apo state are indistinguishable.

In analogy to the same analysis already performed for full-length calmodulin (Section 6.10.4, also see Section 4.3), it is important to ensure that the apo-states are purely stochastically populated. A test for the stochasticity is based on the distributions of consecutive transitions from the unfolded state to the apo state (Figure 44A) and from the apo state to the holo state (Figure 44B). Both follow geometric distributions (shaded areas), a prerequisite for a stochastic memoryless process. The hypothesis that the observation of the short-lived apo dwells could be an artifact induced by coupling of external vibrations into the measurement can therefore be ruled out.

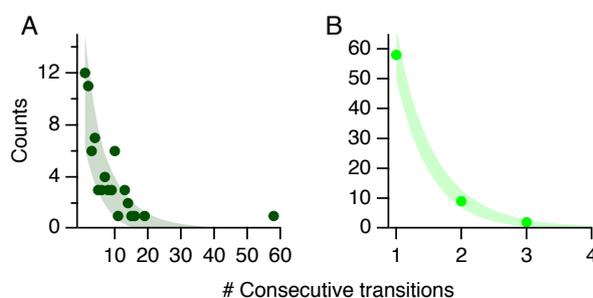


Figure 44: The transitions into the C-domain's apo states are purely stochastic. Shown is data for the isolated C-terminal domain at a calcium concentration of $50 \mu\text{M}$. The distributions for the number of consecutive transitions into the apo-state (A) and the holo-state (B) both follow geometric distributions (shaded areas, 1σ prediction intervals).

7.6.1 When can the transient apo-states be observed?

The transient populations of the apo-state at intermediate concentrations of calcium were clearly visible for the isolated C-terminal domain. Interestingly however, the N-terminal domain did not show such behavior. (See, for example, the scatter plots of Figure 39B, or the clearly single exponential lifetime histogram CDFs of the N-terminal domain shown in Figure 43A). Is there a reason why only one of the domains would show such an effect?

There are indeed reasons why the effect can be present only in one of the domains and not in the other. To observe distinct dwells in the apo-state, the apo unfolding rate constants must be comparable to the binding rate constant of calcium ions. To illustrate this, if the binding of

calcium is fast compared to the lifetime of the apo state, all apo dwells will immediately bind calcium and proceed to the holo state.

Based on the assumption of diffusion-rate limited calcium binding, the on-rate at $100 \mu\text{M Ca}^{2+}$ is $\approx 10^4 \text{ s}^{-1}$. Under load, the lifetime of the apo state now can be in the range where transiently populated apo states unfold before they can bind calcium and are detected just as the short apo dwells described earlier. For example, under the conditions of [Figure 40B](#), middle, the folded state is at a force of about 6 pN. The apo C-domain will unfold with a rate constant of about 2000 s^{-1} at this force ([Figure 40D](#), right). Hence, a fraction of $1 - \exp(-2000 \text{ s}^{-1}/10000 \text{ s}^{-1}) \approx 18\%$ of the dwells in the apo conformation will unfold before calcium is bound and thus be recorded as a short dwell.

In contrast, at $100 \mu\text{M Ca}^{2+}$, the folded state of the N-terminal domain lies at about 7.5 pN ([Figure 39B](#), middle). At this force, the apo-state of the N-terminal domain unfolds at a rate constant of about 400 s^{-1} . Hence, the majority of the dwells in the apo state (in this particular example, $\exp(-400 \text{ s}^{-1}/10000 \text{ s}^{-1}) \approx 96\%$) will bind a calcium ion before they unfold. These dwells will therefore be long holo-dwells.

7.7 DISCUSSION

The titration of calcium closer to the physiologically relevant range revealed a number of insights into the folding behavior of calmodulin, which I will discuss here.

7.7.1 The folding intermediates at low calcium

An intriguing feature of the folding of full-length calmodulin at high calcium concentration was the presence of off-pathway intermediates that were predicted to effectively slow down folding. While all of the intermediates were still present at intermediate calcium concentrations of $100 \mu\text{M}$, under apo conditions, only the folding of the N-terminal domain could be clearly observed. It may come as a surprise that the C-terminal domain, although clearly resolvable in experiments with the isolated domain, could not be resolved in experiments of full-length calmodulin. However, this observation is in accord with earlier findings by [Masino et al.](#) In this study it was found that a folded N-terminal domain can destabilize the C-terminal domain in full-length calmodulin [69].

Separate experiments with other truncation mutants indicated that the construct CaM_{23} (see [Section 6.2](#)) is folded with a free energy of $4 k_B T$ at $100 \mu\text{M Ca}^{2+}$, but unfolded at experimentally resolvable forces above 2-3 pN at 10 mM EDTA conditions. This is also in agreement with findings by [Lakowski et al.](#), who determined the calcium affinities of CaM_{23} and found dissociation constants of $30 \mu\text{M}$ for the binding of the first ion and $> 1000 \mu\text{M}$ for the binding of the second ion [58].

Measurements with CaM₂₃₄ showed that all folded states (F₂₃ and F₃₄) of this construct exist also at calcium concentrations down to 100 μ M. Experiments at lower Ca²⁺ concentrations were not performed with this construct.

As the state F₂₃ was populated in many of the truncation mutants, the equilibrium free energy of folding of this state could be compared between the constructs full-length calmodulin, CaM₂₃ and CaM₂₃₄. At a fixed calcium concentration, its energy was consistent between all constructs, indicating that for all these constructs, the structure of the fold is identical. Interestingly, the free energy of folding of this state scales with the calcium concentration at a slope of 4.5 ± 0.2 k_BT per decade, indicating that two calcium ions bind to F₂₃ in the probed range of concentrations (Figure 45). This stands in contrast to the findings by Lakowski *et al.*, which suggest that in the range between $\approx 30 - 1000$ μ M, only one calcium ion is bound to F₂₃.

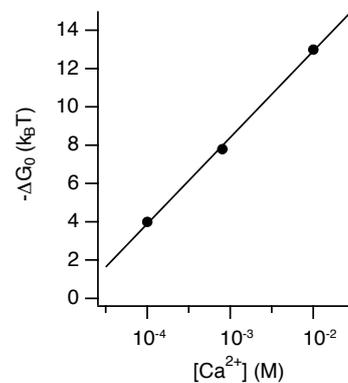


Figure 45: Calcium-dependent free energies of folding for the state F₂₃. Shown are data from the constructs of full-length calmodulin, CaM₂₃ and CaM₂₃₄. The fit is a simple line fit with a slope of 4.5 ± 0.2 k_BT per decade of calcium concentration, indicating that two calcium ions bind to F₂₃ in the probed range.

7.7.2 Accuracy of the determined values

Since the observable range of folding and unfolding rate constants depend on the calcium concentration, it was necessary to rely on models to extrapolate the observed rate constants in order to compare them between different calcium concentrations.

The folding and unfolding rate constants reported here were all obtained by extrapolating force-dependent rate constants down to zero load, a process which obviously depends on the choice of extrapolation model. For the data presented here, a Bell-model was chosen for unfolding, a model including all linker elasticities for folding (see Section 4.6).

Other approaches, such as the application of the elasticity model also to the unfolding branch, or the unfolding model by Dudko *et al.* [33],

yielded similar results. Our extrapolation of the folding and unfolding rate constants for the apo C-terminal domain is close to values reported in temperature jump experiments ($k_{\text{unf}} \approx 300 \text{ s}^{-1}$, $k_{\text{refold}} \approx 4500 \text{ s}^{-1}$, compare [Figure 41B, D](#)) [84].

Thermodynamics and the principle of detailed balance require that the natural logarithm of the ratio of the folding and unfolding rate constants be equal to the total energy of the system [47]. Hence, at zero force,

$$-\frac{\Delta G_0}{k_B T} = \ln \frac{k_{\text{fold}}}{k_{\text{unf}}}. \quad (55)$$

[Figure 46](#) shows the free energies calculated using [Equation 55](#) for the isolated N- and C-terminal domains. At all calcium concentrations, they closely follow the energy values determined under load (continuous lines). The near perfect agreement justifies our choice of rate extrapolation models.

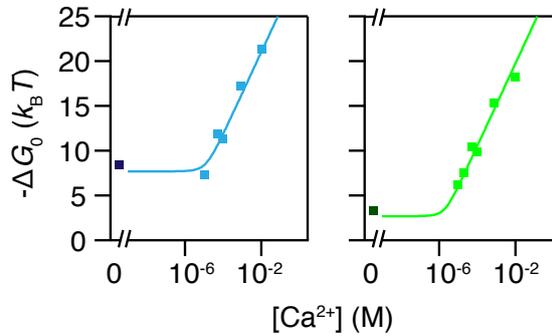


Figure 46: Equilibrium free energies calculated from zero-force extrapolated rates of the isolated N-terminal domain (left) and the isolated C-terminal domain (right) according to [Equation 55](#). Continuous lines are the same as in [Figure 41E–F](#).

As discussed earlier, not only the apo folding/unfolding rate constants at zero force, but also the equilibrium free energies of folding and the calcium affinities to the native state agree with values from the literature.

The determined values for K_{TS1} and K_{TS2} are only well defined under the condition that the barrier for folding and unfolding is the same. Since all measurements are performed under equilibrium conditions, this is a sound assumption. This allowed us to fit the calcium-dependent folding and unfolding rate constants globally and hence obtain a more accurate estimate for K_{TS1} and K_{TS2} . The best fits, shown in [Figure 41A–D](#), all show the general feature of an overall slope of one at calcium concentrations greater than K_{TS1} , indicating that the value of K_{TS2} is much larger than the highest measured calcium concentration. Hence, the evidence that only one calcium ion is bound to the transition state in the model presented in [Section 5.3](#) is strong. However, the exact value of K_{TS1} , which is indicated with vertical grey lines in [Figure 41A–D](#) are less well-defined. Nevertheless, the general conclusions still hold.

7.7.3 Conformational changes

It is important to note that due to limited resolution, we could not observe conformational changes that putatively occur upon binding or release of calcium. Such structural rearrangements have been postulated from X-ray crystallography and were later confirmed by rapid mixing techniques [79]. They have also been postulated in simulations [23]. Since the assay is only sensitive to the end-to-end distance of the molecule, we are inherently blind to structural rearrangements within the molecule. Further, rearrangements with time constants faster than a few hundred microseconds are below the temporal resolution of the instrument.

The model presented in Section 5.3 that was applied to the data contains no structural rearrangements, even though they are expected to occur [79].

Hence, we also cannot rule out possible changes in calcium affinity that might occur during these structural rearrangements. In the same vein, further complications could arise from possible changes of calcium-affinity with applied force. Nevertheless, the excellent agreement of data determined here with the literature suggests that these effects, if present, are small and our general conclusions unaffected.

Hints for possible dynamic structural differences depending on the calcium concentrations can be found in the slopes of the Chevron plots of the single domain constructs (see Figure 39D and Figure 40D). A more detailed presentation of this effect is shown in Figure 47.

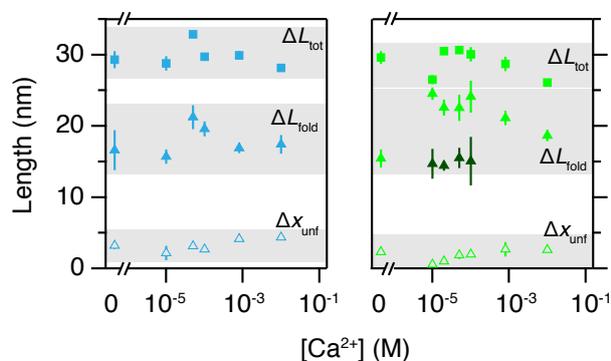


Figure 47: Calcium dependence of the total length and barrier positions of the isolated N-terminal domain (left) and the isolated C-terminal domain (right). Shown are the total measured contour length (squares, ΔL_{tot}), the position of the barrier for folding (closed triangles, ΔL_{fold} , Equation 35) and the position of the barrier for unfolding (open triangles, Δx_{unf} , Equation 34). The values for the short-lived dwells of the C-terminal domain at low calcium concentrations are shown in a darker shade of green.

Especially the C-terminal domain shows an intriguing dependence of the transition barrier position on the calcium concentration. Both the distance to the barrier for unfolding (open green triangles) as well as

the barrier position for folding (full green triangles) show a clear trend. Interestingly, the values under EDTA conditions more closely resemble those at high calcium, where values obtained at low calcium deviate. The short-lived dwells of the folded C-domain at low calcium concentrations, which were attributed to transient population of the apo-state, also agree in the distance to the transition barrier with the real apo state (compare dark green triangles with the filled light green triangle at 0 mM Ca^{2+}).

Long distances to the transition state of unfolding have recently been linked to a more compliant molten globule state of the protein [37]. Nevertheless, even though the observed shift of the transition state position could in principle also arise from structural differences that depend on calcium, I will argue in the following section that the effect is, in fact, expected and can be explained based on simple arguments.

7.7.4 Apparent ligand-dependent shift of the transition state barrier position

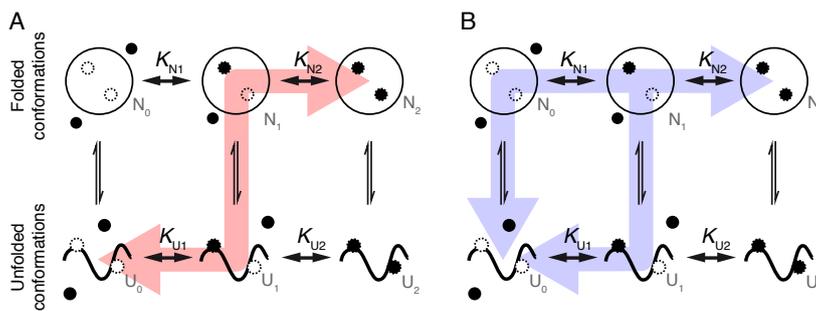


Figure 48: Primary folding/unfolding pathways of the C-terminal domain CaM₃₄. (A) Primary pathway for folding and unfolding at high calcium concentrations (greater than $\approx 100 \mu\text{M}$). (B) Primary pathways for folding and unfolding at low calcium concentrations (less than $\approx 100 \mu\text{M}$).

As illustrated earlier, the primary folding/unfolding pathway of the C-terminal domain at high calcium concentrations populates a transition state where only one of the two calcium ions is bound. Figure 48A shows this pathway in the model network introduced in Section 5.3.

However, the observed transient populations of the apo-conformation (Section 7.6) show that another pathway is also populated at lower calcium concentrations. Here, the folding can also proceed by, first, folding into the apo-conformation and, second, binding of calcium ($U_0 \rightarrow N_0 \rightarrow N_1 \rightarrow \dots$). An illustration of the coexistence of both these pathways at low calcium concentrations is shown in Figure 48B. The principle of detailed balance (cf. Section 4.6.3) requires that, if the folding can occur via the pathway $U_0 \rightarrow N_0 \rightarrow N_1 \rightarrow \dots$, also the unfolding must populate this pathway. The significant population of this folding/unfold-

Binding of calcium to the unfolded state, as postulated by the model, is conceivable: At typical forces, the unfolded polypeptide is stretched to less than 50 % of its contour length. The calcium binding sites are located in loops with no tertiary structure.

ing via U_0 and N_0 at low calcium is one of the reasons for the appearance of the change in the distance to the transition state Δx_{unf} .

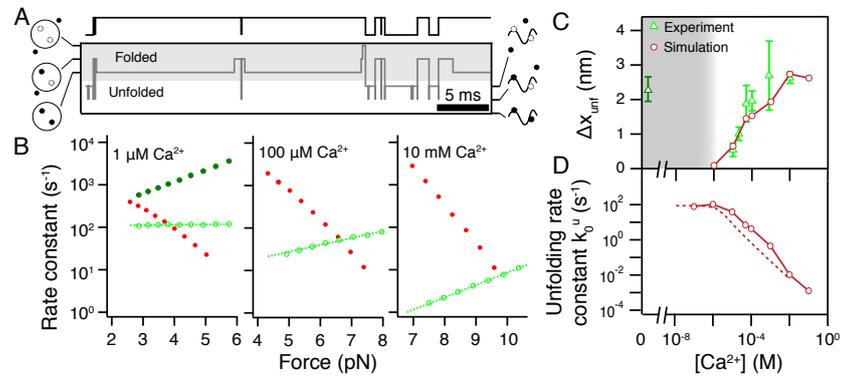


Figure 49: Simulations explain the apparent shift of the transition state position of the long-lived state in the C-terminal domain. (A) Simulated trajectory in the six-state model of Figure 16. (B) Simulated folding (red) and unfolding rate constants for the long-lived state (light green) and the short-lived state (dark green) at different calcium concentrations. The apparent force-dependent slope for the long-lived state clearly depends on the calcium concentration. (C) Transition state position of unfolding. Green: Experimental data (cf. Figure 47), Red: Simulation data. In the shaded region, $[\text{Ca}^{2+}] < \sqrt{K_{N1}K_{N2}}$ and hence, the long-lived state does not get populated. (D) Zero-force extrapolated values from simulations (circles) reproduce the experimental fit (dashed line, cf. Figure 41B).

For simulation details,
see Section C.3.

Simulations of the network shown in Figure 48A–B under load directly reproduce the measured effect. Figure 49A shows an example trajectory for CaM₃₄ at 10 μM Ca²⁺ at an intermediate force bias where both the folded and unfolded conformations are populated. The grey trace shows the populations of the various substates, the black trace is the corresponding trajectory of folded and unfolded state conformations that would be measured in an experiment.

The transition rate constants that were extracted from the simulations are shown in Figure 49B for different calcium concentrations. Notably, in the simulations, as well as in experiments, the slope for unfolding from the canonical long-lived state (light green circles), i. e. Δx_{unf} , decreases at low calcium concentrations.

The results are summarized in Figure 49C–D. The simulation data are—without additional free parameters—able to directly replicate the experimentally determined shift of the transition barrier position Δx_{unf} and also reproduce the experimental results of the calcium-dependent zero-force unfolding rate constant (dashed line and Figure 41B).

An explanation for the effect can be found in the population of the second pathway at low calcium (see Figure 48B). Let's assume the protein is in the folded state with two calcium ions bound (N_2). While at high cal-

cium, the dissociation of one ion (transition into N_1) is rapidly followed by fast rebinding, and only rarely by unfolding, the situation is different at low calcium. Here, the timescale for rebinding is slower than both the time constant for the unfolding process $N_1 \rightarrow U_1$ and the unbinding of the second ion, followed by unfolding ($N_1 \rightarrow N_0 \rightarrow U_0$). Hence, at low calcium, the dissociation of calcium is likely immediately followed by unfolding. In other words, calcium dissociation “triggers” unfolding. Since the dissociation of calcium is independent of force, also the unfolding has to be force-independent, resulting in a flat unfolding branch, that is interpreted as low Δx_{unf} .

Notably, for the N-terminal domain, the intrinsic unfolding rates $N_i \rightarrow U_i$ are always smaller than the binding rate constants for calcium in the experimentally accessible range. Hence, the effect of a calcium-dependent shift of Δx_{unf} is not present, which could be confirmed with according simulations.

Additionally, it is important to note that possible differences between the barriers under load become less important at zero force. The method of using zero-force extrapolated rate constants instead of rate constants under load thus greatly simplifies the description of the system. It has further been argued that the “true” barrier is that at zero load [35].

7.7.5 Influence of the ubiquitin handles

At high calcium concentrations, we could detect a measurable influence of an unfolded ubiquitin domain on the energetics, kinetics and lengths of full-length calmodulin (see Section 6.10.6). A similar effect could be seen for the isolated N-terminal and C-terminal domains at low calcium concentrations. Even though unfolding events of the ubiquitin handle domains were rare, a direct influence on the kinetics was evident, whenever they occurred. Figure 50A shows the force-dependent rate constants for the folding and unfolding of the isolated N-terminal domain at $50 \mu\text{M Ca}^{2+}$. Light colors indicate values for folded ubiquitins. Upon unfolding of one of the ubiquitin domains, the unfolding rate constants are slowed by about two orders of magnitude (arrow, dark colors).

A similar effect could be observed for the isolated C-terminal domain at $50 \mu\text{M Ca}^{2+}$ (Figure 50B). Here, again, the unfolding rate is slowed when a ubiquitin domain is unfolded. In addition, a clear acceleration of the folding rate constants could be observed. For both domains, an unfolded ubiquitin domain increased the free energy of folding by $\approx 6 k_B T$. It should be noted that there is no information as to which of the ubiquitins had unfolded.

As for the ubiquitin-dependent effects in full-length calmodulin, it remains unclear what the microscopic mechanism of the stabilization is. Again, an incorporation of part of the unfolded polypeptide into the calmodulin domains is a possibility. An expected difference in folding length could be observed for the N-terminal domain (Table 9).

A graphical representation of the rate constants involved is shown in the appendix in Figure 53.

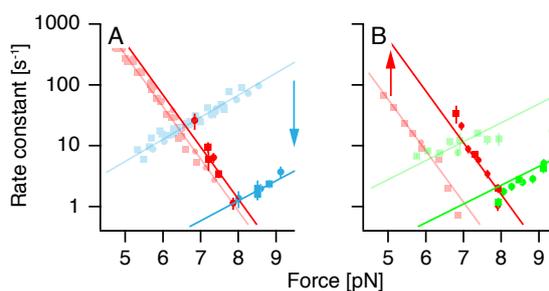


Figure 50: Influence of an unfolded ubiquitin on the kinetics of the isolated domains. (A) Folding (red) and unfolding (blue) rate constants for the N-terminal domain. (B) Folding (red) and unfolding (green) rate constants for the C-terminal domain. Light colors are the kinetics in the case of folded ubiquitins. Dark colors are in the case of an unfolded ubiquitin. Different molecules are shown with different symbols. Lines are intended as guide to the eye. Upon the unfolding of a handle ubiquitin, the rate constants change (arrows).

CONSTRUCT	Unfolded contour length [nm]		N
	Both ubi- quitins folded	One ubi- quitin unfolded	
CaM ₁₂	32.6±0.7	35.4±0.2	3
CaM ₃₄	31.5±1.0	32.0±0.2	2

Table 9: Contour length differences depending on the folding state of the ubiquitin handles.

Control constructs without ubiquitin handles were indistinguishable from constructs with two folded ubiquitin domains. This suggests that when the two ubiquitin handles are properly folded, they do not influence the measurements on the isolated domains of calmodulin.

7.7.6 Open questions

During the measurements, a few more questions arose that remained up to now unanswered. One observation was that at low calcium concentrations, the N-terminal domain often showed long dwells in the unfolded conformation, incompatible with the expected lifetime of the unfolded conformation. We called these dwells “dead” states. Sometimes, attempts to fold into the native conformations were observed (Figure 51). The free energies of different “attempt-states”, however, were not equal even in a single molecule. Molecules could not be simply recovered from “dead” by changing the external biasing force. The observations point to a reversible chemistry, for example, an isomerization of single residues that affects the folding behavior. Interestingly, the effect was not found for the

isolated C-terminal domain. The construct CaM₂₃, however, did show it. Initial guesses, that the isomerization of proline residues could be responsible for the effect, could be ruled out by measurements on the mutant P₄₃M, P₆₆G of the N-terminal domain, where all prolines were replaced with methionine or glycine residues. This construct was energetically less stable than wildtype N-terminal domain, yet remained unaltered in terms of the presence of “dead” states.

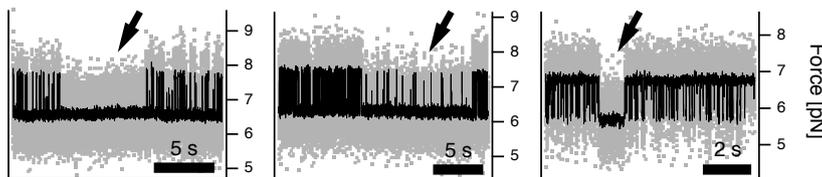


Figure 51: “Dead” states in the isolated N-terminal domain. Three occurrences in the same molecule measured at 50 μ M are shown.

Effects of proline isomerization of residue P₄₃ have been shown with the help of NMR and X-ray crystallography for the protein calbindin D_{9k} [22, 95]. Calbindin D_{9k} is structurally very similar to either of the single domains of calmodulin and was therefore investigated with the optical tweezers assay at various calcium concentrations, in the wildtype, as well as in the P₄₃M variant. Both showed the “dead” states, suggesting that, as for calmodulin, isomerization of proline residues are not responsible for the occurrence of the “dead” state phenomenon.

The effect showed mostly at low calcium concentrations, where the kinetics of the intact domains are rapid and the long “dead” dwells were easily recognizable. As for the wildtype calmodulin measurements at 10 mM Ca²⁺, control experiments showed that replacing the oxygen scavenger system from GLOXY to PCA/PCD did not abolish the “dead” state behavior. Likewise, measurements of a construct without flanking ubiquitin molecules were indistinguishable from measurements with ubiquitins.

In general, the dead states were easily recognizable and left out of the analysis. The data reported here only concern the “live” state of the protein.

7.8 SUMMARY

In conclusion, this chapter presented a study on the folding and unfolding properties of calmodulin close to the biologically relevant range of calcium concentrations. The complex network of six states found at 10 mM Ca²⁺ was also present at concentrations one hundred fold lower. However, at zero calcium, only folding of the N-terminal domain was experimentally detectable. Nevertheless, the kinetic traps F₂₃ and F₁₂₃ are likely not formed at zero calcium conditions.

The folding and unfolding properties of the isolated N- and C-terminal domains were investigated. The N-terminal domain folded in a two-state

fashion with a clear stabilizing effect of calcium. The kinetics of the transition became faster with decreasing calcium concentration. For the accurate determination of transition rate constants, the missed event correction as presented in [Section 4.2.2](#) became crucial.

The C-terminal domain folded in a two-state fashion at very low and very high calcium concentrations. In the intermediate range, short dwells at an apparently folded conformation were detected and attributed to the transient population of the apo state, that was undetectable for the N-terminal domain due to different kinetics of binding.

The calcium-dependent folding and unfolding data of the two domains could be described by a kinetic model ([Section 5.3](#)). It was found that both the folding and unfolding rate constants change with calcium and the data could be globally described in a transition state theory where the transition state only binds one calcium ion and the second ion subsequently binds to the folded conformation.

In the experiments performed here it was still not possible to determine the folding and unfolding behavior at calcium concentrations less than $\approx 10 \mu\text{M}$. Residual calcium from glass surfaces (e. g. in the measurement chamber) tends to accumulate in the buffer and artificially increase its calcium content. The use of low-calcium buffers (see, e. g. [\[63\]](#)) could solve this problem in future experiments.

8

OUTLOOK

The experiments presented here provided insights into the detailed folding mechanism of small ligand-binding multi-domain proteins. Although the precision and stability during the measurements exceeded the performance that is possible using AFM by far, there is still room for improvement. The resolution of the optical tweezers instrument is currently limited by thermal noise fluctuations. While the influence can be reduced by differential detection (Section 3.2), the amount of the noise reduction is determined by the stiffness of the linker connecting the protein to the beads [72].

In the world of DNA linkers, the effective stiffness can be increased by shortening the linker length. However, this is at the expense of smaller bead separations during the experiment and its accompanying issues: increased crosstalk of detection between the two bead channels and increased hydrodynamic coupling between the beads.

The DNA origami technique was used in recent years to design and produce three-dimensional nano-sized objects based on DNA [31, 32]. There is ongoing work to produce handles based on DNA origami structures with predicted persistence lengths in the micrometer range (compared to measured dsDNA persistence lengths in this study of about 25 nm), which are expected to effectively behave as stiff rods.

Still on the technical side, a flow system in the instrument would allow to change buffer conditions online while measuring a single molecule. An approach of this type would make the comparison between different conditions tested on *different* molecules—and its associated issues in comparability of the calibration—unnecessary. Especially for calcium-binding proteins, it would be interesting to observe how they react on a rapid increase or decrease of calcium concentration, that occurs *in vivo* during calcium signaling. While such an experiment could be realized with parallel buffer flows of different calcium concentrations and rapidly moving the dumbbell from one flow to another, the rate at which the concentration varies is still limited by, first, the gradient that builds up between the flows, and, second, by the velocity with which the dumbbell can be moved. Another possibility to quickly increase calcium concentrations would be caged calcium, that can be quickly released into solution with a laser flash [36]. In combination with a flow system, this would allow to—repeatedly—quickly increase the calcium concentration through a laser flash, and then slowly wash away the free calcium again with the flow.

The above mentioned improvements on the setup would also allow more in-depth studies of the ligand binding properties of the calmodulin superfamily. Although calcium sensing is thought to be calmodulin's

primary function, it also possesses low affinity to magnesium. It would be interesting to study the kinetics of magnesium binding itself, and, in addition, to investigate calcium binding, if competing magnesium ions are present.

Another interesting set of experiments could focus on the interactions between calmodulin and target peptides. Since calmodulin interacts with multiple binding partners, a feasible option is to tether the binding partners to the protein and study the binding/unbinding under load. Experiments of this sort have been performed using AFM by Junker and Rief [52] and are directly transferable to a high-resolution optical tweezers experiment.

From a physics perspective, calmodulin is a good model protein to study thermodynamics at the microscopic level. Optical tweezers is an adequate system to study problems of this sort and already has been used to investigate fluctuation theorems [78] or the violation of the second law of thermodynamics on short timescales [102]. In this context it would be particularly interesting to study how the measurements are influenced by the dumbbell pulling apparatus. In the current view, the pulling apparatus solely serves as a probe for the motions of the sample protein. While the dynamics of this probe are fast compared to typical dwelltimes of protein in their metastable states [67], this is not true for the transition times. Hence, as a test for the generality of the method, it would be interesting to see, whether the tethering of comparably huge beads to the protein influences the protein's transition path.

This information is especially valuable for methods, where the energy landscape of the protein is inferred from the position distribution of the beads $p(x)$ by a Boltzmann inversion: $G(x) = -k_B T \ln p(x)$. The *measured* energy landscape in this case is broadened by noise of the measurement apparatus. There are approaches to remove the broadening by a deconvolution procedure [43, 106]. However, the deconvolution approach is based on two major assumptions: First, the measurement probe (i. e. linker and tethered beads) follows the motion of the protein rapidly and second, the response of the probe is purely reactive and does not influence the sample. Experiments with shorter linkers and smaller beads would decrease the friction of the probe and allow to test both assumptions.

Part III
APPENDIX

A

RESOLVING NUMERICAL ISSUES FOR HIDDEN MARKOV MODELS

Hidden Markov analyses of long stretches of data with the methods described in [Section 4.1](#) are numerically challenging. Due to the recursive properties of the equations, especially the forward and backward probabilities often suffer from numerical underflow.

A renormalization technique as presented by [Press et al.](#) can compensate for these numerical issues [82]. In the following discussion, $\tilde{\alpha}(i, t)$ is the renormalized equivalent for $\alpha(i, t)$ and $\tilde{\beta}(i, t)$ is the equivalent for $\beta(i, t)$.

We define a threshold $c < 1$ and renormalize the $\tilde{\alpha}(i, t)$ during the recursion (starting from $t = 0$) ([Equation 2](#)) whenever the value of $\hat{\alpha}(t) = \sum_{i=0}^{N-1} \tilde{\alpha}(i, t)$ becomes smaller than c by setting

$$\tilde{\alpha}_{\text{new}}(i, t) := \frac{\tilde{\alpha}_{\text{old}}(i, t)}{c}, \forall i \in \{0, 1, \dots, N-1\}. \quad (56)$$

The number of such renormalizations until t are counted in $R_{\alpha}(t)$.

Similarly, we renormalize the backward probability during the recursion starting from $t = M-1$ ([Equation 3](#)) whenever $\hat{\beta}(t) = \sum_{i=0}^{N-1} \tilde{\beta}(i, t)$ becomes smaller than c by calculating

$$\tilde{\beta}_{\text{new}}(i, t) := \frac{\tilde{\beta}_{\text{old}}(i, t)}{c}, \forall i \in \{0, 1, \dots, N-1\} \quad (57)$$

and collecting the total count of these renormalization steps $R_{\beta}(t)$. Note that $\alpha(i, t) = \tilde{\alpha}(i, t)c^{R_{\alpha}(t)}$ and $\beta(i, t) = \tilde{\beta}(i, t)c^{R_{\beta}(t)}$.

The overall likelihood $\mathcal{L} = \sum_i \alpha(i, 0)\beta(i, 0) = \sum_i \tilde{\alpha}(i, 0)\tilde{\beta}(i, 0) \cdot c^{R_{\alpha}(0)+R_{\beta}(0)}$ can be written in a renormalized form by letting $\tilde{\mathcal{L}} = \sum_i \tilde{\alpha}(i, 0)\tilde{\beta}(i, 0)$ and $R_L = R_{\alpha}(0) + R_{\beta}(0)$. It follows that $\mathcal{L} = \tilde{\mathcal{L}} \cdot c^{R_L}$. Note that since $\tilde{\mathcal{L}}$ is independent of t , we only need to calculate it once.

Using these definitions, the equations of the Baum-Welch and Forward-Backward algorithms can be rewritten in terms of the renormalized probabilities:

[Equation 4](#) becomes

$$p(n, t) = \frac{\tilde{\alpha}(n, t) \cdot \tilde{\beta}(n, t)}{\sum_{i=0}^{N-1} \tilde{\alpha}(i, t) \cdot \tilde{\beta}(i, t)}. \quad (58)$$

[Equation 7](#) becomes

$$\hat{T}_{ij} = T_{ij} \cdot \frac{\sum_t \frac{\tilde{\alpha}(i, t) \cdot E_j(o_{t+1}) \cdot \tilde{\beta}(j, t+1)}{\tilde{\mathcal{L}}} \cdot c^{R_{\alpha}(t)+R_{\beta}(t+1)-R_L}}{\sum_t \frac{\tilde{\alpha}(i, t) \cdot \tilde{\beta}(i, t)}{\tilde{\mathcal{L}}} \cdot c^{R_{\alpha}(t)+R_{\beta}(t+1)-R_L}}. \quad (59)$$

Equation 5 becomes

$$\hat{E}_i(x) = \frac{\sum_{t|x=o_t} \frac{\tilde{\alpha}(i,t) \cdot \tilde{\beta}(i,t)}{\tilde{\mathcal{L}}} \cdot c^{R_\alpha(t)+R_\beta(t)-R_L}}{\sum_t \frac{\tilde{\alpha}(i,t) \cdot \tilde{\beta}(i,t)}{\tilde{\mathcal{L}}} \cdot c^{R_\alpha(t)+R_\beta(t)-R_L}}. \quad (60)$$

Although some of the factors in the equations above cancel out, it is numerically advantageous to calculate all figures as given here. Note that fractions of the form $(\alpha \cdot \beta) / \mathcal{L}$ are on the order of 1 and do not cause numerical underflow.

It should be noted that in addition to the renormalization approach presented here, other solutions for dealing with numerical underflows are viable, such as calculating the logarithms of $\alpha(i, t)$ and $\beta(i, t)$ instead of their real values.

To account for numerical underflow in the Viterbi algorithm, it is helpful to maximize $\log v(i, t)$ instead of $v(i, t)$ (Equation 8). The products in this case are replaced by sums.

B | NON-EQUILIBRIUM METHODS

Whenever equilibrium measurements are not possible, e. g. if the folding and unfolding kinetics are slow, it is often easier to obtain better statistics when non-equilibrium methods are used. This chapter presents a few.

B.1 DIRECT MEASUREMENT OF RATE CONSTANTS IN PULLING CYCLES

Force-dependent transition rate constants can be measured directly in pulling cycles. Following a proposal by [Oberbarnscheidt et al.](#), a force-ramp also can be described as a series of short constant trap separation experiments, each at a different biasing force [75]. Folding and unfolding rate constants can then be obtained by repeating a series of identical pulling cycle experiments.

To illustrate, consider a series of N stretch cycles on a two-state folding protein. The ramp is divided into a number of bins. The unfolding rate constant k_i in bin i can then be calculated as

$$k_i = \frac{M_i}{N \cdot \Delta t_i}, \quad (61)$$

where M_i is the number of unfolding events in bin i and Δt_i is the time spent while ramping through bin i . The rate constants for refolding can be calculated accordingly.

Technically, it is also possible to apply the method to systems that fold in a non-two state fashion. To achieve this, the identity of the state which the system is in needs to be known at any time point, which is challenging but practicable.

B.2 FLUCTUATION THEOREMS

In recent years, several fluctuation theorems have been developed which can help to obtain equilibrium information about systems that were measured under non-equilibrium conditions. Two notable theorems are the Crooks fluctuation theorem and Jarzynski's equality, which I will briefly present here.

B.2.1 The Crooks fluctuation theorem

Crooks developed a theorem that allows the determination of the equilibrium free energy from the distributions of (non-equilibrium-) work exerted on a system [29]:

$$\frac{p(W_+)}{p(W_-)} = \exp \frac{W - \Delta G_0}{k_B T} \quad (62)$$

Here, $p(W_{\pm})$ are the distributions of the work during folding and unfolding, respectively. Experimentally, the equilibrium free energy ΔG_0 can be obtained by measuring a number of non-equilibrium (force-ramp) folding and refolding experiments (Figure 52A), determining the work (i. e. the area under the curves), and building histograms to obtain an estimate of the distributions (Figure 52B). The value for ΔG_0 is the energy for which the histograms intersect ($W = \Delta G_0$ for $p(W_+) = p(W_-)$).

Note that to be thermodynamically correct, the curves need to be force-distance curves and not force-extension curves [73].

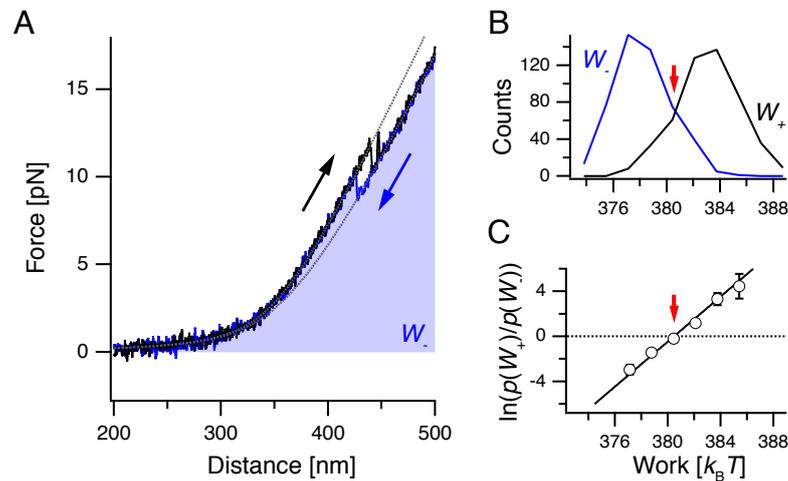


Figure 52: The Crooks fluctuation theorem applied to simulated data of a two-state folder.

(A) Simulated stretch (black) and relax (blue) cycle. The shaded area is the work of relaxation W_- . (B) Histogram of works for 500 realizations of the stretch-and-relax experiment. The histograms intersect at the equilibrium free energy (red arrow). (C) A plot of $\ln \frac{p(W_+)}{p(W_-)}$ in units of $k_B T$ yields a line with slope one. The intersection with the horizontal axis yields the equilibrium free energy ΔG_0 (red arrow).

The energy ΔG_0 , as defined here, still includes the energy for the stretching of the linker and the unfolded polypeptide. This can be easily corrected for by subtracting the area under the fit for the unfolded state from ΔG_0 .

Alternatively, it is helpful to calculate and fit $\ln \frac{p(W_+)}{p(W_-)}$ with a line (Figure 52C). The intersection point with the horizontal axis gives ΔG_0 . The line is predicted to have a slope of $1/k_B T$. Deviations from this value can indicate incorrect calibration of the instrument or the possibly incorrect assumption that the molecule is in the same thermodynamic state before each cycle.

Nevertheless, even though the general Crooks fluctuation theorem breaks down if the molecule is not in the same state (usually, its native confor-

mation) at the beginning of each cycle, a recently published modification to Crooks' formula allows even the description of such cases (e. g. when the initial conformation is occasionally misfolded) [2].

B.2.2 Jarzynski's equality

In some cases, experimental data is only available on either the stretch or the relax part of the cycle (e. g. in the case of very slow folding proteins, a feasible experiment might be to unfold the protein in a force-ramp, go back to zero-force, wait for refolding and repeat).

The equilibrium free energy can still be extracted from this data using Jarzynski's equality [50], which is phenomenologically very similar to the Crooks fluctuation theorem (Equation 62):

$$\exp\left(-\frac{\Delta G_0}{k_B T}\right) = \left\langle \exp\left(-\frac{\Delta W}{k_B T}\right) \right\rangle. \quad (63)$$

Here, $\langle \cdot \rangle$ is an average over many realizations of the same pulling experiment. The equation can be used to analyze either the folding or the unfolding work distributions to obtain the equilibrium free energy ΔG_0 . Since the convergence of the Jarzynski estimator is statistically biased, convergence can be very slow, i. e. a large number of cycles is needed. If a priori information is known about the work distribution, a method to correct for the bias has been proposed recently [78].

C | SIMULATIONS

C.1 BROWNIAN DYNAMICS SIMULATIONS

For the analysis of traces it was often helpful to compare the data with simulations. To this end, an approach based on Brownian Dynamics was chosen. Since the drag on the beads is the slowest component in the dumbbell configuration [67], it was assumed that all other components are in relative equilibrium.

The movement of one bead $x(t)$ can be calculated based on their equation of motion

$$m\ddot{x}(t) + \gamma\dot{x}(t) + kx(t) = \sqrt{2k_B T \gamma} v(t) \quad (64)$$

where $v(t)$ is uncorrelated (white) noise with average zero and an amplitude of one ($\langle v(t) \rangle = 0$, $\langle v(t)v(t') \rangle = \delta(t - t')$) [8]. In practice, $v(t)$ is best chosen by an uncorrelated noise function with gaussian distribution and $\sigma = 1$. The drag coefficient γ is given by the Stokes friction term $\gamma = 6\pi\eta r$, where r is the radius of the bead and η is the viscosity.

Since the motion of the beads is strongly overdamped, the inertial term in Equation 64 can be neglected. In a discretized form, the movement of the individual beads ($i = 1, 2$) is [42, 67]

$$dx_i(t) = \frac{dt}{\gamma} \left(-kx_i(t) + F_{\text{ext}} + \sqrt{2k_B T \gamma / dt} \xi(t) \right). \quad (65)$$

Here, $\xi(t)$ is random number from a generator that produces an uncorrelated series with standard deviation $\sigma = 1$.

The external force F_{ext} is exerted by the linker onto the bead and can be calculated straightforwardly. At a set trap separation z , the extension of the linker is $x_L = z - 2r - x_1 - x_2$. The force can now be computed from a series of eWLC DNA linker (Equation 28) and WLC unfolded polypeptide (Equation 27).

For simulations including folding/unfolding transitions, it is easy to have a Monte Carlo step decide on the current state of the protein and adapt the contour length of the WLC model accordingly:

- For a simulation with N states, first generate the transition rates $k_{ij}(F)$ according to models (e.g. Equation 34, Equation 35). The contour lengths of unfolded polypeptide for each of the states L_p^i , $i = 0 \dots N - 1$ are required for the calculations as are all the necessary parameters of the DNA linker and trap potentials.

- Set the beads to their equilibrium positions. Decide on an initial state i .
- At each time step dt :
 - Calculate the extension of the linker x_L according to $x_L = z - 2r - x_1 - x_2$. The linker can be modeled with an **eWLC** for the DNA in series with a **WLC** of the contour length L_p^i for the unfolded polypeptide. Under the assumption that the DNA and the unfolded polypeptide are in equilibrium, the force F_{ext} acting on both is the same. We can therefore make use of the inversions of [Equation 28](#) and [Equation 27](#) to calculate $x_L = x_D(F_{\text{ext}}) + x_p(F_{\text{ext}})$, which can be inverted to obtain $F_{\text{ext}}(x_L)$.
 - Update the bead positions x_1 and x_2 according to [Equation 65](#).
 - Draw a random number and decide on a new state i based on k_{ij} for the protein. This will in the next step adapt the unfolded contour length.
 - Record the bead positions x_1 and x_2 and their sum $x_1 + x_2$.

Since the analytical inversion of this term is difficult, numerical inversion was chosen here.

C.2 FOLDING TIME SIMULATIONS

The folding times as shown in [Figure 28](#) were calculated assuming a Markov-process of transitions between states with the transition rate matrix k_{ij} , where the diagonal elements were zero: $k_{ii} = 0 \forall i$.

At each simulation step, the dwelltime in the current state i was drawn from an exponential distribution with off-rate $k_{\text{off}} = \sum_{j \neq i} k_{ij}$. The subsequent state l was chosen with probabilities $P_l = \frac{k_{il}}{k_{\text{off}}}$. The dwell times were summed up until the final state was reached.

C.3 SIMULATION OF FOLDING AND UNFOLDING OF LIGAND-BINDING PROTEINS

The network of a single ligand-binding domain (e. g. CaM₁₂ or CaM₃₄) was simulated using the six-state Markov model shown in [Figure 16](#).

First, the 6×6 transition rate matrix k was initialized with the *zero-force* rate constants (cf. [Section 5.1](#) and [Section 5.3](#)):

$$\begin{aligned}
 k_{X_0 \rightarrow X_1} &= 2k_{\text{on}} [L] \\
 k_{X_1 \rightarrow X_0} &= K_{X1} k_{\text{on}} \\
 k_{X_1 \rightarrow X_2} &= k_{\text{on}} [L] \\
 k_{X_2 \rightarrow X_1} &= 2K_{X2} k_{\text{on}} \\
 k_{N_0 \rightarrow U_0} &= k_{\text{u}}^0 \\
 k_{U_0 \rightarrow N_0} &= k_{\text{f}}^0 \\
 k_{N_1 \rightarrow U_1} &= k_{\text{u}}^0 \frac{K_{N1}}{K_{\text{TS1}}} \\
 k_{U_1 \rightarrow N_1} &= k_{\text{f}}^0 \frac{K_{U1}}{K_{\text{TS1}}} \\
 k_{N_2 \rightarrow U_2} &= k_{\text{u}}^0 \frac{K_{N1} K_{N2}}{K_{\text{TS1}} K_{\text{TS2}}} \\
 k_{U_2 \rightarrow N_2} &= k_{\text{f}}^0 \frac{K_{U1} K_{U2}}{K_{\text{TS1}} K_{\text{TS2}}}
 \end{aligned} \tag{66}$$

where $X \in \{N, U\}$. The calcium on-rate k_{on} was assumed to be diffusion limited at a value of $10^8 \text{ M}^{-1}\text{s}^{-1}$. The calcium concentration $[L]$, the zero-force apo folding/unfolding rates $k_{\{f,u\}}^0$ and the equilibrium constants K_{\dots} are experimentally predetermined.

The transition rate constants from the folded states N_i to the unfolded states U_i and back *under load* were then determined following a Bell model ([Equation 34](#)) or a full-energy model ([Equation 35](#)), respectively. It was assumed that the ligand binding and unbinding transitions are independent of force. [Figure 53](#) shows a graphical representation of the rate constants for calmodulin's N-terminal domain at high calcium and the C-terminal domain at low calcium at a biasing force where all states are populated.

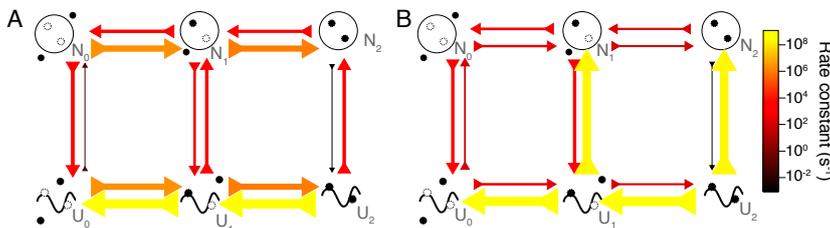


Figure 53: Representation of the transition rate matrix for a domain with two binding sites under load. Line sizes and colors represent the value of the corresponding transition rate constant. (A) Rate constants for the N-terminal domain CaM_{12} at 10 mM Ca^{2+} and an intermediate biasing force. (B) Rate constants for the C-terminal domain CaM_{34} at $1 \mu\text{M Ca}^{2+}$ and an intermediate biasing forces. Here, the transient population of the apo conformation N_0 can be observed.

Using this transition rate matrix k , trajectories can be generated as follows:

At each simulation step, the dwelltime in the current state i was drawn from an exponential distribution with time constant $\tau = \left(\sum_{j \neq i} k_{ij}\right)^{-1}$. The subsequent state l was chosen with probabilities $P_l = k_{il}\tau$.

D | MATERIALS AND METHODS

D.1 CONSTRUCT PRODUCTION

MOLECULAR CLONING The basis for most cloning strategies was a pET28a(+) vector containing two ubiquitins with terminal cysteine residues that serve as attachment sites for single stranded DNA oligos (see below). The ubiquitins serve as spacers to inhibit circular disulfide bond formation if the termini of the insert lie structurally close together. The proteins were inserted via *SACI* and *KPN1* restriction sites between the ubiquitins (see [Section D.3](#)).

Point mutations were performed using a QuikChange kit (Agilent Technologies).

EXPRESSION AND PURIFICATION Proteins were expressed in *E.coli* and initially purified using His₆ Ni-NTA affinity. After this first purification step, the terminal cysteine ends were capped with a ten-fold molar excess of **DTDP**. After a reaction time of about one hour, the reacted protein was separated from unreacted **DTDP** by size exclusion chromatography on a Superdex 200 column. After this step, the protein was frozen with 15 % glycerol at -80 °C. Buffer conditions for size exclusion were 50 mM Tris, 150 mM KCl, pH 8.0.

OLIGO ATTACHMENT The short single-stranded oligo was reacted with 10 mM Tris(2-carboxyethyl)phosphine (**TCEP**) for one hour and the reduced oligos were recovered using EtOH precipitation. After resuspension of the DNA oligo pellet in Tris buffer, the DNA oligos were mixed with the **DTDP** capped protein at a free cysteine:oligo ratio of 1:1 and incubated over night. Finally, the reacted oligo-protein-oligo constructs were separated from the unreacted oligos, unreacted protein, oligo dimer and protein with only one oligo by size exclusion chromatography on a Superdex 200 column. Again, oligo-protein chimeras could be frozen with 15 % glycerol at -80 °C.

The sequence for the oligo was

OLIGO GGCAGGGCTGACGTTCAACCAGACCAGCGAGTCG-Thiol

HANDLE PRODUCTION The dsDNA linkers were produced using **PCR** on a λ -phage template with a 1:1 mixture of biotin and anti-digoxigenin modified sense primers and an anti-sense primer containing an abasic site. The abasic site produced a single stranded overhang that is complementary to the protein-bound oligo. Handles were purified using a Qiagen PCR purification kit and EtOH precipitation. Sequences were

* designates the site of modifications

X designates an abasic site modification

LINKER PRIMER SENSE *-GGCGA*CTGG*CGTTGATTTG

LINKER PRIMER ANTISENSE C GACTCGCTGGTCTGGTTGAACGTCAGCCCT-GCCXCCTGCCCGGCTCTGGACAGG

D.2 MEASUREMENT PROCEDURE

SAMPLE PREPARATION All measurements were performed in 50 mM Tris, 150 mM KCl, pH 8.0 with various amounts of CaCl₂ or EDTA.

First, the protein-oligo chimera construct was mixed with the dsDNA linkers and subsequently mixed with suspensions of silica microspheres (Polysciences, diameter 1 μ m) that were in-house covalently coated with a fluorescent label and anti-digoxigenin Fab fragments, and incubated for a few minutes. Finally, the resulting suspension was mixed with purchased 1 μ m streptavidin-coated silica microspheres (Bangs laboratories) and an oxygen scavenger system was added (typically, in the case of GLOXY, 26 U/ml glucose oxidase, 17000 U/ml catalase, 0.65 % glucose).

The final mixture was filled into a measurement chamber made of a cover slip that was fixed to a microscope slide using Nescofilm. The chamber was previously passivated by short incubation with 10 mg/ml Bovine Serum Albumin (BSA). The ends of the chamber were finally sealed with silicone grease.

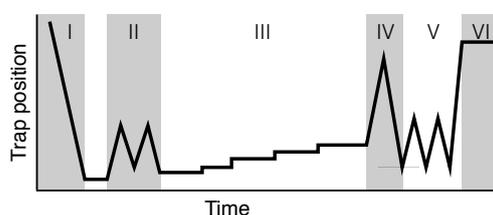


Figure 54: Pulling protocol of a typical measurement. In section (I) Initial approach, (II) Stretch-and-relax, (III) Constant distance, (IV) Forced breaking of tether, (V) Untethered cycles, (VI) Calibration for trap stiffness and sensitivity.

MEASUREMENT DETAILS All signals (displacement of the mobile bead from the trap center, displacement of the fixed bead from its trap center and mobile trap position) were typically recorded at 20 kHz bandwidth, which is well above the typical corner frequencies due to hydrodynamic damping.

Figure 54 shows the pulling protocol of a typical measurement. Section (I) shows the initial approach, where the beads are initially brought into close proximity and baselines are recorded (for a detailed description, see below). Section (II) shows stretch-and-relax cycles. Here, the overall behavior is observed and the integrity of the tether is assured. In section (III), measurements at constant trap separation are recorded. Af-

ter a sufficient amount of data has been collected, the tether is forcibly broken in area (IV). The tethers usually break at high loads, presumably at the hybridization site between oligo and dsDNA handle. Tethers that did not break usually indicated improper attachment and were discarded. In section (V), more cycles are pulled, now without a formed tether. If the corrected signals of both the mobile and fixed bead deflection are *not* zero across the entire range of trap position, the signals have drifted and the results of the measurement should be regarded with caution. Finally, in section (VI), the beads are separated and calibrated for sensitivity and stiffness (see below).

INITIAL APPROACH CALIBRATION For an experiment, the mobile trap was placed 5 μm from the fixed trap (for the experimental setup, see [Figure 5](#) and one bead was trapped in each. The fluorescent label on the anti-digoxigenin coated beads was excited in epi-fluorescence using a 532 nm laser and helped to make sure that the two trapped beads carried different functionalizations.

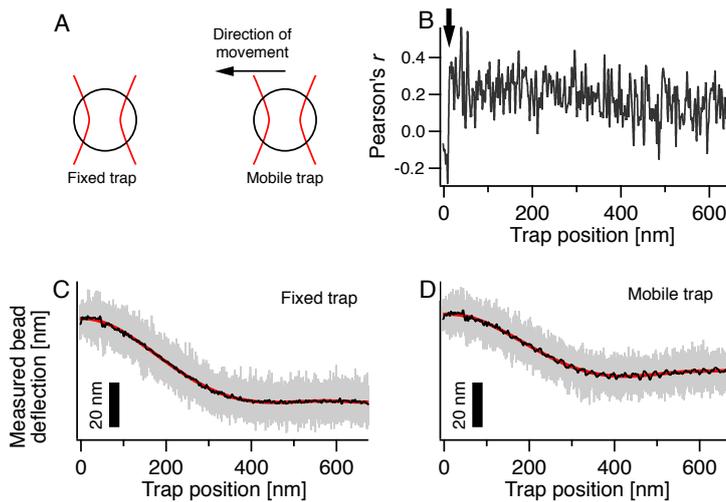


Figure 55: Experimental calibration during the initial bead approach. (A) Sketch of the initial bead approach. (B) Pearson's r coefficient of between the signals of the fixed and mobile trap. The signals are such that they are positive when the beads move towards each other. A sudden drop in r occurs when the bead surfaces touch (arrow). (C), (D) Measured bead displacements of the fixed and mobile trap, respectively (grey trace, filtered: black trace) and a polynomial fit (red trace) during the initial bead approach.

The mobile bead was then moved with constant velocity toward the fixed bead until their surfaces touched ([Figure 55A](#)). The bead contact point could be determined by continuously calculating Pearson's r correlation coefficient between the mobile (x_{mob}) and fixed signal (x_{fix}) during the initial bead approach (in a windowed manner). r measures the

Note that x_{mob} and x_{fix} are defined such that they are positive if the beads are deflected closer to each other.

correlation between two signals and is defined by

$$r = \frac{\sum_i (x_{\text{mob},i} - \langle x_{\text{mob}} \rangle) (x_{\text{fix},i} - \langle x_{\text{fix}} \rangle)}{\sum_i (x_{\text{mob},i} - \langle x_{\text{mob}} \rangle)^2 \sum_i (x_{\text{fix},i} - \langle x_{\text{fix}} \rangle)^2}. \quad (67)$$

At close distances, r is positive, probably owing to hydrodynamic interactions of the beads [71]. When the bead surfaces touch, the bead motions become symmetric and r suddenly turns negative (arrow in Figure 55B). The recorded trap distance is the actual zero-point of measurements.

Figure 55C–D shows the recorded signal of the fixed and mobile bead deflection during this initial approach. Since the mobile beam is moved at a speed of 500 nm, the mobile bead is not significantly moved from its equilibrium position due to hydrodynamic drag. However, probably due to inexact alignment of the instrument, overlap of the light beams or interferences, the signals of both the mobile and the fixed bead vary with the trap distance. To correct for this, the signals are set to zero by subtracting a trap distance dependent polynomial from each (red lines in Figure 55C–D).

*The Stokes friction
 $\eta = 6\pi\eta rv$ is on the
order of femtonewtons.*

CALIBRATION OF SENSITIVITY AND STIFFNESS The two factors that are needed to convert the voltage signal from the bead position detectors into a force signal are the sensitivity (nanometers per volt) and the trap stiffness (piconewtons per nanometer) for each of the trap. All parameters can be obtained simultaneously by using a measurement protocol introduced by Tolić-Nørrelykke et al. [97]. Since the results of the method are influenced by crosstalk between the two beams, correction factors were determined earlier. The correction accounts for the differences between *measured* trap stiffnesses and sensitivities, when two beads are calibrated simultaneously and the *true* values when they are calibrated in isolation.

CORRECTION FOR PARASITIC SIGNALS FROM THE AOD The choice of an AOD over a piezo-driven mirror to steer one of the beams has the distinct advantage of speed. Since there are no mechanically movable parts, the trap positions can be quickly altered. However, at the detection stage, the AOD-modified beam shows artificial signals that do not appear in a piezo mirror based setup. In the measurement process of Figure 54, section (V) is used to correct for this effect.

Figure 56A shows the signals on the two detectors when there are two beads trapped but no tether has formed. While the mobile trap shows an oscillating dependence on the trap distance, the fixed signal is seemingly unaffected.

This oscillatory behavior of the mobile trap signal is an artifact that appears at the detection stage and does not reflect a “real” movement of the bead. This can be verified using a tethered construct (e. g. dsDNA handles, as shown in Figure 56B). If the mobile bead really followed the

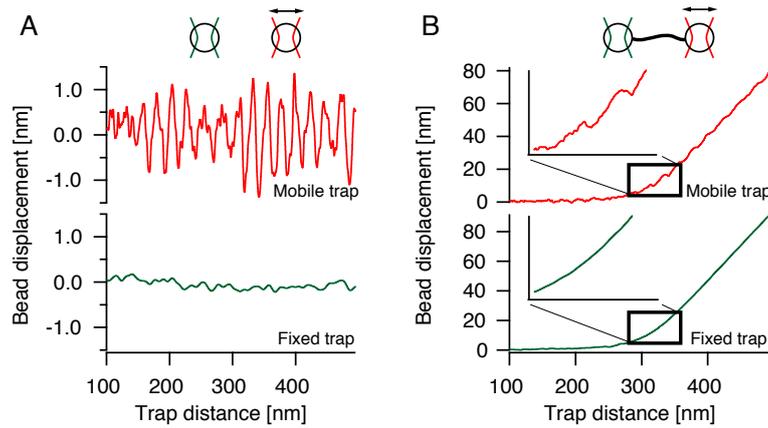
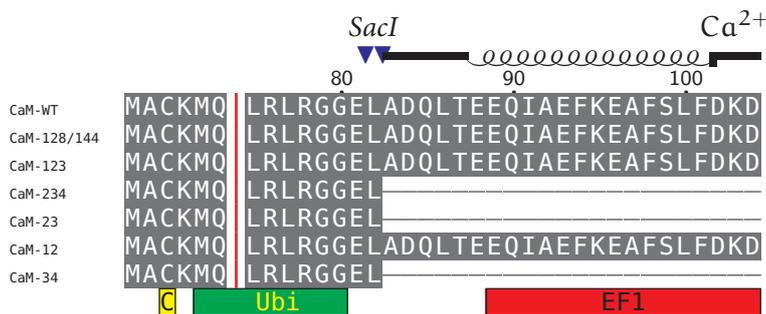


Figure 56: Background signal from the AOD. (A) Measured distance-dependent signals of the mobile and fixed trap with no formed tether. (B) Measured distance-dependent signals of the mobile and fixed trap with a formed tether of dsDNA.

oscillatory path that is indicated in the red trace, the tether would transduce the movement onto the other bead, which should then exhibit the same motion. However, no such behavior can be seen, indicating that the oscillatory signal is an artifact in the detection process. It has been suggested that the oscillations arise from unwanted backreflections inside the AOD crystal [99]. Nevertheless, it is straightforward to remove the parasitic signal by calibration, e. g. by recording “empty” stretch-and-relax cycles as suggested in the measurement protocol (Figure 54), section (V). To this end, the signals shown in Figure 56A were used as background signal information and subtracted from all measured data.

D.3 SEQUENCES



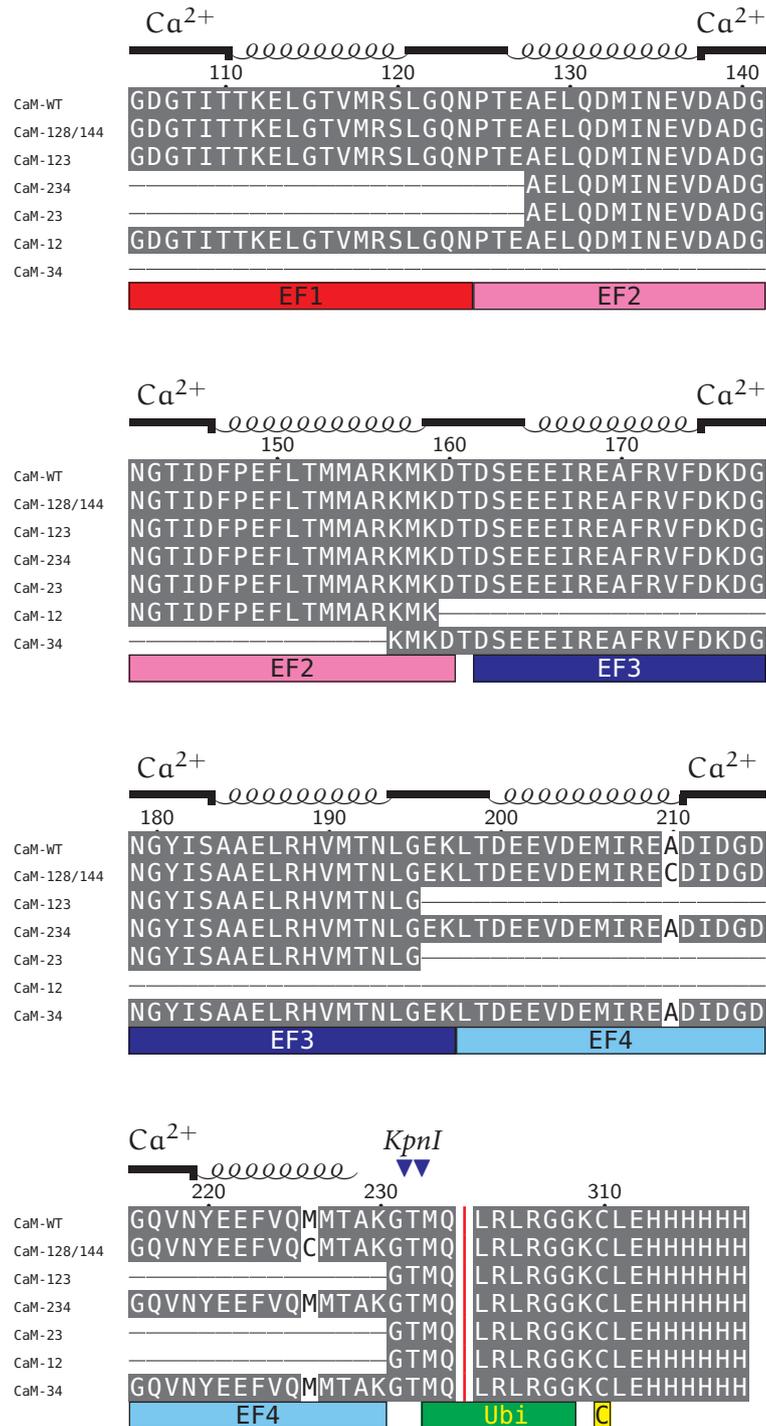


Figure 57: Sequences of the calmodulin constructs. Indicated are helical content and the locations of the EF hands according to the UNIPROT information of P62158.

BIBLIOGRAPHY

- [1] Colin Echeverría Aitken, R Andrew Marshall, and Joseph D Puglisi. An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments. *Biophysical Journal*, 94(5):1826–1835, March 2008. (Cited on page 61.)
- [2] Anna Alemany, Alessandro Mossa, Ivan Junier, and Felix Ritort. Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. *Nature Physics*, 8(8):1–7, July 2012. (Cited on page 99.)
- [3] Peter C Anthony, Christian F Perez, Cuauhtémoc García-García, and Steven M Block. Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proceedings of the National Academy of Sciences of the United States of America*, January 2012. (Cited on page 9.)
- [4] A Ashkin and J M Dziedzic. Optical trapping and manipulation of viruses and bacteria. *Science*, 235(4795):1517–1520, March 1987. (Cited on page 9.)
- [5] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41:164–171, 1970. (Cited on page 20.)
- [6] Peter M Bayley, W A Findlay, and Stephen R Martin. Target recognition by calmodulin: dissecting the kinetics and affinity of interaction using short peptide sequences. *Protein science : a publication of the Protein Society*, 5(7):1215–1228, July 1996. (Cited on pages 7, 77, and 78.)
- [7] G I Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, May 1978. (Cited on page 32.)
- [8] K Berg-Sørensen and Henrik Flyvbjerg. The colour of thermal noise in classical Brownian motion: a feasibility study of direct experimental observation. *New Journal of Physics*, 7:38, 2005. (Cited on page 101.)
- [9] Felix Berkemeier, Morten Bertz, Senbo Xiao, Nikos Pinotsis, Matthias Wilmanns, Frauke Gräter, and Matthias Rief. Fast-folding alpha-helices as reversible strain absorbers in the muscle protein myomesin. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34):14139–14144, August 2011. (Cited on page 17.)

- [10] Christophe Bodenreider and Thomas Kiefhaber. Interpretation of protein folding psi values. *Journal of molecular biology*, 351(2):393–401, August 2005. (Cited on pages 39 and 77.)
- [11] Volker Bormuth, Anita Jannasch, Marcel Ander, Carlos M van Kats, Alfons van Blaaderen, Jonathon Howard, and Erik Schäffer. Optical trapping of coated microspheres. *Optics Express*, 16(18):13831–13844, September 2008. (Cited on page 9.)
- [12] Thomas Bornschlögl and Matthias Rief. Single molecule unzipping of coiled coils: sequence resolved stability profiles. *Physical Review Letters*, 96(11):118102, 2006. (Cited on page 3.)
- [13] C Bouchiat, M D Wang, J Allemand, T Strick, Steven M Block, and V Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophysical Journal*, 76(1 Pt 1):409–413, 1999. (Cited on page 31.)
- [14] Melanie Brunnbauer. *Mechanische Untersuchungen heterodimerer Kinesin-2 Motoren*. PhD thesis, Technische Universität München, March 2011. (Cited on page 9.)
- [15] Carlos Bustamante, J F Marko, E D Siggia, and Steven B Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, September 1994. (Cited on page 30.)
- [16] Yi Cao and Hongbin Li. Dynamics of protein folding and cofactor binding monitored by single-molecule force spectroscopy. *Biophysical Journal*, 101(8):2009–2017, October 2011. (Cited on page 3.)
- [17] Yi Cao, Kai Shih Er, Rakesh Parhar, and Hongbin Li. A Force-Spectroscopy-Based Single-Molecule Metal-Binding Assay. *Chemphyschem : a European journal of chemical physics and physical chemistry*, 10(9-10):1450–1454, July 2009. (Cited on page 3.)
- [18] Mariano Carrion-Vazquez, Hongbin Li, Hui Lu, Piotr E Marszalek, Andres F Oberhauser, and Julio M Fernández. The mechanical stability of ubiquitin is linkage dependent. *Nature Structural Biology*, 10(9):738–743, 2003. (Cited on page 3.)
- [19] Ciro Cecconi, Elizabeth A Shank, Carlos Bustamante, and Susan Marqusee. Direct observation of the three-state folding of a single protein molecule. *Science*, 309(5743):2057–2060, September 2005. (Cited on pages 4 and 9.)
- [20] Ciro Cecconi, Elizabeth A Shank, Frederick W Dahlquist, Susan Marqusee, and Carlos Bustamante. Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J*, 37(6):729–738, July 2008. (Cited on page 13.)

- [21] R Chattopadhyaya, W E Meador, A R Means, and F A Quioco. Calmodulin structure refined at 1.7 Å resolution. *Journal of molecular biology*, 228(4):1177–1192, December 1992. (Cited on page 6.)
- [22] W J Chazin, J Kördel, T Drakenberg, Eva Thulin, P Brodin, T Grundström, and S Forsén. Proline isomerism leads to multiple folded conformations of calbindin D9k: direct evidence from two-dimensional ¹H NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 86(7):2195–2198, April 1989. (Cited on page 89.)
- [23] Yng-Gwei Chen and Gerhard Hummer. Slow conformational dynamics and unfolding of the calmodulin C-terminal domain. *Journal of the American Chemical Society*, 129(9):2414–2415, March 2007. (Cited on pages 7 and 84.)
- [24] D Chin and A R Means. Calmodulin: a prototypical calcium sensor. *Trends in cell biology*, 10(8):322–328, August 2000. (Cited on page 6.)
- [25] John D Chodera, Phillip J Elms, Bettina Keller, Christian M Kaiser, Aaron Ewall-Wice, Susan Marqusee, Carlos Bustamante, and Nina Singhal Hinrichs. Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. *arXiv.org*, cond-mat.stat-mech, August 2011. (Cited on pages 20 and 24.)
- [26] H S Chung, K McHale, J M Louis, and W A Eaton. Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times. *Science*, 335(6071):981–984, February 2012. (Cited on page 3.)
- [27] Hoi Sung Chung and Andrei Tokmakoff. Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy. *Proteins: Structure, Function, and Bioinformatics*, 72(1):474–487, April 2008. (Cited on page 65.)
- [28] David E Clapham. Calcium Signaling. *Cell*, 131(6):1047–1058, December 2007. (Cited on pages 5 and 71.)
- [29] G E Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 60(3):2721–2726, September 1999. (Cited on page 98.)
- [30] S C Crouzy and Fred J Sigworth. Yet another approach to the dwell-time omission problem of single-channel analysis. *Biophysical Journal*, 58(3):731–743, September 1990. (Cited on page 26.)

- [31] Hendrik Dietz, Shawn M Douglas, and William M Shih. Folding DNA into twisted and curved nanoscale shapes. *Science*, 325(5941):725–730, August 2009. (Cited on page 91.)
- [32] Shawn M Douglas, Hendrik Dietz, Tim Liedl, Björn Högberg, Franziska Graf, and William M Shih. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, 459(7245):414–418, May 2009. (Cited on page 91.)
- [33] Olga K Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical Review Letters*, 96(10):108101, March 2006. (Cited on pages 33 and 82.)
- [34] Olga K Dudko, Gerhard Hummer, and Attila Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41):15755–15760, October 2008. (Cited on page 33.)
- [35] Olga K Dudko, Thomas G W Graham, and Robert B Best. Locating the Barrier for Folding of Single Molecules under an External Force. *Physical Review Letters*, 107(20):208301, November 2011. (Cited on page 87.)
- [36] G C Ellis-Davies, J H Kaplan, and R J Barsotti. Laser photolysis of caged calcium: rates of calcium release by nitrophenyl-EGTA and DM-nitrophen. *Biophysical Journal*, 70(2):1006–1016, February 1996. (Cited on page 91.)
- [37] Phillip J Elms, John D Chodera, Carlos Bustamante, and Susan Marqusee. The molten globule state is unusually deformable under mechanical force. *Proceedings of the National Academy of Sciences of the United States of America*, 109(10):3796–3801, March 2012. (Cited on pages 9 and 85.)
- [38] E Evans and K Ritchie. Dynamic strength of molecular adhesion bonds. *Biophysical Journal*, 72(4):1541–1555, 1997. (Cited on page 32.)
- [39] Jennifer L Fallon and Florante A Quioco. A closed compact structure of native Ca(2+)-calmodulin. *Structure*, 11(10):1303–1307, October 2003. (Cited on pages 6 and 65.)
- [40] Alan R Fersht. Phi value versus psi analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 101(50):17327–17328, December 2004. (Cited on pages 39 and 77.)
- [41] G David Forney Jr. The Viterbi Algorithm. *Proceedings of the IEEE*, 61:268–278, March 1973. (Cited on page 21.)

- [42] J Christof M Gebhardt. *Energielandschaften einzelner Proteine gemessen mit der optischen Pinzette*. PhD thesis, Technische Universität München, Technische Universität München, 2009. (Cited on pages 7, 12, 14, 29, and 101.)
- [43] J Christof M Gebhardt, Thomas Bornschlögl, and Matthias Rief. Full distance-resolved folding energy landscape of one single protein molecule. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5):2013–2018, February 2010. (Cited on pages 4, 33, 52, 69, and 92.)
- [44] William J Greenleaf, Michael T Woodside, Elio A Abbondanzieri, and Steven M Block. Passive all-optical force clamp for high-resolution laser trapping. *Physical Review Letters*, 95(20):208102, November 2005. (Cited on page 11.)
- [45] Zy Guo and D Thirumalai. Kinetics of protein-folding - Nucleation Mechanism, Time Scales and Pathways. *Biopolymers*, 36(1):83–102, 1995. (Cited on page 69.)
- [46] D B Heidorn and J Trewhella. Comparison of the crystal and solution structures of calmodulin and troponin C. *Biochemistry*, 27(3): 909–915, February 1988. (Cited on pages 6 and 65.)
- [47] J Howard. *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer Associates, Publishers, 2001. (Cited on pages 63 and 83.)
- [48] Josep Maria Huguet, Cristiano V Bizarro, Núria Forns, Steven B Smith, Carlos Bustamante, and Felix Ritort. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 107(35):15431–15436, August 2010. (Cited on page 9.)
- [49] Anita Jannasch, Mahamdeh Mahamdeh, and Erik Schäffer. Inertial Effects of a Small Brownian Particle Cause a Colored Power Spectral Density of Thermal Noise. *Physical Review Letters*, 2011. (Cited on page 9.)
- [50] C Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*, 78:2690–2693, April 1997. (Cited on page 99.)
- [51] Jan Philipp Junker. *Ligandenabhängige Gleichgewichtsfluktuationen einzelner Proteine*. PhD thesis, Technische Universität München, August 2009. (Cited on pages 39, 48, and 78.)
- [52] Jan Philipp Junker and Matthias Rief. Single-molecule force spectroscopy distinguishes target binding modes of calmodulin. *Proceedings of the National Academy of Sciences of the United States of America*, 106(34):14361–14366, August 2009. (Cited on pages 3, 7, and 92.)

- [53] Jan Philipp Junker, Fabian Ziegler, and Matthias Rief. Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science*, 323(5914):633–637, 2009. (Cited on pages 3, 7, 48, 72, and 78.)
- [54] H Kawasaki, S Nakayama, and R H Kretsinger. Classification and evolution of EF-hand proteins. *Biomaterials*, 11(4):277–295, December 1998. (Cited on pages 5 and 6.)
- [55] R H Kretsinger and C E Nockolds. Carp muscle calcium-binding protein. II. Structure determination and general description. *J Biol Chem*, 248(9):3313–3326, May 1973. (Cited on page 5.)
- [56] M Kruithof and J van Noort. Hidden Markov analysis of nucleosome unwrapping under force. *Biophysical Journal*, 96(9):3708–3715, May 2009. (Cited on pages 17 and 18.)
- [57] H Kuboniwa, N Tjandra, S Grzesiek, H Ren, C B Klee, and A Bax. Solution structure of calcium-free calmodulin. *Nature Structural Biology*, 2(9):768–776, September 1995. (Cited on page 6.)
- [58] TM Lakowski, GM Lee, M Okon, RE Reid, and LP McIntosh. Calcium-induced folding of a fragment of calmodulin composed of EF-hands 2 and 3. *Protein science : a publication of the Protein Society*, 16(6):1119–1132, 2007. (Cited on pages 6, 48, 53, 81, and 82.)
- [59] Markita P Landry, Patrick M McCall, Zhi Qi, and Yann R Chemla. Characterization of photoactivated singlet oxygen damage in single-molecule optical trap experiments. *Biophysical Journal*, 97(8):2128–2136, October 2009. (Cited on pages 14 and 61.)
- [60] Thomas J Lane, Gregory R Bowman, Kyle Beauchamp, Vincent A Voelz, and Vijay S Pande. Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long MD Trajectories. *Journal of the American Chemical Society*, 133(45):18413–18419, November 2011. (Cited on pages 3 and 69.)
- [61] Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, October 2011. (Cited on page 3.)
- [62] Sara Linse. *Biophysical Studies of the Cooperativity of Calcium Binding in Calmodulin and Calbindin D9k*. PhD thesis, Lund University, Lund University, May 1993. (Cited on page 37.)
- [63] Sara Linse. Calcium Binding to Proteins Studied via Competition with Chromophoric Chelators. *Methods in Molecular Biology*, vol. 173: *Calcium-Binding Protein Protocols, Volume 2*, pages 15–24, 2002. (Cited on page 90.)

- [64] Sara Linse, A Helmersson, and S Forsén. Calcium binding to calmodulin and its globular domains. *J Biol Chem*, 266(13):8050–8054, May 1991. (Cited on page 7.)
- [65] Sara Linse, Eva Thulin, and Peter Sellers. Disulfide bonds in homo- and heterotimers of EF-hand subdomains of calbindin D9k: Stability, calcium binding and NMR studies. *Protein science : a publication of the Protein Society*, 2:985–1000, 1993. (Cited on page 7.)
- [66] A Malmendal, J Evenäs, S Forsén, and M Akke. Structural dynamics in the C-terminal domain of calmodulin at low calcium levels. *Journal of molecular biology*, 293(4):883–899, November 1999. (Cited on page 7.)
- [67] M Manosas, Jin-Der Wen, Pan T X Li, Steven B Smith, Carlos Bustamante, Ignacio Tinoco, and Felix Ritort. Force unfolding kinetics of RNA using optical tweezers. II. Modeling experiments. *Biophysical Journal*, 92(9):3010–3021, May 2007. (Cited on pages 29, 92, and 101.)
- [68] Stephen R Martin, A Andersson Teleman, Peter M Bayley, T Drakenberg, and S Forsén. Kinetics of calcium dissociation from calmodulin and its tryptic fragments. A stopped-flow fluorescence study using Quin 2 reveals a two-domain structure. *European journal of biochemistry / FEBS*, 151(3):543–550, September 1985. (Cited on pages 7 and 79.)
- [69] L Masino, Stephen R Martin, and Peter M Bayley. Ligand binding and thermodynamic stability of a multidomain protein, calmodulin. *Protein science : a publication of the Protein Society*, 9(8):1519–1529, August 2000. (Cited on pages 7, 62, 72, 77, and 81.)
- [70] Sean A McKinney, Chirlmin Joo, and Taekjip Ha. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal*, 91(5):1941–1951, September 2006. (Cited on page 17.)
- [71] JC Meiners and SR Quake. Direct measurement of hydrodynamic cross correlations between two particles in an external potential. *Physical Review Letters*, 82(10):2211–2214, 1999. (Cited on page 108.)
- [72] Jeffrey R Moffitt, Yann R Chemla, David Izhaky, and Carlos Bustamante. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proceedings of the National Academy of Sciences of the United States of America*, 103(24):9006–9011, June 2006. (Cited on pages 11, 12, and 91.)
- [73] Alessandro Mossa, Sara de Lorenzo, Josep Maria Huguet, and Felix Ritort. Measurement of work in single-molecule pulling experiments. *The Journal of chemical physics*, 130(23):234116, June 2009. (Cited on page 98.)

- [74] S Nakayama and R H Kretsinger. Evolution of the EF-hand family of proteins. *Annu Rev Biophys Biomol Struct*, 23:473–507, 1994. (Cited on page 5.)
- [75] Leoni Oberbarnscheidt, Richard Janissen, and Philipp Oesterhelt. Direct and Model Free Calculation of Force-Dependent Dissociation Rates from Force Spectroscopic Data. *Biophysical Journal*, 97(9):L19–L21, 2009. (Cited on page 97.)
- [76] Andres F Oberhauser, Piotr E Marszalek, Mariano Carrion-Vazquez, and Julio M Fernández. Single protein misfolding events captured by atomic force microscopy. *Nature Structural Biology*, 6(11):1025–1028, November 1999. (Cited on page 3.)
- [77] José Nelson Onuchic, Z Luthey-Schulten, and Peter G Wolynes. Theory of protein folding: the energy landscape perspective. *Annual review of physical chemistry*, 48:545–600, 1997. (Cited on pages 3 and 69.)
- [78] Matteo Palassini and Felix Ritort. Improving Free-Energy Estimates from Unidirectional Work Measurements: Theory and Experiment. *Physical Review Letters*, 107(6), August 2011. (Cited on pages 92 and 99.)
- [79] H Y Park, S A Kim, J Korlach, E Rhoades, L W Kwok, W R Zipfel, M N Waxham, W W Webb, and L Pollack. Conformational changes of calmodulin upon Ca²⁺ binding studied with a microfluidic mixer. *Proceedings of the National Academy of Sciences of the United States of America*, 105(2):542–547, January 2008. (Cited on pages 7 and 84.)
- [80] Erwin J G Peterman, Frederick Gittes, and Christoph F Schmidt. Laser-induced heating in optical traps. *Biophysical Journal*, 84(2 Pt 1):1308–1316, February 2003. (Cited on page 14.)
- [81] Menahem Pirchi, Guy Ziv, Inbal Riven, Sharona Sedghani Cohen, Nir Zohar, Yoav Barak, and Gilad Haran. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nature Communications*, 2:493–7, October 2011. (Cited on pages 3 and 17.)
- [82] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 3 edition, September 2007. (Cited on pages 20 and 95.)
- [83] Lawrence R Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989. (Cited on pages 17 and 20.)

- [84] Carl-Roland Rabl, Stephen R Martin, Eberhard Neumann, and Peter M Bayley. Temperature jump kinetic study of the stability of apo-calmodulin. *Biophysical Chemistry*, 101-102:553–564, December 2002. (Cited on page 83.)
- [85] E Racca, F Laio, D Poggi, and L Ridolfi. Test to determine the Markov order of a time series. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 75(1), January 2007. (Cited on page 64.)
- [86] Christoph Ratzke, Moritz Mickler, B Hellenkamp, Johannes Buchner, and Thorsten Hugel. Dynamics of heat shock protein 90 C-terminal dimerization is an important part of its conformational cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 107(37):16101–16106, September 2010. (Cited on page 3.)
- [87] Christoph Ratzke, Felix Berkemeier, and Thorsten Hugel. Heat shock protein 90's mechanochemical cycle is dominated by thermal fluctuations. *Proceedings of the National Academy of Sciences of the United States of America*, December 2011. (Cited on pages 3 and 17.)
- [88] Matthias Rief, Mathias Gautel, Filipp Oesterhelt, Julio M Fernández, and Hermann E Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109–1112, May 1997. (Cited on page 3.)
- [89] Michael Schlierf, Hongbin Li, and Julio M Fernández. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19):7299–7304, May 2004. (Cited on pages 3, 14, and 65.)
- [90] Michael Schlierf, Felix Berkemeier, and Matthias Rief. Direct observation of active protein folding using lock-in force spectroscopy. *Biophysical Journal*, 93(11):3989–3998, 2007. (Cited on pages 33 and 52.)
- [91] Elizabeth A Shank, Ciro Cecconi, Jesse W Dill, Susan Marqusee, and Carlos Bustamante. The folding cooperativity of a protein is controlled by its chain topology. *Nature*, 465(7298):637–640, June 2010. (Cited on pages 4 and 9.)
- [92] David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, October 2010. (Cited on pages 3 and 69.)
- [93] Andrea Soranno, Brigitte Buchli, Daniel Nettels, Ryan R Cheng, Sonja Müller-Späth, Shawn H Pfeil, Armin Hoffmann, Everett A

- Lipman, Dmitrii E Makarov, and Benjamin Schuler. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, April 2012. (Cited on page 3.)
- [94] Johannes Stigler, Fabian Ziegler, Anja Gieseke, J Christof M Gebhardt, and Matthias Rief. The complex folding network of single calmodulin molecules. *Science*, 334(6055):512–516, October 2011. (Cited on pages 72 and 78.)
- [95] L Anders Svensson, Eva Thulin, and S Forsén. Proline cis-trans isomers in calbindin D9k observed by X-ray crystallography. *Journal of molecular biology*, 223(3):601–606, February 1992. (Cited on page 89.)
- [96] K Svoboda and Steven M Block. Biological applications of optical forces. *Annu Rev Biophys Biomol Struct*, 23:247–285, 1994. (Cited on pages 9, 10, and 12.)
- [97] SF Tolić-Nørrelykke, Erik Schäffer, J Howard, FS Pavone, F Jülicher, and Henrik Flyvbjerg. Calibration of optical tweezers with positional detection in the back focal plane. *Review of Scientific Instruments*, 77:103101, 2006. (Cited on page 108.)
- [98] Swarnendu Tripathi and John J Portman. Inherent flexibility determines the transition mechanisms of the EF-hands of calmodulin. *Proceedings of the National Academy of Sciences of the United States of America*, 106(7):2104–2109, February 2009. (Cited on page 7.)
- [99] Megan T Valentine, Nicholas R Guydosh, Braulio Gutiérrez-Medina, Adrian N Fehr, Johan O Andreasson, and Steven M Block. Precision steering of an optical trap by electro-optic deflection. *Opt Lett*, 33(6):599–601, March 2008. (Cited on page 109.)
- [100] Claudia Veigel, Fei Wang, Marc L Bartoo, James R Sellers, and Justin E Molloy. The gated gait of the processive molecular motor, myosin V. *Nature cell biology*, 4(1):59–65, 2002. (Cited on page 55.)
- [101] A J Viterbi. Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *Ieee Transactions on Information Theory*, 13(2):260–+, 1967. (Cited on page 21.)
- [102] G M Wang, E M Sevick, Emil Mittag, Debra J Searles, and Denis J Evans. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Physical Review Letters*, 89(5):050601, July 2002. (Cited on page 92.)
- [103] M D Wang, H Yin, R Landick, J Gelles, and Steven M Block. Stretching DNA with optical tweezers. *Biophysical Journal*, 72(3):1335–1346, March 1997. (Cited on page 31.)

- [104] M D Wang, M J Schnitzer, H Yin, R Landick, J Gelles, and Steven M Block. Force and velocity measured for single molecules of RNA polymerase. *Science*, 282(5390):902–907, October 1998. (Cited on page 9.)
- [105] Jin-Der Wen, Maria Manosas, Pan T X Li, Steven B Smith, Carlos Bustamante, Felix Ritort, and Ignacio Tinoco. Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophysical Journal*, 92(9):2996–3009, May 2007. (Cited on page 4.)
- [106] Michael T Woodside, PC Anthony, William M Behnke-Parks, K Larizadeh, Daniel Herschlag, and Steven M Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314(5801):1001, 2006. (Cited on pages 9, 69, and 92.)
- [107] Michael T Woodside, William M Behnke-Parks, Kevan Larizadeh, Kevin Travers, Daniel Herschlag, and Steven M Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciences of the United States of America*, 103(16):6190–6195, April 2006. (Cited on page 4.)
- [108] Jeffries Wyman and Stanley J Gill. *Binding and Linkage: Functional Chemistry of Biological Macromolecules*. University Science Books, August 1990. (Cited on page 37.)
- [109] T Yokouchi, Y Izumi, T Matsufuji, and Y Jinbo. Unfolding intermediate of a multidomain protein, calmodulin, in urea as revealed by small-angle X-ray scattering. *FEBS letters*, 2003. (Cited on page 3.)
- [110] Hao Yu, Xia Liu, Krishna Neupane, Amar Nath Gupta, Angela M Brigley, Allison Solanki, Iveta Sosova, and Michael T Woodside. Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proceedings of the National Academy of Sciences of the United States of America*, March 2012. (Cited on page 58.)