Single molecule investigations of chaperone assisted protein folding

Alireza Mashaghi Tabari

Single molecule investigations of chaperone assisted protein folding

PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op woensdag 12 december 2012 om 10.00 uur

door

ALIREZA MASHAGHI TABARI

Master of Science, Tehran University

geboren te Iran

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. S. J. Tans

Samenstelling promotiecommissie:

Rector Magnificus,	chairman
Prof. dr. S. J. Tans,	Delft University of Technology, promotor
Prof. dr. C. Dekker,	Delft University of Technology
Prof. dr. B. Mulder,	University of Wageningen
Prof. dr. P. Bolhuis,	University of Amsterdam
Prof. dr. B. Maier,	University of Cologne
Dr. M-E. Aubin,	Delft University of Technology
Prof. dr. S. de Vries,	Delft University of Technology

The work described in this thesis was performed at the FOM institute for Atomic and Molecular Physics (AMOLF) in Amsterdam, The Netherlands. This work is part of the research program of the 'Stichting voor Fundamenteel onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'.

Nederlandse titel: Enkele molecule studies van chaperone gestuurde eiwit vouwing

ISBN 9789077209691

A digital version of this thesis can be obtained from http:// www.amolf.nl and from http://repository.tudelft.nl. Printed copies can be obtained by request via email to library@amolf.nl.

Printed by Ipskamp Drukkers, the Netherlands

Contents

CHAPTER 1

Single molecule investigations of chaperones and chaperone assisted protein folding, a review

Most newly synthesized proteins require the action of molecular chaperones to avoid aberrant folding and aggregation in the crowded cellular environment. Our current understanding of chaperones and their function is largely based on insights obtained from bulk biochemical techniques. We still have limited insight into the conformational dynamics and transient intermediate states that are thought to be central to the folding process. Recently developed single-molecule techniques are now starting to address these issues. Here we review key studies in this upcoming field, explore the possibilities and limitations of the single-molecule approach, and discuss the many open questions for this new and upcoming research direction.

1.1 Protein folding and chaperones

Newly synthesized polypeptides interact with a multitude of cellular components before they gain their native fold. Folding occurs in a highly crowded environment that strongly facilitates undesired interactions that can irreversibly harm the folding process. There are ample opportunities for such undesired interactions: while secondary structures form on the order of microseconds or faster, tertiary folding may take up to minutes [1, 2] with proteins diffusing through the cellular environment on sub second timescales [3]. Unfolded protein chains can self assemble into aggregates with rates that are sensitive to concentrations, which are high at the translation sites [4, 5]. Once properly folded, proteins are at a reduced risk of pathological interactions. Proteins generally have significant number of hydrophobic residues and assume folded conformations that bring the hydrophobic residues close to one another within the protein core [6]. A misfolding error [7] typically leads to exposure of hydrophobic residues that increases the probability of unwanted inter-molecular interactions in a crowded surrounding.

To avoid non-productive interactions of proteins, all cells are equipped with a wide spectrum of molecular chaperones that are believed to help mitigating the risk of pathological interactions during folding. Many newly translated proteins interact with chaperones beginning at the ribosome exit tunnel and/or after release from the ribosome. Simultaneous deletion of these chaperones often leads to protein aggregation with lethal consequences [8]. Chaperones possibly influence the conformational dynamics of proteins at critical points along the folding pathways. Considerable knowledge on the biology of chaperones has been amassed over the past decades, yet many fundamental questions remain unanswered. Key questions concern disentangling the roles of chaperones as suppressors of aggregation and folding catalysts and deciphering the underlying physical principles that govern the functioning of chaperones: How do they affect the conformation of a folding protein chain? What is the role of energy input in the chaperone-assisted folding process? What intermediate states are visited along the folding pathway? How can chaperones act in a generic manner on many different proteins?

Such processes are difficult to study with bulk techniques because the underlying molecular mechanisms involve interactions that are often transient and heterogeneous. Single-molecule approaches have allowed the study of isolated proteins in real time, thereby revealing the rich conformational dynamics of proteins during the folding process. These techniques have in recent years begun to being applied to study chaperone assisted protein folding. While many excellent reviews on

unassisted protein folding and on bulk studies of chaperones have been written, the emerging single-molecule studies of chaperones and assisted protein folding have so far received little attention. Here we review a number of studies that address the folding of proteins aided by chaperones at the single molecule level. This will be preceded by a discussion of the nature of single molecule assays as compared to bulk measurement techniques. We highlight how single-molecule methods can be used to answer the key questions related to assisted-protein folding reactions.

1.2 Single molecule versus bulk studies

Conventional biochemical assays involve millions of billions of molecules often reacting within an aqueous environment. The measured molecular properties are then summed or averaged over an ensemble of molecules as described in statistical thermodynamics (Fig 1.1a). This leads in some cases to an improved signal to noise ratio in bulk studies, and in other cases to information loss (Fig 1.1b). To illustrate this point, consider the following example. Proteins in solution carry kinetic energy and often have significant electric dipole moments that govern their interactions. Measurement of the kinetic energy –i.e. the temperature of the protein solution - can be done precisely in bulk, while fluctuations appear and signal-to-noise ratio reduces when the measurements are performed on a small subsystem (Fig 1.1b). Fluctuations may however carry hidden information and might be themselves of interest to the observer. When the net dipole moment of the system is measured in bulk, the dipoles will sum to zero for a solution of proteins with negligible inter-molecular coupling. In this case, measurement at smaller scale will help the observer to distinguish between the protein solution and a solution containing non-polar molecules. In exceptional cases some bulk measurements allow one to partly circumvent the loss of information due to ensemble averaging and lack of synchrony. For example, when molecules are embedded in anisotropic environments, their residual inter-nuclear magnetic dipolar coupling (which is averaged to zero in solution NMR studies) will be detectable in bulk due to partial alignment of molecules. The use of this approach to recover structural information from dipolar coupling has provided new insights into the folding of the spliceosome, an RNA-protein complex responsible for mRNA splicing in cells [9].

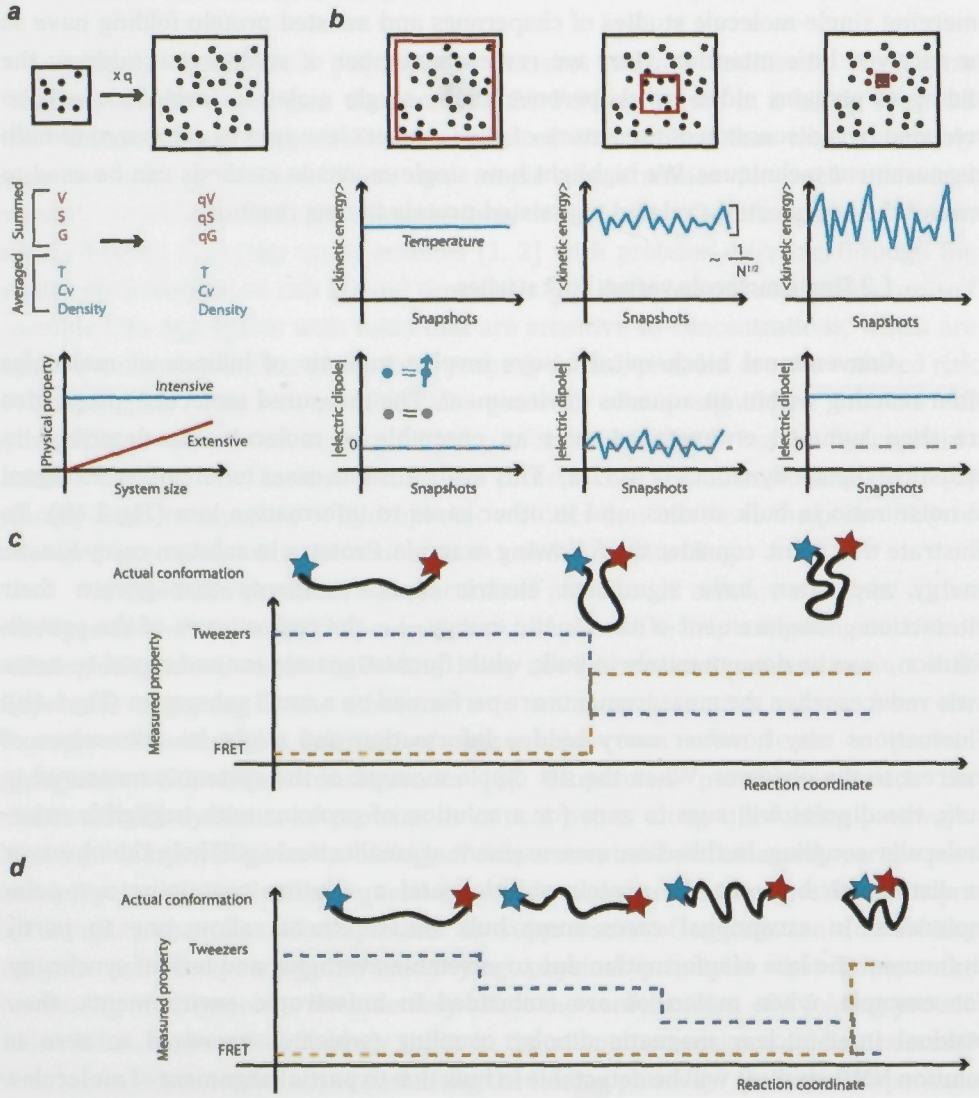


Figure 1.1: Limitations of single molecule techniques. (a) In thermodynamics, extensive and intensive properties of a large system are the result of sum and averaging performed on individual molecular properties. For a system composed of gas molecules, extensive properties such as free energy (G), entropy (S) and volume (V) scale with the size of the system. Intensive properties such as temperature (T) do not scale with the system size and remain invariant upon scaling. (b) illustrates the key differences when an intensive property is measured at bulk and single-molecule scales. Average kinetic energy measurement at large scale provides a signal with negligible noise. The noise varies inversely with the square root of the number of molecule under study. The noise however might carry useful information. Electric dipole

moment (per molecule) measurements in bulk for example cannot distinguish between a solution of polar molecules and a solution of non-polar molecules, as the dipoles will cancel out due to random orientation of molecules. When measurements are performed on a mesoscopic sample, the electric dipole of the polar molecule appears as the measurement noise. Single-molecule measurement provides the magnitude of the dipole moment precisely. (c,d) An example conformation of a protein along a folding trajectory. In single-molecule experiments, the state of a molecule with many degrees of freedom is identified by one or a few detected properties, for example FRET signal or end-to-end distance measured by optical tweezers. (c) Trajectory with a single folding transition. In this case a long-range interaction is formed early in the folding process. The end-to-end distance and FRET signal are both degenerate properties after the transition. (d) Trajectory with multi-folding transitions. For a protein that folds by nucleation and growth from one initial nucleation site, the End-to-end distance is a one-to-one function while the FRET signal is degenerate over many states.

Measurements on a single molecule when performed over a long time period can in principle provide the same information as bulk measurements. This property is common to molecular systems (with some exceptions like systems with symmetry breaking [10]) and is referred to as the ergodic property. The advantage of single molecule assays is in cases where ensemble averaging hides reaction intermediates [11]. In particular, single molecule studies are assumed to be beneficial in measuring non-equilibrium transient processes, while bulk studies are often done on quasi-static processes and equilibrium systems (with a few exceptions such as time-resolved X-ray crystallography and two-dimensional infrared spectroscopy that can be used to follow transient molecular events at fast time scales [12, 13]). From that perspective, bulk studies provide a simple picture of large systems described by a large number of degrees of freedom at the cost of information on transient intermediates. Interestingly, the information gathered from the study of out-of-equilibrium processes at the single molecule level can be used to estimate equilibrium properties in certain cases [14, 15].

Transient intermediate molecular states are typically hidden from the observer in bulk measurements on protein folding because individual molecules are in different folded states i.e. they are not in synchrony. In chaperone-assisted protein folding processes, the system under study is even more heterogeneous. Individual copies of a protein may not only be in different folded states but they also may be at different stages of interaction with the chaperone during the chaperone cycle. Different single molecule techniques vary in their ability to detect intermediates and may provide complementary information (Fig 1.1c, 1.1d). To detect intermediate conformations, one might need to design an individualized single molecule experiment. In Figs 1.1c and 1.1d, two folding processes are displayed schematically: one in which folding commences by local interactions and another in which a long range interaction is first established and followed by local ones. Detectability of the folding intermediates depends on the single molecule approaches used, the design of

the experiment and the protein under study. For example, when folding occurs via nucleation and growth by forming local interactions, the end-to-end distance of the molecule reduces along the folding pathway and can be monitored by optical tweezers (Fig 1.1d). For the protein shown in Fig 1.1c that folds by a long-range interaction followed by short range ones, folding steps that involve short-range interactions are not resolved when the end-to-end distance is measured by optical tweezers. FRET based assays with dye molecules bound to appropriate locations within the protein structure can in principle help in resolving the hidden conformations.

Single-molecule kinetic studies may also provide information about the hidden intermediates [16]. In Fig 1.2a, three scenarios are compared: in the first scenario (type I) a single protein with two-state folding pathway is studied: the observer follows the state of the protein as a function of time. This allows one to measure the transition time (dwell time) between the two states and compile the transition probability (dwell time) histogram, which is indicated in Fig 1.2a. A chaperone could reduce the transition energy barrier between the two states and simply increase the folding rate. Alternatively, a chaperone may change the folding pathway of the protein, e.g. by stabilizing a transition intermediate. Two experimental single-molecule schemes are presented in Fig 1.2a to study such changes. In one assay (type II), the intermediate is not detected directly while in another one (type III) the intermediate and the individual reaction rates are independently measured. The shape of the compiled transition time histogram in type II experiment does indicate the existence of the intermediates and even in special cases the reaction rates and the number of intermediates can be inferred [11, 16]. However, when multiple on-pathway intermediates with varying transition energy barriers are visited, and when off-pathway intermediates are populated, direct detection of the intermediates (as in type III) are required to measure the kinetics and the nature of the intermediates.

Single-molecule assays further allow one to simplify complex reactions by controlling the number of possible intermolecular interactions (Fig 1.2b). In complex molecular systems with multiple components and several competing reaction pathways, it is often impossible to disentangle the roles of individual parts as intermolecular interactions can lead to emergent properties. For instance in bulk

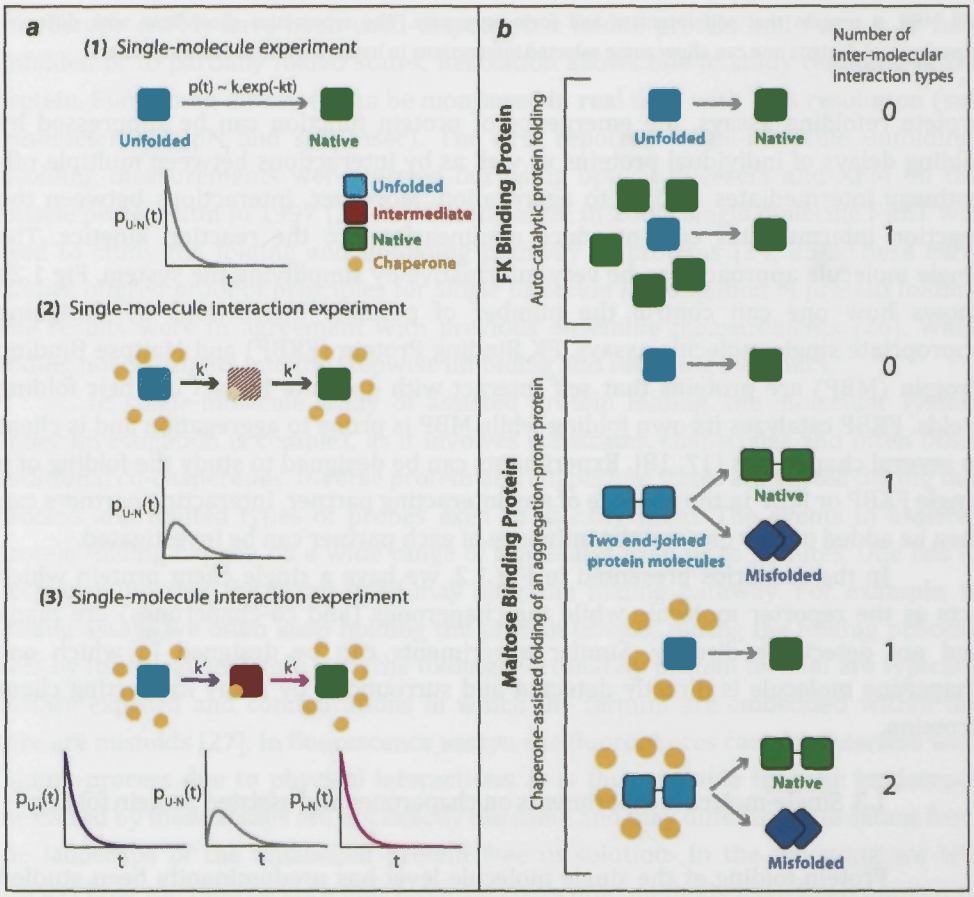


Figure 1.2: Benefits of single-molecule approaches. (a) Single-molecule techniques allow for indirect and direct detection of chaperone induced folding intermediates. Three scenarios are presented. 1. Single step (two-state) protein folding: the state of the protein is monitored in time as the protein transits from unfoled state to folded state. The constructed distribution of measured transition (dwell) times follows a single exponential decay. When chaperones are present, they may change the folding rate without changing the number of conformational states visited during folding. In this case, the constructed distribution of dwell times follows a single exponential decay with a different rate constant. Alternatively the folding pathway of a two-state folder might change in the presence of chaperones by emergence of intermediate states as indicated in schemes 2 and 3. The emergence of intermediates renders the probability distribution of the transition time non exponential. In scenario 2 the intermediate molecule is not directly observed (striped red square). However from the shape of the probability distribution, it might be possible to decipher the number of intermediate steps. In scenario 3 all molecular species are directly detectable and reaction rates can be measured directly. (b) Single-molecule techniques allow for controlling the number of possible intermolecular interactions and taking a reductionist approach. Inter-molecular interactions may lead to nonlinearity of the kinetics and emergence of competing reaction pathways. Two examples are illustrated: FKBP prolyl isomerase, a protein that self interacts and catalyzes its own folding

and MBP, a protein that self interacts and form aggregate. The schematics show how with different experimental designs one can allow some selected interactions to happen while preventing others.

protein refolding assays, the emergence of protein function can be suppressed by folding delays of individual proteins as well as by interactions between multiple off-pathway intermediates leading to aggregation. Moreover, interactions between the reaction intermediates can introduce nonlinearity into the reaction kinetics. The single molecule approach can be very informative by simplifying the system. Fig 1.2b shows how one can control the number of possible interactions by designing appropriate single molecule assays. FK Binding Protein (FKBP) and Maltose Binding Protein (MBP) are proteins that self interact with opposite impact on their folding yields. FKBP catalyzes its own folding while MBP is prone to aggregation and is client to several chaperones [17, 18]. Experiments can be designed to study the folding of a single FKBP or MBP in the absence of any interacting partner. Interacting partners can then be added one by one and the influence of each partner can be investigated.

In the scenarios presented in Fig 1.2, we have a single client protein which acts as the reporter molecule while the chaperones (and co-chaperones) are many and not detectable directly. Similar experiments can be designed in which one chaperone molecule is directly detected and surrounded by many interacting client proteins.

1.3. Single-molecule experiments on chaperones and assisted protein folding

Protein folding at the single molecule level has predominantly been studied using force spectroscopy and fluorescence methods. Fluorescence methods can be readily applied to proteins that are naturally fluorescent. The introduction of artificial fluorescent labels into the structure of a protein or chaperone, allows one to employ Förster Resonance Energy Transfer (FRET). FRET is a fluorescence method based on radiationless energy transfer between a donor and an acceptor with a transfer efficiency that depends on the local environment, the spatial distance ($\sim r^{-6}$) between the donor and the acceptor and their relative orientation [19]. FRET-based assays thus probe the proximity of two regions on a single molecule or a pair of interacting molecules. When an acceptor approaches the donor, the emission and lifetime of the donor, as well as the polarization of the emission (with respect to excitation), changes [20]. In the next section, we explain how these properties can be used to study assisted protein folding.

In force spectroscopy, the system under study can be perturbed mechanically, for example by pulling the end of linear molecules. Optical tweezers and atomic force

microscope (AFM) have been used to perturb a folded protein and force it to fully unfolded or to partially folded states. Relaxation allows one to study refolding of the protein. Forces and distances can be monitored in real time with high resolution (sub nanometer, sub pN and sub msec). The first reported single-molecule unfolding-refolding measurements were carried out using optical tweezers and AFM on the muscle protein titin in 1997 [21-23]. Shortly after, in 2000, single molecule FRET was used to study the folding and unfolding pathway of proteins [24, 25]. These early studies offered proof of principles for single molecule investigation of protein folding. The results were in agreement with previous ensemble measurements [26], while adding novel insights into the stepwise unfolding and refolding dynamics.

In single-molecule study of assisted protein folding, the molecular system under investigation is complex, as it involves substrates, chaperones and often other additional co-chaperones. Diverse protein and chaperone states are visited during this process and limited types of probes exist to identify them. The events in assisted protein folding happen on a wide range of timescales from μs to minutes. One has to keep in mind that the assay itself may alter the folding pathway. For example, in pulling assays we often keep holding the protein termini during the folding process. This by itself may interfere with the folding. Fortunately protein termini are typically surface exposed and configurations in which the termini are embedded within the core are misfolds [27]. In fluorescence assays, the fluorophores can also interfere with folding process due to physical interactions. It is thus possible that the landscapes measured by these assays are not exactly the same and may differ to some extent from the landscape of the unlabelled protein free in solution. In the following we will discuss how the fluorescence and force spectroscopic methods have been applied to investigate chaperone-assisted folding processes.

1.3.1 Fluorescence experiments

Single-molecule FRET has been used to follow chaperone assisted folding and unfolding of proteins in real time with high temporal (millisecond) resolution at the single molecule level. Sharma et al. [28] engineered three donor-acceptor pairs within a slow folding mutant of Maltose Binding Protein (DM-MBP) while ensuring the modifications did not affect the folding rate of the protein. Stopped-flow ensemble FRET measurements showed that in the absence of the barrel-shaped chaperonin GroEL, the protein collapses to a compact state within milliseconds. In the presence of GroEL, after the initial collapse the protein expands rapidly with $t_{1/2}$ of 100 ms. Single molecule FRET (smFRET) was used to study structural heterogeneity in the ensemble

of client proteins upon GroEL binding (Fig 1.3). In the absence of chaperone the protein adopts a uniformly compact conformation within the first 200 s of spontaneous refolding.

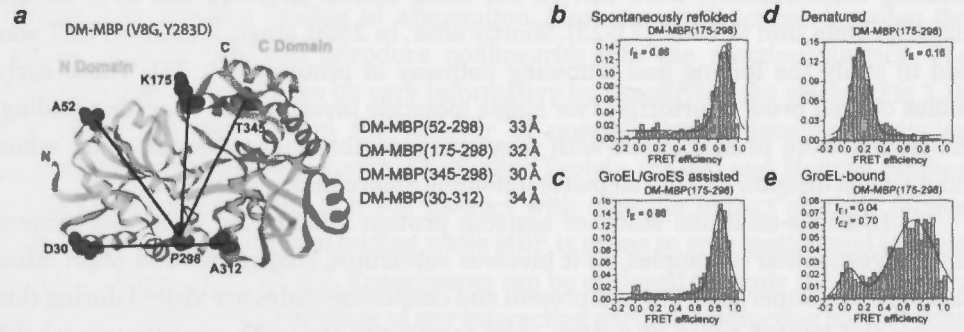


Figure 1.3: Single-Molecule FRET Analysis of DM-MBP in Spontaneous and Chaperonin-Assisted Folding (a) Ribbon diagram of the structure of MBP (pdb 1OMP) with the N-terminal domain shown in yellow and the C-terminal domain in blue. The positions of engineered cysteines are indicated in red (N domain) and blue (C domain). (b–e) Single-molecule FRET measurements of double-labeled DM-MBP (175-298). GdnHCl-denatured double-labeled DM-MBP (3 nM) was diluted 50-fold (60 pM final concentration) either into buffer alone (b) or into buffer containing 3 μM GroEL/6 μM GroES/2 mM ATP (c) or 3 M GdnHCl (d) or 3 μM GroEL alone (e). Peak values of a Gaussian fit to the FRET efficiency distributions (f_E) are indicated. Representative histograms of at least two independent measurements are shown. (Figure is adapted from [28] with permission).

GroEL binding produces fast unfolding and the emergence of a heterogeneous conformational distribution of the client protein with molecules populating both compact and locally expanded states. Finally steady-state anisotropy measurements have been used to study modulation of local protein mobility by chaperones and to show that segmental mobility of the MBP mutant is affected upon GroEL binding.

Chakraborty et al. [29] presented smFRET data that supports GroEl–GroES can rescue a protein from a kinetically trapped state. This is consistent with the idea that GroEL unfolds a misfolded protein, and gives it new opportunity to refold properly.

Hofmann et al. [30] reported sm FRET results that show differential effects of this chaperonin system on folding rates of different domains of a substrate protein. The results showed that confinement in the chaperonin decelerates the folding of the C-terminal domain of the substrate protein rhodanese, but leaves the folding rate of the N-terminal domain unaffected. The observed variable effects of chaperonin [30, 31] pose a challenge on how to encompass the effects of chaperonin into a universal mechanism.

An example of non-equilibrium measurements, is probing the binding of the proteins Albumin and Casein to stretched hydrophobic polypeptides [32]. Using a stretch assay along with FRET, a correlation was found between bindings of these proteins with the force-induced unfolding of Fibronectin.

SmFRET studies have also revealed the structural heterogeneity of the chaperones, as encoded in the width of the distribution of FRET efficiencies. Mapa et al. [33] used smFRET to study the structural heterogeneity of mitochondrial Ssc1, a member of the Hsp70/DnaK family. Surprisingly while the ATP state shows structural uniformity, the ADP state was found to be heterogeneous in conformation. Similar results were obtained in smFRET mesurements for the endoplasmic eticulum Hsp70 homolog [34] and in bulk electron paramagnetic resonance measurements for *E. coli* Hsp70 [35]. In another study Ratzke et al. [36] used smFRET with Hsp90 dimers caged in lipid vesicles, and studied the dimerization of the chaperone at its C- and N-terminal interfaces. They surprisingly found that, the C-terminal dimer opens and closes with fast kinetics and that the C- and N-terminal dimerizations are anti-correlated. The dimerization of C-terminal open state is shown to be modulated by nucleotide binding.

To understand how chaperones modulate protein folding, it is important to be able to count the number of chaperones bound to a single client protein. Fluorescence correlation spectroscopy has been used to assess the stoichiometry of chaperone-protein complexes. In late 1990s, fluorescence correlation spectroscopy was first applied to single molecule experiments [37, 38]. A decade later, Sharma et al. [28] used pulsed interleaved excitation (PIE) in conjunction with fluorescence cross-correlation spectroscopy (FCCS) to indentify the stoichiometry of GroEL-MBP complexes and screen for complexes with 1:1 ratio for conformational analysis with FRET. FCS can also be combined with FRET to increase the temporal resolution to sub microseconds. Neuweiler et al. [39] used single-molecule fluorescence quenching by photo-induced electron transfer, detecting short-range events (~2 nm), in conjunction with fluorescence correlation spectroscopy (PET-FCS) to explore folding dynamics of the small binding domain BBL with nanosecond time resolution. The experimental scheme can potentially be extended to study the impact of chaperones on the folding dynamics.

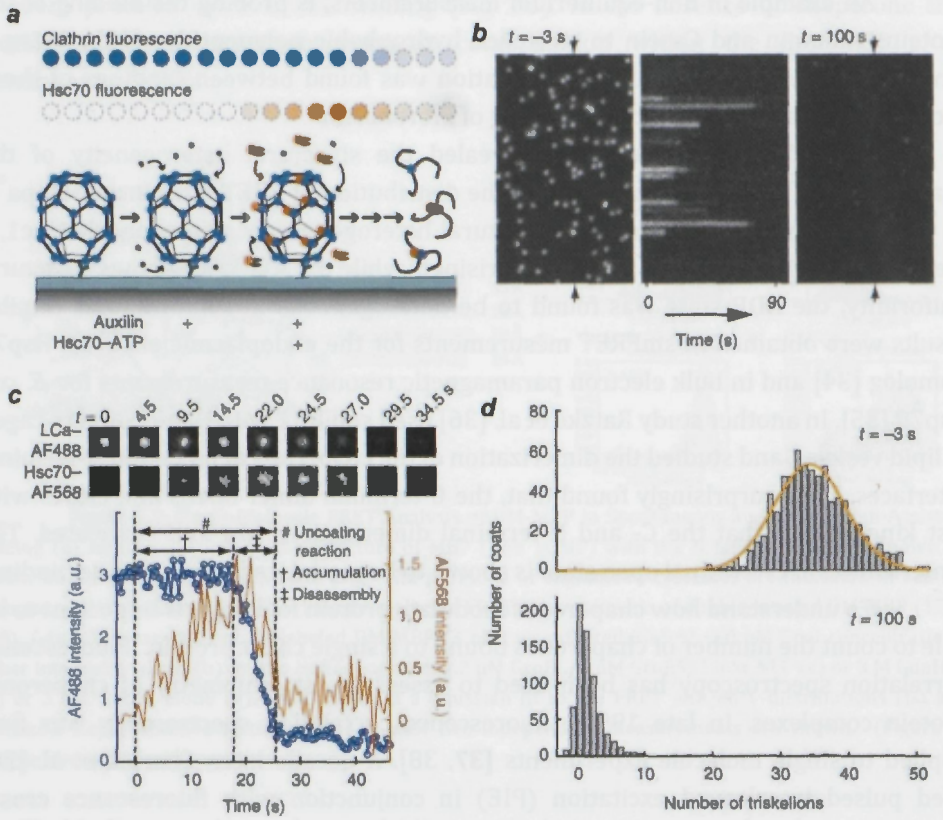


Figure 1.4: (a) Schematic representation of the single-particle uncoating assay. The fluorescence intensities from labeled clathrin and Hsc70 were monitored by TIRF microscopy for clathrin/AP-2 coats captured on the surface of a PEG-modified glass cover slip. (b) Representative time series of a single-particle uncoating assay. Left and right panels, first and last frames, respectively, of the fluorescence channel used to monitor the signal from coats tagged with clathrin LCa–AF488. Middle, kymograph generated from the vertical axis indicated by the arrows in the left panel, showing the unsynchronized disappearance of clathrin fluorescence. Hsc70–ATP (1.2 μM) arrived in the flow chamber at t = 0. (c) Uncoating profile from a single coat. The selected snapshots from the time series (top) show the fluorescence from clathrin and Hsc70 in the selected coat, at various time points during the uncoating reaction carried out with 0.9 μM Hsc70. The snapshots are background-corrected averages of three successive frames. The plot shows intensity traces of the clathrin (blue) and Hsc70 (orange) signals. The t = 0 time point is the moment at which a rapid increase in Hsc70 background signal is recorded; this event corresponds to the arrival of Hsc70 within the evanescent field at the cover slip. (d) Histogram of the number of trimers (triskelions) per coat at the beginning (top) and the end (bottom) of the single-particle uncoating assay carried out with 1.2 μM Hsc70. The number of trimers in intact coats follows a normal distribution with a mean of 34 triskelions per coat (top). In most cases, only one or two trimers remained at the site of a coat at the end of the reaction (bottom). Objects with overlapping point-spread functions were excluded from this analysis. A.u., arbitrary units. (Figure is adapted from [40] with permission.)

Single-particle fluorescence imaging combined with micro fluidics and TIRF illumination has been used to study chaperone-assisted disassembly processes (Fig 1.4). Böcking et al. [40] studied how Hsc70, a member of Hsp70 family, catalyzes a large-scale disassembly reaction in real time. Single clathrin coats were fluorescently labeled and immobilized on the surface of particles attached to a cover slip within a micro fluidic channel. Fluorescently labeled chaperone was infused and uncoating was tracked by TIRF and photobleaching.

1.3.2 Force spectroscopy and pulling experiments

In mechanical manipulation approaches, proteins are held via their termini between a flat surface and a cantilever (in the case of AFM) or between two micron size beads (in the case of an optical tweezers). By applying mechanical force to the molecular ends, the proteins are stretched and forced to unfold. The end-to-end distance of the protein can be monitored during the experiment. The methods can also be used to study protein folding, by relaxing the unfolded protein and then monitoring both force and the effective protein end-to-end distance by consecutive pulls with different waiting times.

Bechtluft et al. [18] used this approach to demonstrate how SecB, the ATP-independent *E. coli* chaperone involved in protein translocation across the plasma membrane, modulates the folding pathway of MBP (Fig 1.5). MBP unfolding was found to occur in two steps: first a C-terminal part (~ 28 nm) was unfolded resulting in an MBP core intermediate. The core intermediate then unfolded in one step. During MBP folding the extended peptide is compacted to a molten globule state followed by folding of MBP core. The effect of SecB was found to be specific: tertiary contacts were effectively blocked in the transition to the core state, while the transition from the core to the native state was unaffected by SecB (Fig 1.5e,f). Similar effects had been also observed for GroEL [41, 42]. By analyzing a tandem 4MBP construct that during refolding is highly prone for misfolding (Fig 1.2b), SecB was found to prevent the stable aggregation interactions between MBP molecules and thus to significantly alter the folding pathway of MBP.

In two other studies, Aubin-Tam et al. and Maillard et al. demonstrated how ClpXP and ClpX unfold individual protein domains in a highly

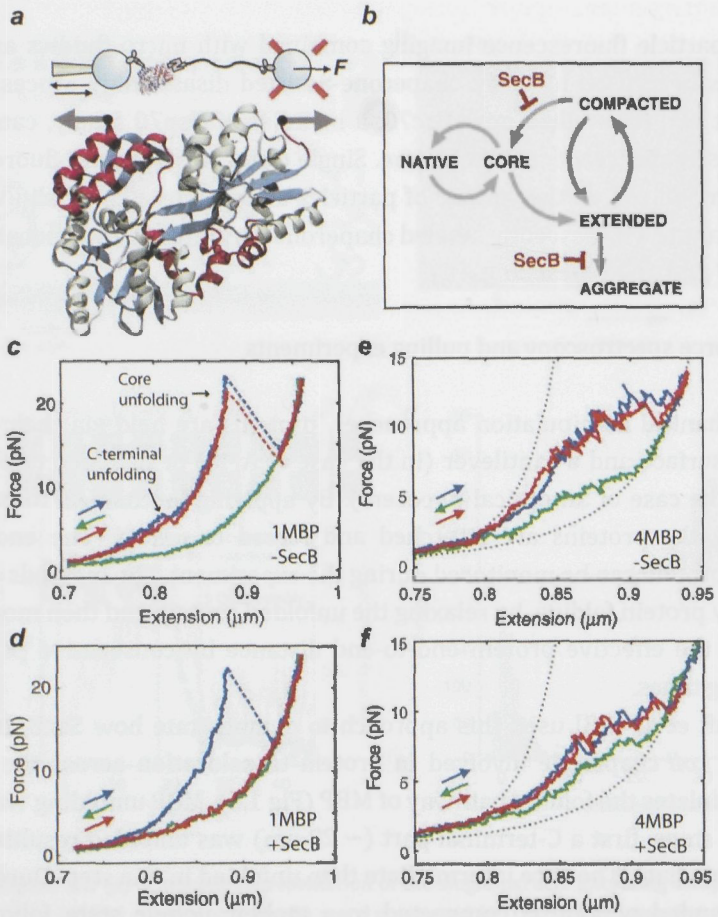


Figure 1.5: Experimental setup and SecB dependence of MBP force-extension curves. (a) MBP is tethered between two beads: One is held on a position-controlled micropipette; the other is held by an optical trap allowing force detection. At the C terminus, MBP is attached via an antibody-myc-tag connection, whereas at the N-terminus it is attached via streptavidin-biotin linkages to a DNA tether, which in turn is attached to the bead surface via an antibody-digoxigenin connection. (b) Force-induced MBP folding and unfolding transitions observed in the experiments and their dependence on SecB. (c) Force-extension curves in the absence of SecB showing unfolding at high force (blue), refolding at low forces (green), and again unfolding at high force (red). (d) Force-extension curves in the presence of SecB (0.1 μM). The second stretching curve (red) lacks the typical unfolding features, showing that stable tertiary interactions are absent. (e) Force-extension curves of the C-terminal regions of 4MBP in absence of SecB. The construct was first stretched, resulting in the predicted gradual unfolding of the external α helices (red helices in panel a), and then relaxed before the core structures would unfold. During relaxation the near-equilibrium refolding of the same α helices was observed directly as a shortening of the tether. The dotted lines indicate the WLC behavior: The first denotes the DNA alone, whereas in the second a 4x91 residue-compliant polypeptide was added, representing the unfolded external α helices. (f) Force-extension curves of the C-terminal

regions of 4MBP in presence of SecB (0.1 μM). The SecB interactions do not alter this refolding transition. (Figure is adapted from [18] with permission.)

cooperative manner prior to their degradation [43, 44]. Their finding supports a power-stroke model of denaturation in which enzyme-mediated unfolding of stable protein domains involves concurrence of mechanical pulling of the template through the enzyme pore and a transient stochastic reduction in protein stability.

AFM has been used to image chaperones to gain structural insight or to assay their binding and localization: Structural analysis of GroEL, GroES and oligomeric GroES have been performed using liquid AFM [45-48]. It has also been used to confirm the presence of Hsp60 on the membrane of stressed cells at a high lateral resolution by detecting specific single molecule binding events [49]. In another study Zhu et al. [50] found that the small heat shock protein αB-crystallin interacts and protects cardiac titin from damage. This is achieved by lowering the persistence length of the titin N2B-Us element and by reducing the unfolding rate of the Ig domain flanking the N2B-Us.

AFM has been used to probe binding of chaperones to client proteins as well as stabilization of the clients. Here, one of the interaction partners is attached on the surface, while the other is connected to the AFM tip. An AFM study showed that interaction forces between substrates (destabilized mutant of citrate synthase and RTEM beta-lactamase) and GroEL decrease in the presence of ATP (but not ATPgammaS) and that the force is smaller for native-like proteins than for the fully denatured ones [48]. AFM has also been used to analyze interactions of alpha-synuclein with microbial esterases. The interaction of alpha-synuclein with esterase appears to be highly specific and can protect the native conformation of esterase [51].

1.3.3 In silico single-molecule methods

Although molecular simulations have yielded important insights into mechanistic aspects of folding of individual proteins in isolation, they have rarely been used to study chaperone-assisted protein folding problem [52], either co- or post translationally. Such simulations have not been tractable due to the large system size and the enormous computational cost involved. Some in silico studies exist on the structure and dynamics of chaperones without client and on chaperone assisted protein folding. Unbiased all atom molecular dynamics simulations of the GroEL subunit protein have been performed with and without ATP. The study revealed inter-ring cooperativity and nucleotide-dependent conformational transitions in GroEL [53-55].

Fan et al. [56] studied the facilitation of folding of a partially folded protein using a coarse grained model of GroEL/ES. The chaperone was simulated as a hydrophobic box that allows repeated binding and release of substrate. Folding then proceeds under spatial confinement. The study led to the conclusion that folding of encapsulated proteins is facilitated under spatial confinement.

The average structure of DnaK-DnaJ complex was obtained by MD simulation [57]. Liwo et al simulated the full transition of an Hsp70 from the SBD-closed to the SBD-open conformation by conducting canonical and multiplexed replica exchange simulations of the conformational dynamics of Hsp70s using a coarse-grained molecular dynamics approach with the UNRES force field [58]. The results confirmed the experimentally observed influence of ATP-binding on the transition of Hsp70s from the SBD-closed to the SBD-open form .

The nucleotide dependence of the dynamics and structural heterogeneity of Hsp90 have been studied recently using novel simulation methods. In one study, the authors performed multi scale comparison of the dynamics of Hsp90 from different sources. They considered the fluctuations of the distances of all pairs of amino acids (based on several hundred ns atomistic MD simulations) and thereby found the regions subject to the largest deformation of their structural neighborhood upon change of the ligand [59]. To gain insight to motions at the scale of domains, a minimalistic quasi-rigid domain description was used. The results pointed towards two functional sites important in nucleotide mediated structural changes, one being at the interface of middle-domain and N-terminal domain and a second one within the middle domain.

In another study, using molecular dynamics simulations combined with principal component analysis and the energy landscape model, the authors identified a network of conserved regions with possible functional role in coordinating functional dynamics, principal collective motions and allosteric signaling of Hsp90 and its homologs [60]. With new developments in coarse graining protocols and efficient atomistic modeling approaches, we expect a surge in single-molecule in silico studies of chaperones and chaperone-assisted protein folding.

1.4 Challenges, proposals and outlook

Understanding how protein chains fold into their functional state has been referred to as one of the grand challenges of modern science. Understanding the decisive events of this elementary process would open the door to predicting complex protein structures and functions from DNA sequence data, and to engineering non-

natural proteins with novel functions. Moreover, mechanistic insights into the folding errors and protein aggregation that underlie many medical and ageing conditions are essential for eventually developing rational approaches to therapeutic intervention.

Chaperone-assisted protein folding poses considerable challenges. Biochemical approaches have been highly successful in unraveling the life-cycle of chaperones, their structures, binding partners, and their effects on the yield of natively folded protein. However, they are not well-suited to study the protein conformational dynamics and multiple transitions between folded states that determine folding pathways and outcomes. Simulations of protein-chaperone dynamics remain scarce as they are limited by the corresponding large reaction volumes and timescales. As a result, basic questions on the effect of chaperones on folding dynamics remain unanswered.

Single molecule studies have already contributed to a deeper understanding of assisted protein folding. In particular, they revealed the structure and dynamics of the intermediate states in the folding pathways of a few model proteins and the direct interaction of the folding intermediates and chaperones. A future important advance would be to investigate protein folding mechanisms in more complex environments in the presence of a network of chaperones and co-chaperones. The chaperones can be added in a controlled manner or by performing folding experiment within the cytosolic extract. Such studies would help to gain insights into the consequences of a failure of the proteastasis system caused by imbalance or even absence of sufficient chaperone capacity, e.g. the development of protein misfolding diseases [61].

In addition to probing processes with direct relevance to biology, single-molecule methods will also increasingly target the underlying physical principles that chaperones exploit to assist folding. Single-molecule methods can be used to measure chaperone-induced changes in a folding landscape. Folding landscape could be altered in a number of ways. By binding and unbinding substrate proteins, chaperones can protect protein chain regions from unwanted interactions within and between proteins, and hence improves the folding yield. In the folding landscape, this would correspond to increasing the barrier to the aggregated state. It has often been speculated that chaperones have more elaborate mechanisms with a more direct role in protein folding. Chaperones, which present a folding cavity like GroEL may aid protein folding by confinement thereby lowering the entropy and consequently raising the energy of the unfolded state and lowering the barrier to the native state. One may also speculate that chaperone interactions lower the barrier to the native state by stabilizing the folding transition state, in a manner that resembles a classical catalyst. Alternatively chaperones may produce additional intermediates in between

the unfolded and native folds, which would be detectable for example as a state with an intermediate length in the optical tweezers experiment.

With the development of assays with single-molecule detection capabilities, there is a growing interest in their application in protein folding studies. Available techniques can be adapted to address chaperone-assisted protein folding. For example, optical tweezers and Atomic force microscopes with imaging capabilities and thermal control would be valuable tools to study thermo sensitive chaperones and heat and cold stress response. An important step would be to use the intrinsic fluorescence of tryptophan and tyrosine to gain complementary information on the conformation of proteins and their interaction with chaperones. Recently tryptophan has been successfully used as FRET donor in bulk RNA-Protein binding studies [62]. Unfortunately photoliability and low quantum yield of these aminoacids have hindered their usage particularly in single molecule FRET studies [19]. An important advance came from photoinduced electron transfer (PET) studies of folding of a single protein encapsulated in a polymeric vesicle [63]. In this approach, electron is transferred from a dye molecule to a tryptophan with an efficiency that depends on the distance.

An ultimate single molecule assay on chaperone-assisted protein folding will be to follow protein folding inside cells. This would facilitate to study chaperone assisted protein folding in the complex surrounding of the highly crowded cytoplasm or even within organelles that provide specialized microenvironments [64]. To study de novo folding of proteins and (un)folding of proteins under cellular stresses within cells, sensitive fluorescence and spectroscopic methods would be needed. Furthermore, in such experiments one deals with mesoscopic number of molecular species and therefore fluctuations cannot be detected. In recent years, availability of sensitive molecular detection techniques, such as single-molecule fluorescence counting and Bayesian localization microscopy, allows to study low copy number molecules within single cells [65-67]. Recent developments in NMR spectroscopy have facilitated the measurement of 3D protein structure using high-resolution heteronuclear multi-dimensional NMR spectra of macromolecules in living cells [68, 69]. The development of such new technologies for detection of single and mesoscopic number of protein molecules, and the input of fresh ideas, have opened up new perspectives on our understanding of the mechanisms of chaperone-assisted protein folding, some of which could be applicable to protein misfolding diseases.

CHAPTER 2

Noise reduction by signal combination in Fourier space applied to drift correction in optical tweezers

A general method is proposed to reduce noise by combining signals. Different measurements of the same physical quantity often exhibit different noise levels in different frequency ranges. Hence, a single high-fidelity signal can be constructed by combining the low-noise parts of the signals in Fourier space. We demonstrate this method by reducing noise in the measured bead-to-bead distance in an optical tweezers setup.

2.1 Optical tweezers setup

The optical tweezers (Fig 2.1.) was set up based on [18] with some improvements. Briefly the setup contained a single trapping laser (ND:YVO4 laser (Spectra Physics, $\lambda = 1064$ nm, 5,4 W) pumped by a laser diode through an optical fiber. The objective lens (60x water-immersion, Nikon CFI Plan-Apochromat, NA = 1,2) is used both for focusing the trapping laser and for imaging. The flow system was driven by a piezostage (E-710.P4L, Physik Instrumente). Finally the light is focused onto a quadrant photodiode (QPD, SPOT-9DMI, OSI Optoelectronics).

Improvements included: (1) Addition of an optical isolator (Thorlabs IO-3-1064-HP) to ensure that no light is reflected back into the laser. Back reflections need to be avoided as they lead to instabilities in the laser system such as mode hopping. (2) For expansion of the trapping beam, we used a beam expander with variable focus and

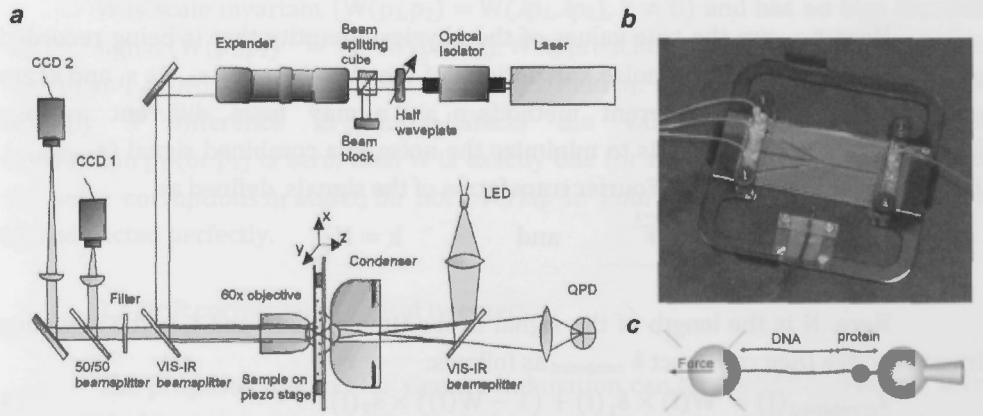


Figure 2.1: (a) Schematic drawing of the optical setup. (b) Photograph of our flow cell. (c) Schematic of a molecule tether grabbed from its two ends by two polystyrene beads one trapped by the focused laser and the other one is immobilized on the tip of a piezo-driven glass micropipette.

magnification (Thorlabs BE05-10-C). Variable magnification is very helpful for matching the beam waist diameter with the back aperture of the objective [70]. (3) Labview implementation to track bead positions with live visualization capabilities (with supports from software engineer S. Wouda). (4) Addition of a sound canceling box, to minimize acoustic noise in the room and thus minimize signal noise in the setup.

This optical tweezers setup was used for all the measurements presented in this thesis. While working with this tweezers new ideas emerged as to how to improve

the setup for chaperone studies. A noise reduction method was developed to correct for reducing noise in the measured bead-to-bead distance in the optical tweezers setup. A new dual beam setup with fluorescence imaging and control on the temperature of the flow cell was then designed.

2.2 Noise reduction by signal combination in Fourier space

Experimental setups may record the same physical quantity by two different methods. We will call these two signals s_1 and s_2. It is always possible to write such signals as

$$\mathbf{s}_1(t) = \mathbf{s}_{true}(t) + \mathbf{n}_1(t). \quad (1)$$

$$\mathbf{s}_2(t) = \mathbf{s}_{true}(t) + \mathbf{n}_2(t). \quad (2)$$

Here, s_{true} are the true values of the physical quantity that is being recorded over time. n_1 and n_2 are the noise corruptions of the signals s_1 and s_2. As s_1 and s_2 are measured using two different methods, n_1 and n_2 may have different average frequency spectra. The goal is to minimize the noise in a combined signal ($s_{combined}$). To this end, we combine the Fourier transforms of the signals, defined as

$$\hat{\mathbf{s}}(k) = \sum_{n=1}^{N} \mathbf{s}(n) e^{-2i\pi(k-1)\frac{n-1}{N}} \quad \text{and} \quad k = N\frac{f}{f_s}. \quad (3)$$

Here, N is the length of the signal in the time domain and f_s is the sampling frequency. We then construct $\hat{s}_{combined}$ as follows:

$$\hat{\mathbf{s}}_{combined}(f) = \mathbf{W}(f) \times \hat{\mathbf{s}}_1(f) + \big(1 - \mathbf{W}(f)\big) \times \hat{\mathbf{s}}_2(f). \quad (4)$$

W is a real-valued function, with values between 0 and 1. By substituting for $\hat{s}_1$ and $\hat{s}_2$ in Eq. (4), we get

$$\hat{\mathbf{s}}_{combined}(f) = \hat{\mathbf{s}}_{true}(f) + \mathbf{W}(f) \times \hat{\mathbf{n}}_1(f) + \big(1 - \mathbf{W}(f)\big) \times \hat{\mathbf{n}}_2(f). \quad (5)$$

The combined signal in the time domain is then the inverse Fourier transform of $\hat{s}_{combined}$. Importantly, $\hat{s}_{true}$ appears as a term that is separate from the corruption terms, which are weighted by W. Thus, noise in the combined signal can be suppressed by choosing the function W such that the terms after $\hat{s}_{true}$ in Eq. (5) are minimized. W must then be close to 0 in frequency regions where n_1 is dominant and close to 1 where n_2 is dominant. We quantify n_1 and n_2 by defining

$$\mathbf{p}_i(f) = \frac{1}{N} \times E(\hat{\mathbf{n}}_i(f) \times \hat{\mathbf{n}}_i(f)^*). \quad (6)$$

Here, E[.] denotes the average of many measurements (expected value) and * denotes complex conjugation. p_i can be seen as the average frequency spectrum amplitude of n_i. We define p_i in such a way that its mean is proportional to the variance of the noise corruption, if the noise has zero mean

$$\overline{\mathbf{p}_i(f)} = \frac{\sum_{k=1}^{N} \mathbf{p}_i(k)}{N} = E(\mathrm{Var}[\mathbf{n}_i(t)]) \qquad \text{if} \qquad E(\overline{\mathbf{n}_i(t)}) = 0. \qquad (7)$$

With $E(\mathrm{Var}[\mathbf{n}_i(t)]) = \mathrm{Var}[\mathbf{n}_i(t)]$ and $E(\overline{\mathbf{n}_i(t)}) = \overline{\mathbf{n}_i(t)}$ for sufficiently long signals. Overbar denotes the mean value. We then propose the following expression for W:

$$\mathbf{W}(f) = \left(\frac{1}{e^a - 1}\right)\left(\frac{1+e^a}{1+e^{a\,x}} - 1\right) \qquad \text{with} \quad \mathbf{x} = \frac{\mathbf{p}_1(f) - \mathbf{p}_2(f)}{\mathbf{p}_1(f) + \mathbf{p}_2(f)} \quad \text{and} \quad \mathbf{a} \in \mathbb{R}^+. \quad (8)$$

W is scale invariant ($W(p_1,p_2) = W(\lambda p_1,\lambda p_2)$, $\lambda \neq 0$) and has no bias towards either signal ($W(p_1,p_2) = 1-W(p_2,p_1)$). W approaches 0 as p_1 becomes larger than p_2 and approaches 1 as p_2 becomes larger than p_1. The factor a determines how strongly a difference in p_1 and p_2 affects the values of W. If the noise distribution p_1 (or p_2) is zero, then W is exactly one (or zero). In the special case that the noise corruptions n_1 and n_2 do not overlap in Fourier space, the signal can be reconstructed perfectly.

2.3 Drift correction in optical tweezers

The proposed method for signal combination can be used to correct for drift in an optical tweezers setup. In the optical tweezers setup considered here [18], one bead is held by a micropipette whose position is controlled by a piezo stage. A second bead is trapped in a focused laser beam. Forces acting on the trapped bead lead to deflections of the laser beam, which are recorded with a quadrant photodiode (QPD). The QPD voltage is proportional to the force acting on the trapped bead and the displacement of the trapped bead [71, 72]. The bead-to-bead distance is measured in two ways. One method uses the piezo position and the QPD voltage ($s^{QPD-piezo}$). This signal has nm spatial resolution and can be sampled at up to 1 kHz. However, on long time scales mechanical drift in the apparatus leads to corrupting changes in the bead-to-bead distance that remain undetected in $s^{QPD-piezo}$.

The second measure of the bead-to-bead distance is based on videotracking of the beads: the beads are imaged with a CCD camera, a particle tracking algorithm (LabVIEW) determines the bead positions, the difference of which is the bead-to-bead

distance (s^{CCD}). This method is limited in temporal and spatial resolution (30 frames/s and 5 nm, respectively), but is comparatively insensitive to drift, since the distance between the beads is determined directly from the CCD image. To apply the method to these two signals, we need to measure the average spectra of their noise corruptions p^{CCD} and $p^{QPD-piezo}$. Importantly, we mainly need to determine which of these two corruptions dominates at a certain frequency. This can be achieved with estimates for p^{CCD} and $p^{QPD-piezo}$.

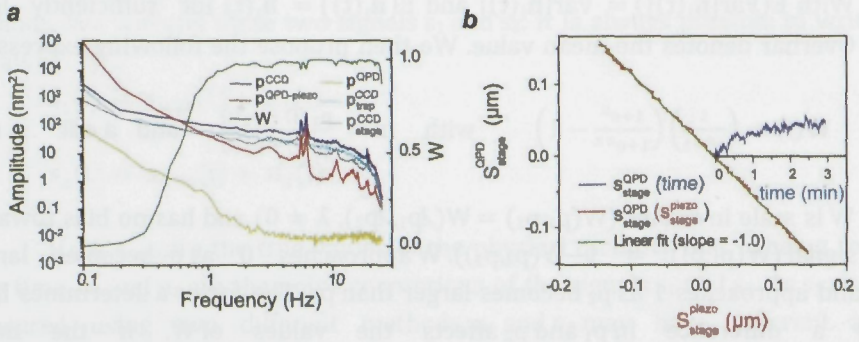


Figure 2.2: (a) Characterization of noise in the optical tweezers setup. p_B^A is the noise distribution of position B measured by method A. The subscript "trap" refers to the bead trapped in the laser focus, "stage" refers to the bead on the micropipette (which is connected to the stage). Bead positions can be measured by the QPD or the CCD camera and can be set by the piezo. Measured noise distributions and the resulting W function are plotted against frequency. W is close to one (or zero) at frequencies where p^{CCD} is much bigger (or smaller) than $p^{QPD-piezo}$. (b) Characterization of the QPD method. Piezo-induced displacements of the bead on the micropipette (s_{stage}^{piezo}) are measured with the laser beam focused on the bead, whose deflection is measured with the QPD (s_{stage}^{QPD}).[2,3] The linear relationship indicates that s_{stage}^{QPD} can accurately measure bead positions. Also displayed is s_{stage}^{QPD} as a function of time in the absence of piezo steering to the scale of the y axis. It shows that drift in this signal stays in the linear regime within the timescale of our experiments.

The noise in $s^{QPD-piezo}$ is composed of QPD and stage noise. Laser beam pointing instability, electronic noise, and mechanical drift in the optical components are the main sources of noise in the QPD signal. This noise (p^{QPD}) can be estimated by measuring the QPD signal with no bead in the trap. As can be seen in Fig 2.2.a, p^{QPD} is orders of magnitude smaller than all other noise sources considered here. Thus, the noise in $s^{QPD-piezo}$ appears to be dominated by the stage noise (n_{stage}). We can measure n_{stage} in the absence of piezo steering by pointing the laser beam onto the bead on the pipette (which is connected to the stage) and measure the beam deflections with the QPD (s_{stage}^{QPD}). This signal properly reports the stage movements, as shown by the linear dependence between the QPD signal and stage movements

induced by piezo steering (Fig 2.2.b). Care was taken to ensure that stage drift remained within the linear regime (Fig 2.2.b). In the absence of piezo steering ($s_{stage}^{piezo} = 0$), we can thus write

$$\mathbf{n}_{stage}(t) \cong \mathbf{s}_{stage}^{piezo}(t) - \mathbf{s}_{stage}^{QPD}(t) = \mathbf{0} - \mathbf{s}_{stage}^{QPD}(t). \quad (9)$$

In Fig 2.2.a it can be seen that $p^{QPD-piezo}$, which is the noise distribution corresponding to n_{stage}, is dominated by low-frequency stage drift.

The second measurement of the bead-to-bead position, s^{CCD}, represents the difference between the position of the bead in the trap (s_{trap}^{CCD}) and of the bead on the pipette (s_{stage}^{CCD}) as determined by videotracking. We can estimate the noise in both these signals, by comparing these videotracking signals with the low-noise QPD measurements of the bead positions s_{trap}^{QPD} and s_{stage}^{QPD}. s_{trap}^{QPD} denotes the position of the trapped bead as recorded by the QPD,

$$\mathbf{n}_{stage}^{CCD}(t) \cong \mathbf{s}_{stage}^{CCD}(t) - \mathbf{s}_{stage}^{QPD}(t). \quad (10)$$

$$\mathbf{n}_{trap}^{CCD}(t) \cong \mathbf{s}_{trap}^{CCD}(t) - \mathbf{s}_{trap}^{QPD}(t). \quad (11)$$

The corresponding p_{stage}^{CCD} and p_{trap}^{CCD} are displayed in Fig 2.2.a. p_{trap}^{CCD} has a slightly higher overall noise level than p_{stage}^{CCD}, which is likely caused by the Brownian motion of the trapped bead. The small noise increases towards lower frequencies in both may be due to slight movements of the camera or changes in focus. We determine p^{CCD} by adding p_{stage}^{CCD} and p_{trap}^{CCD}. Note that this is a slight over-estimate at low frequencies, because n_{stage}^{CCD} and n_{trap}^{CCD} are not fully uncorrelated as camera movements affect both in the same way. Furthermore, at high frequencies our estimate for p^{CCD} is a lower bound, because it was measured for stationary beads, while the movements that occur during molecular transitions lead to loss of sharpness in the CCD images and hence increased noise. However, these limitations of our estimate for p^{CCD} are not critical, as p^{CCD} is already larger than $p^{QPD-piezo}$ at high frequencies. As depicted in Fig 2.2.a, p^{CCD} intersects with $p^{QPD-piezo}$ at about 0.6 Hz. Thus, below 0.6 Hz the CCD method for determining the bead-to-bead distance exhibits the lowest noise, while above 0.6 Hz the noise in the QPD-piezo method is the lowest. This is expressed quantitatively by the W function, which can be determined from p^{CCD} and $p^{QPD-piezo}$ using Eq. (8).

To demonstrate our method, we recorded consecutive stretching and relaxation curves for a 2500 bp-long double stranded DNA with the optical tweezers

setup (Fig 2.3). DNA undergoes a characteristic overstretching transition when stretched with about 65 pN, during which the DNA length increases by about 70% [73, 74]. During relaxation, the DNA displays multiple discrete jumps in rapid succession[5] (Figs. 2.3.a,c), back to its normal (B-DNA) state. In between these stretching-relaxation cycles, the beads were held at close distance for a waiting time of about 1 min.

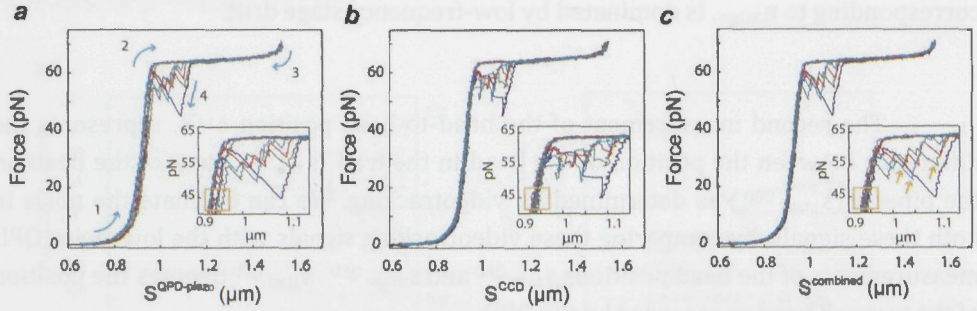


Figure 2.3: Force-extension data on a single DNA molecule held between the beads as measured by optical tweezers. (a) Force applied on the DNA molecule is plotted against $s^{QPD-piezo}$. The multiple stretching curves do not overlap due to stage drift at long time scales (orange box in the inset). The blue arrows indicate the timeline of one stretching-relaxation cycle. (b) The force is plotted against s^{CCD}. The stretching curves do overlap (orange box in the inset), but the fast movements of the bead in the trap during the hysteretic transitions are not well resolved. (c) The force is plotted against $s^{combined}$, the stretching curves overlap and fast movements during the hysteretic transitions are retained.

The three measures for the bead-to-bead distance (s^{CCD}, $s^{QPD-piezo}$, and $s^{combined}$) are plotted against force in Fig 2.3. The $s^{QPD-piezo}$ data show a significant spread along the distance axis for the stretching curves (Fig 2.3) that is not observed in s^{CCD} (Fig 2.3.b), in accordance with the observed higher noise at low frequencies in the QPD-piezo method (Fig 2.2.a). In contrast, the fast bead movements during the relaxation transitions are resolved by $s^{QPD-piezo}$ (Fig 2.3.a), but not by s^{CCD} (Fig 2.3.b). Next, these two signals were combined [75] using Eq. (4) from Sec. 1 and the W function determined in Sec. 2. The choice of a is not crucial because even at a = 0, the lower noise signal is preferred by linear weighting. In principle, the noise in the combined signal will be minimal for $a \to \infty$, as then the lower noise signal is exclusively used at every frequency. In practice, we are limited by the quality of our estimate for the noise distributions and how precise the crossing point in Fig 2.2.a can be located. Thus, we recommend an intermediate value for awhich reduces the effect of errors in noise distribution estimation. Here, we chose a = 5. The resulting $s^{combined}$ displays low long-term drift in the stretching curves (comparable to s^{CCD}) and also resolves the fast

transitions during relaxation (comparable to $s^{QPD-piezo}$). The combined data show that discrete jumps in different relaxation curves now overlap (orange arrows in Fig 2.3.c), indicating that the same DNA state is revisited in different stretching-relaxation experiments. This observation follows from the combined data and cannot be concluded with confidence from any of the two separate measurements individually (Figs. 2.3.a or 2.3.b).

2.4 Discussion and conclusion

The method for signal combination that we propose here does not rely on specific properties of the measured signal, which can be periodic or aperiodic. It does not suffer from the loss of information that occurs in other noise-suppression approaches such as filtering or smoothing. Our approach does rely on estimates for the noise distributions. Obtaining such estimates may not always be straightforward and care must be taken that they properly identify the lowest-noise signal during the actual measurement. Nevertheless we find that our method is relatively robust to inaccuracies in the noise estimates. For example, we approximated the function W displayed in Fig 2.2.a with different step-functions that step from 0 to 1 anywhere in between 0.25 Hz and 2 Hz. The resulting combined signals of the DNA stretching and relaxation experiments were then similar to the result displayed in Fig 2.2.c.

Noise distributions were calculated based on Fourier analysis. One may alternatively consider using the so-called Allan variance [75], which can quantify noise with higher resolution than Fourier analysis for noise with characteristic frequencies lower than about 1 Hz. Our method can be used with noise distributions based on Allan variance, though for the application discussed here the resolution of the noise distributions was not the limiting factor.

The proposed method assumes that the noise distributions do not change over time. When a source of noise is switched on and off during the experiment, this assumption does not hold. One could then consider combining signals in the time domain. Similar cases have been studied extensively in signal transmission, where this is called diversity combining. Diversity combining in Fourier space has recently been considered for in-car microphone systems [76]. There noise reduction is achieved by magnitude combining without phase estimation. If the noise distributions do not change over time, our method could be used for real-time signal combination in Fourier space. When combining the N most recent sample points of the respective signals, N should be large enough for the calculated Fourier transform to have

sufficient frequency resolution. In practice, the time needed for computing Fourier transforms of length N may significantly limit the sampling frequency of the combined signal.

The method described here is general and can be applied to any two signals that measure the same physical quantity. It might, for instance, be applied to AFM (or generally SPM) systems, when stage and tip positions are not only inferred from capacitive sensors in the piezo, but also from imaging [77, 78] or laser scattering [78]. Other areas where our method might be useful are chemometrics [79, 80], where drift correction for chemical sensors has been researched, and signal transmission, where diversity combining schemes similar to the method proposed here could be considered.

CHAPTER 2

CHAPTER 3

A polypeptide-DNA hybrid with selective linking capability applied to single molecule measurements using optical tweezers

Many applications in biosensing, biomaterial engineering and single molecule biophysics require multiple non-covalent linkages between DNA, protein molecules, and surfaces that are specific yet strong. Here, we present a novel method to join proteins and dsDNA molecule at their ends, in an efficient, rapid and specific manner, based on the recently developed linkage between the protein StrepTactin (STN) and the peptide StrepTag II (ST). We introduce a two-step approach, in which we first construct a hybrid between DNA and a tandem of two STs peptides (tST). In a second step, this hybrid is linked to polystyrene bead surfaces and Maltose Binding Protein (MBP) using STN. Furthermore, we show the STN-tST linkage is more stable against forces applied by optical tweezers than the commonly used biotin-STV linkage. It can be used in conjunction with Neutravidin (NTV)-biotin linkages to tether individual DNA molecules that can sustain applied forces above 65 pN for tens of minutes. The method is general and can be applied to construct other surface-DNA and protein-DNA hybrids. The reversibility, high mechanical stability and specificity provided by this linking procedure make it highly suitable for single molecule mechanical studies, as well as biosensing and lab on chip applications.

3.1 Introduction

Many experiments involving the manipulation of nucleic acids and proteins require multiple strong linkages that can be established *in-situ*, and can be used together and thus must be specific. For certain applications the molecules involved are immobilized on surfaces, either because the experimental setup requires fixing and controlling the position of the molecular ends or because the molecular phenomenon is measured using surface sensitive techniques [81, 82]. An example of an experiment demanding such supramolecular structures at surfaces includes the binding of liposome-ssDNA hybrids to surface immobilized-DNA in order to detect single nucleotide polymorphism using total internal refection [83]. Another example is the large-scale positioning of self-assembled functional DNA nanoarrays on surfaces [84], which have been used to construct arrays of quantum dots, proteins, and DNA targets. Supramolecular constructs that link micron-sized beads have been used to engineer molecular wires and to guide the assembly of nano and microstructures [85-88]. Metal wires have been fabricated by depositing metals on multi-protein and DNA constructs connecting the surfaces of two electrodes [89, 90].

Single molecule techniques such as optical tweezers have enabled the kinetic and thermodynamic characterization of DNA and protein molecules, as well as their interaction [18, 91-93]. In these methods, two ends of the molecule of interest typically are manipulated by linking them to surfaces, either directly or via molecular handles. Here molecular linkages are preferably established *in-situ* while still being able to sustain large forces over long timescales. Different classes of linkages have been used: Antibody-antigen linkages [18], the Streptavidin (STV)-biotin linkage [18, 93], covalent disulfide linkages [93] and covalent binding proteins (HaloTag or SNAP-tag) [94]. Each has its own strength and drawbacks. Antibody-antigen interactions are specific and diverse but affinities are affected by buffer condition, pH and temperature, and most are comparatively weak and thus limit the measurement time. Examples are Myc/AntiMyc and Dig/AntiDig. Moreover, many commercially available antibodies are polyclonal, causing variability in the force that the linkage can sustain. The Dig/AntiDig connection can be stable mechanically, and has therefore been used extensively to link DNA to surfaces. However, this system is less suitable for interfacing to proteins, while digoxigenin is also prone to oxidation and thus can deteriorate over time. Disulfide bonds are very strong but involve long preparation times and the molecules of interest must be resistant to redox reactions, which limits its applicability.

The biotin-STV interaction is one of the most broadly used, as it is strong and efficiently established. STV is one of the most stable proteins showing high resistance to temperature, urea, guanidine, and proteases [95]. This is in contrast to linkages such as HaloTag or SNAP-tag that unfold, aggregate and encourage nonspecific binding under these harsh conditions [94]. In the presence of SDS, Streptavidin begins to break up into monomers only at temperatures above 60 °C [96]. Because of the usefulness of biotin-STV interactions, efforts have been made to engineer variants and further optimize this system. Avidin is a glycosylated and positively charged protein (at neutral pH) which usually appears as a tetrameric biotin-binding molecule. Neutravidin (NTV) is a deglycosylated form of Avidin which is developed to decrease non-specific interactions [97]. It has recently been reported that Traptavidin, a mutant of STV, dissociates biotin more than tenfold slower, has increased mechanical strength and improved thermostability[94].

StrepTactin (STN) is another, recently engineered version of STV which has high affinity to biotin and in particular to its peptide ligand ($K_d \approx 1\mu M$), named StrepTag II (ST), which is 8 amino acids long (WSHPQFEK)[98]. STN has a tetrameric structure that provides four binding sites for ST. Additionally the binding can be reversed by adding Desthiobiotin which can in turn be removed by washing or dialysis. This feature has made the system popular for the purification and detection of proteins by affinity chromatography [98]. Interestingly, ST cannot bind Avidin, nor probably NTV, while STN does have affinity for biotin and ST can bind to STV (Kd ≈ 72μM) [99]. It has been reported that the binding affinity of ST to STV can be further increased to nanomolar levels when using multiple tandem STs [100]. It is also shown that in protein purification, having multiple tandem STs improves the binding affinity [98]. ST can be cleaved enzymatically, and the interaction is resistant to reducing agents (DTT and mercaptoethanol), denaturing agents (urea 1M), chelating agents (EDTA 50 mM) and detergents (SDS 0.1% and Triton X100 2%). ST is proteolytically stable, biologically inert and does not interfere with membrane translocation or protein folding [98]. The strength of the STN-ST linkage has been recently studied by Atomic Force Microscopy [101, 102], in which one single ST was fused to a protein and STN was anchored to a surface via PEG-based [102] or long protein-based [101] handles. The linkage showed an average dissociation force of 40 and 60 pN at pulling rates of 337 and 200 nms^{-1}, respectively [101, 102]. It is unclear what the dissociation force is for STN that is immobilized directly on the surface, and for multiple ST binding to a single STN.

The properties of the ST-STN linkage show promise for use in optical tweezers experiments and biomaterial engineering. These applications typically

require multiple linkages that are specific and strong, which ST-STN can potentially deliver. One challenge is to construct polypeptide-DNA hybrids, which would be required for such an approach. Oligonucleotides (6-16mers) conjugated to a tripeptide have been used for PCR amplification to successfully construct hybrids of DNA with short polypeptides [103]. The feasibility of synthesizing oligonucleotides conjugated to long polypeptides, and using them to amplify DNA segments, remains unclear.

We present a straightforward method to efficiently construct end-joined molecular hybrids in a manner that is mechanically stable and specific. Our method uses a tandem two STs (tST)-STN linkage to couple two arbitrary molecules A and B, where both A and B can be either DNA or protein of arbitrary size. In our method the assembly is carried out in a stepwise manner which leads to promotion of the heterodimer and suppression of other multimeres. We find that DNA molecules can be coupled well to the surface via tST-STN linkage. The linkage is more stable against applied force than the biotin-STV linkage and can be used in conjunction with biotin-NTV to stably tether DNA and to construct protein-DNA hybrids.

3.2 Results and discussions

To construct DNA molecules linked to polypeptide (tST-DNA), we used a primer covalently linked to the polypeptide. The PCR conditions were optimized to efficiently amplify the DNA from the template plasmid. By using gradient PCR, and testing several polymerases (Taq Polymerase and Phusion) and different PCR conditions, we found that comparatively long annealing and extension time (1 and 3 min per cycle respectively) allowed efficient amplification, resulting in a final yield of about 500 ng. The resulting construct was then characterized by agarose gel electrophoresis (Fig 3.1b) and later tested with an optical tweezers assay (Fig 3.2a).

To synthesize a protein-polypeptide hybrid, we chose Maltose Binding Protein (MBP) as our model protein. MBP is a protein with a variety of applications in biotechnology and biological research, widely used to prototype a variety of biosensing platforms [104]. It is also a model protein for folding and export studies and is commonly used as fusion partner in protein biochemistry [18, 105].

Escherichia coli strain BL21.1 was used to overproduce the MBP construct. Two repeats of the sequence encoding the ST (tST) were introduced at the C-terminus of MBP using plasmid pNN226 as a template. The tST-MBP hybrid was expressed at 37º C and purified from the crude cell extract using amylose resin affinity

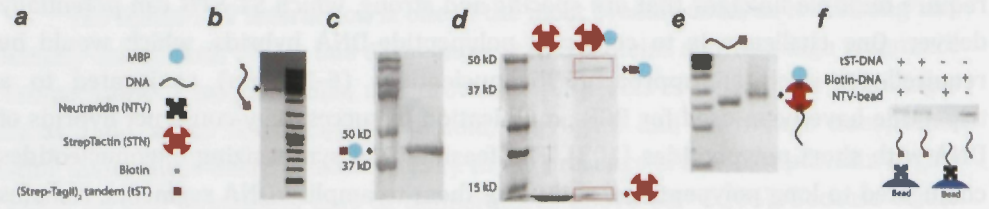


Figure 3.1: Hierarchical synthesis of protein-DNA hybrids. (a) Schematic drawing of the building blocks (b) 1% agarose gel demonstrating construction of tST-DNA-biotin hybrid at 2553 bps (c) SDS-PAGE analysis illustrating over production of tST-MBP in *Ecoli* BL21.1 (d) SDS-PAGE characterization of STN-tST-MBP hybrid after amylose column purification. STN decomposes into monomers upon boiling (e) 1% agarose gel confirming the formation of multi protein-DNA hybrid (f) 1% agarose gel showing the presence and absence of DNA strand in the supernatant of incubated NTV beads by tST-DNA and biotin-DNA respectively. Biotinylated DNA easily binds to NTV, tST labelled DNA does not and remains in the supernatant.

chromatography (New England BioLabs). The hybrid was then tested with SDS-PAGE (Fig 3.1c), which showed a molecular size between 37-50 kDa that corresponded well with the expectations for tST-MBP (~42.5 kDa).

Here we aimed to optimize the specific formation of a hybrid between MBP and DNA using ST-STN linkages (MBP-tST-STN-tST-DNA). To make this construct we first mixed STN and tST-MBP in 10:1 ratio. Unbound STN was removed by amylose column purification. tST-MBP bound to amylose column was then eluted with maltose. SDS-PAGE (Fig 3.1d) showed two bands for eluted sample, with one corresponding to tST-MBP and one to STN only, thus showing that STN had successfully been bound to MBP. The previously constructed tST-DNA was then mixed with a large excess of the MBP-tST-STN hybrid (> 30 molar excess) in order to favour binding of a single DNA molecule to each MBP. Agarose gel analysis showed a band distinctly above from tST-DNA, consistent with the formation of a MBP-tST-STN-tST-DNA hybrid (Fig 3.1e). As expected, MBP-tST-STN-tST-DNA hybrid shows a significantly reduced mobility as compared to tST-DNA due to its larger size and higher molecular weight. The successful formation of the complex hybrid, also confirms the chemical structure of the constituting hybrids synthesized in the previous steps.

To study the specificity of NTV binding, NTV-coated beads were incubated either with tST-DNA or with biotin-DNA. After 30 min, beads were removed by centrifuging and supernatants were loaded onto an agarose gel (Fig 3.1f). The results showed that biotinylated DNA bound the beads efficiently, as no DNA could be detected in the supernatant. In contrast, all of the input tST-DNA remained in

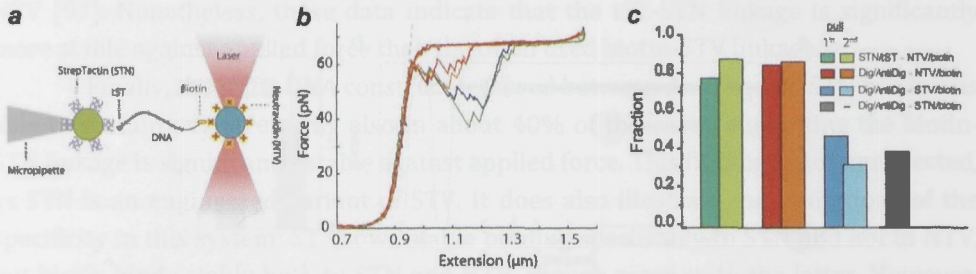


Figure 3.2: Mechanical stability analysis. (a) Optical tweezers setup (b) Force-extension curve of dsDNA showing overstretching at 65 pN, as well as the characteristic step-wise relaxation. The measured DNA stretching curves did not display additional steps that might have arisen from STN unfolding or its detachment from the surface. (c) Fraction of tethers that resisted 60 pN in first and second pull, compared between several commonly used linkage strategies and our proposed linkage strategies based on STN. For the AntiDig/Dig-DNA-biotin/STN system, almost all tethers broke at the first pull, and hence the subsequent pulls are not indicated.

supernatant, showing no affinity to the beads. These results indicate that NTV binds selectively to biotin and not to tST, which is a central requirement for instance for efficiently tethering tST-DNA-biotin constructs between STN- and NTV-coated beads. Without high selectivity, incubation of the tST-DNA-biotin construct with NTV beads would have resulted in the attachment of both DNA ends to the same (NTV) bead.

To measure the mechanical stability of the linkage between tST and surface-bound STN, we pulled on a single synthesized DNA-tST-STN hybrid using optical tweezers. First, we immobilized tST-DNA-biotin constructs on NTV-coated beads using the biotin-NTV linkage while keeping the tST-end free (Fig 3.2a). The NTV beads were titrated with varying amount of tST-DNA-biotin so that only few DNA constructs were linked to one bead. Next, the tST-STN linkage was established *in-situ* to beads coated with STN. Pulling curves showed overstretching at 65 pN, which indicated the presence of a single tether, and showed the tST-STN linkage was able to sustain such forces without breaking (Fig 3.2b). The measured DNA stretching curves did not display additional steps that might have arisen from STN unfolding or its detachment from surface.

Next, we performed a quantitative comparison of the mechanical stability of the tST-DNA-biotin and the Dig-DNA-biotin handles. The latter is often used in optical tweezers studies in conjunction with STV- and AntiDig-coated beads (14,30). Note that in general, NTV-coated beads have advantages compared to STV-coated beads,

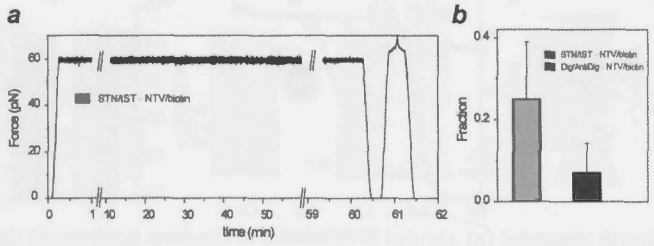


Figure 3.3: Mechanical stability analysis at constant force. (a) An example of stretched tether kept under a constant force of 60 pN by means of a force feedback for more than one hour. A stretching and relaxation cycle at the end of the experiment displays a typical behaviour of dsDNA. (b) Fraction of the tethers resisting more than 10 min at 60 pN.

given the higher affinity of NTV for biotin (18). Pulling experiments were thus performed on Dig-DNA-biotin constructs tethered between AntiDig- and NTV-coated beads, and on tST-DNA-biotin tethered between STN- and NTV-coated beads. We considered a tether was established when the connections could sustain 20pN. Connections that broke below 20pN were disregarded. The constructs were then stretched and relaxed multiple times with a displacement speed of 50 nm/sec to just beyond the DNA overstretching regime at about 65pN, until the connection broke (N=111 for the tST construct, N=230 for the Dig construct). We monitored the fraction of tethers able to sustain DNA overstretching, and distinguished first and subsequent pulls. Overall, we found quite similar results for the two constructs, with about 80% of the tethers able to sustain overstretching (Fig 3.2c). These data suggest that the *in-situ* tST-STN linkage has similar stability against applied force as Dig-AntiDig in the first pull.

The stretching experiments indicated a number of additional points. For instance, for the tST-STN construct, subsequent pulls show a slight increase in the fraction of times the tether survives overstretching (Fig 3.2c, from 77% to 87%). A possible explanation for this increase could be the proposed bimodality of the ST-STN interaction [102]. The origin of this bimodality is believed to lie in the interaction of a single ST with a single or multiple sites on STN, where the latter is supposed to be somewhat less stable. Next, additional experiments indicated that the biotin-STV linkage was significantly less stable than the biotin-NTV linkage: Dig-DNA-biotin tethered between AntiDig and STV beads were able to sustain overstretching only in 40% of the cases, about half of what was found when using AntiDig and NTV beads. An increased stability of biotin-NTV compared to biotin-STV is consistent with bulk values for the equilibrium binding constant, which are significantly lower for biotin-

NTV [97]. Nonetheless, these data indicate that the tST-STN linkage is significantly more stable against applied force than the often used biotin-STV linkage.

Finally, the same DNA construct tethered between AntiDig and STN beads was able to sustain overstretching also in about 40% of the cases, suggesting the biotin-STN linkage is significantly stable against applied force. This finding is not unexpected, as STN is an engineered variant of STV. It does also illustrate the limitations of the specificity in this system: ST shows stable binding specifically to STN and not to NTV, but biotin binds stably both to STN and NTV, though more so to the latter. However, our protocol shows that these limitations can typically be overcome in practice, by first establishing the connection to biotin, which is less specific, and only then form the connection to tST which is specific.

In order to probe the difference in stability between Dig-AntiDig and tST-STN more exhaustively, we tested these linkages for their ability to sustain high forces for long periods of time. Fig 3.3a illustrates a case of STN/tST-DNA-biotin/NTV handle that could survive this load for an hour. The handle was stretched to 60pN (in under 1 min) and kept under force feedback for 60 min without breaking. Next, it was relaxed (Fig 3.3a, in between 60:00 and 60:30 min:sec) and showed a characteristic cycle of DNA overstretching (Fig 3.3a, in between 60:30 and 61:30 min:sec). The tether broke after additional 22 pulling cycles. Importantly, the fraction of strong tethers resisting more than 10 min at 60 pN was found to be significantly higher with tST-STN as compared to Dig-AntiDig (Fig 3.3b). Thus, the tST-STN linkage is able to withstand high forces for longer than the Dig-AntiDig linkage.

3.3 Conclusions

We have presented a simple procedure to specifically attach a protein to a DNA molecule, using STN-tST linkages. The method is rapid and straightforward, and can be established *in-situ* within biologically relevant buffers. Binding of the DNA-tST construct to surface immobilized STN shows high mechanical stability, and can readily tolerate forces as high as 65 pN for tens of minutes. The engineered linkage can be used as a reliable linker for optical tweezers studies of proteins and nucleic acids, both in constant pulling rate and force modes (32-34).

The motivation to use STN to end-join two molecules was based on the fact that the linkage is among the strongest noncovalent intermolecular bonds known. The specificity, stability, and rapid in-situ formation of the STN-tST complex allows it to be used in combination with other well-used linkages that can also be stably formed in-situ, such as NTV-biotin. Dig-AntiDig linkages of similar stability can be formed, but

they require bulk incubation. Thus, choice of linkage depends on the precise application and formation possibilities. We find that tST-STN is more stable against applied force than the commonly used biotin-STV linkage. Moreover, we show that tST-STN can be used for surface attachments as well as for linkage between DNA and protein molecules, which has not been achieved for Dig-AntiDig linkages. Because of the high stability of STN, this complex could potentially also be used in a broad thermal range and harsh conditions.

We have shown that constructing tST-DNA hybrids is straightforward using PCR amplification, making our method suitable for broad applications. For single molecule studies, the presented approach could be applied in combination with other peptide-DNA hybrids. For example, halo tags-DNA hybrid could be constructed as a handle and be linked covalently to halogenase-coated beads. Similarly, a peptide substrate to ubiquitin ligase could be used to generate peptide-DNA hybrid and then be linked to the protein ligase-coated bead. The reversibility of the ST-STN reaction, using Desthiobiotin [99], will make the ST-STN linkage also highly suitable for biologically inspired soft matter systems, where reversibility could open up new possibilities.

3.4 Materials and methods

Design and synthesis of the oligo-peptides

A tandem arrangement of two STs (tST: WSHPQFEKWSHPQFEK) was chemically synthesized and was linked to the primer (5'GTC TCG CGC GTT TCG GTG ATG ACG GTG 3') from its 5' end via a linker (-Cys-SMCC-C6) (BioSynthesis Inc.). The product was purified by HPLC and characterized by mass spectrometry (Applied Biosystems Voyager System 2051).

Synthesis of dsDNA-tST

The 2553 bps DNA handles were generated by PCR using Taq DNA polymerase and pUC19 plasmid DNA (New England BioLabs) as template. 500 ng of handles were generated at a time using 50 µl of PCR reaction. The two types of handles (with and without biotin) were generated using the above oligo-peptide as a forward primer together with the primer 5' TA6GTA6CCGCTCATGAGAC 3' as a reverse (6 is biotin-dT for biotinylated DNA). Polymerase chain reaction reagents for each 50 microliter reaction volume included: 1 unit of Taq polymerase (New England BioLabs), 5 µl of 10x PCR buffer (New England BioLabs), 10 pmol of the forward primer and 10 pmol of reverse primer, 5 µl of 2mM dNTPs (Fermentas), and 50 ng of

the plasmid DNA. The PCR profile was as follows: 1 min at 94°C, 30 cycles of 30 s at 94°C, 60 s at 52°C and 3 min at 72°C, finally followed by 10 min at 72°C and a 4°C soak.

Gel analysis

DNA samples were analyzed by gel electrophoresis (Fig 3.1b, 3.1e and 3.1f) in non-denaturing 1% agarose gels in 0.5xTBE buffer at 80 V/cm. Agarose gels were stained with ethidium bromide (EtBr). Protein samples were applied on an 8% SDS-PAGE gel in 1x running buffer (190 mM Glycine, 25 mM Tris-base and 0.1% SDS) at 180 V/cm. SDS-PAGE gels were stained with Coomassie InstantBlue (Expedeon Ltd.)

Functionalization of polystyrene colloids with STN and NTV

Carboxylated polystyrene beads (Polysciences Inc.) were covalently linked to STN (and NTV) via Carbodiimide reaction (PolyLink Protein Coupling Kit, Polysciences Inc.). Briefly, the beads were washed and then mixed with freshly prepared EDCA and STN (NTV) and mixture was incubated for 3 hours. STN-coated beads were stored at 4°C until use. DNA-coated microspheres were made by mixing ~70 ng of tST-dsDNA and 1 µl NTV-coated beads in 10 µl HMK (50 mM Hepes, pH 7.6, 100 mM KCl, 5 mM $MgCl_2$) buffer. After 30 minutes incubation on a rotary mixer (4°C), the beads were dissolved in 400 µl HMK buffer for use in optical tweezers experiments.

CHAPTER 4

Topological determinants of protein folding and unfolding

Contact topology is introduced to simplify and explain structure, complex folding and unfolding pathways of large proteins. The topology of a protein, a nucleic acid or a structured polymer puts fundamental constraints on its unfolding transitions and can be used to find general selection rules irrespective of the chemical nature of molecules. Protein topology together with contact energies mapped on the topology can be used to predict the exact length states and the most probable unfolding path in pulling experiments. Contact topology can also be used in the context of assisted protein folding. Motivated by recent finding on Trigger factor-protein contacts, we show how these chaperones can be seen and modeled as topological modulators.

4.1 Introduction

4.1.1 Topology

Topology is a mathematical notion that has been used to describe properties of objects that remain unchanged under a certain kind of continuous, invertible and one-to-one transformation (so called homeomorphism). Examples of such transformations include bending, stretching and shrinking. Objects like square, circle and triangle on a sheet of paper are inter-convertible by such transformations and therefore belong to the same topological class. "Figure 8" instead is not inter-convertible to any of these three objects unless a connection is ripped apart (Fig 4.1a). Connectivity is a qualitative property that is invariant under homeomorphism. Topology can therefore be seen as the study of connectivity and continuity. The field has been instrumental in the developments of the general theory of relativity, quantum field theory, cosmology, fluid mechanics and information science among others.

4.1.2 Knot topology

One interesting area of topology is the knot theory that studies the topological features of knots. We know intuitively that upon stretching a rope knot, certain aspects of the knot remains identical and this is referred to as the topology of the knot. In contrast the end-to-end distance of the rope changes upon stretching and thus this property is not a topological one.

For a mathematician, a knot is created by starting with a linear chain, wrapping it around itself arbitrarily, and then fusing its two free ends together to form a closed loop. A knot can then be represented with its projection on a plane, the so called knot diagram. How can one characterize the topology of a knot? For a given knot, there is a projection which minimizes the number of chain crossing (Fig 4.1b). By characterizing these crossings, one can explain the topology of the knot, because these crossings cannot be changed without tearing or gluing the chain. For example, two topologically equivalent knots must have equal number of crossings. The reverse statement does not necessarily hold true. Further characterization of the topology (e.g. analyzing the mirror image and braids) is needed for unique identification.

4.1.3 Molecular topology

Knot theory has been applied to elucidate topological features in molecules and in particular macromolecules. Here the connectivity is defined based on covalent bonds. Several topological classes have been distinguished: branched topologies and

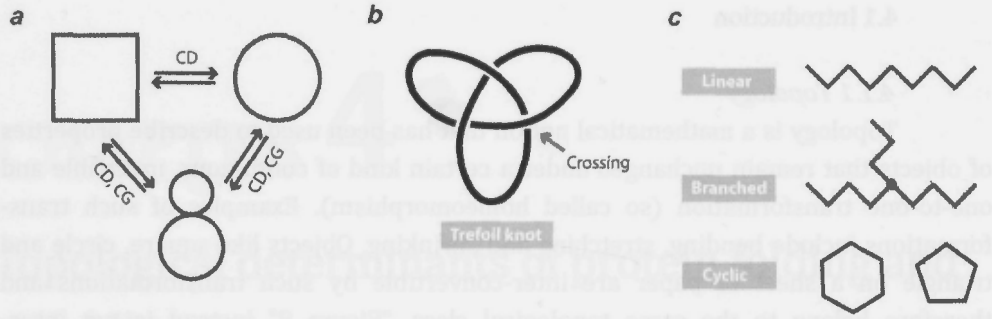


Figure 4.1: Topological polymer chemistry (a) Continuous deformation (CD) changes a shape to another topologically equivalent shape. Gluing/cutting (GC) operation changes a topology to another topology. Circle and rectangle have identical topology which differs from the topology of the "figure 8". (b) The Trefoil knot is a knot topology seen in RNA methyltransferase. The Trefoil knot has three crossings in its minimal representation [106]. (c) Hydrocarbons can be classified by their topologies: Examples of linear, branched and cyclic topologies are presented. Multiple topological invariants can be identified such as the total number of chain ends and of branch points[107].

cyclic topologies [107]. In this topological classification (Fig 4.1c), the following properties are invariants under continuous deformations: the total number of chain ends (termini); the total number of branch points (junctions); the number of branches at each junction and the connectivity of the junction. As such linear polymers will all fall into one topological class.

Topological features of linear polymer chains are not limited to the topology discussed above. Polymers in solution are dynamic entities and thus span a conformational space. These conformations can also be classified based on their topological features. At room temperature, sufficiently long equilibrated polyelectrolyte chains such as DNA molecules show self-entanglement and can form knots [108]. The incidence of knots increases with the chain contour length with an exponential dependence [109]. This topological feature of the linear chains is important for processes such as translocation of the chain through a nano pore [108]. These processes can in turn be used to characterize the topology of a chain.

4.1.4 Protein topology: Topologies based on covalent connectivity

Proteins are examples of linear polymers and thus can be classified into topological classes described above (Fig 4.1b,c). A protein chain has two termini and no branch points and thus falls into the same topological class as n-alkanes. Proteins can also be seen as poly-electrolytes and in appropriate buffers with denaturants, they can also self entangle and form knots.

Folding can generate knots, however most proteins have unknotted topology in their native state. Only 300 proteins in the Protein Data Bank, including rRNA methyltransferases, carbonic anhydrases and ubiquitin hydrolase, are identified as knotted [110]. Knot topology has serious implications for the folding process: knot formation is believed to occur via the formation of intermediate folds with attractive non native interactions [111].

4.1.5 Protein topology: Topologies based on covalent and non-covalent connectivity

Protein folding involves formations of non-covalent contacts between aminoacids. As discussed above, connectivity is a qualitative property that is invariant under homeomorphism. For a linear polymer with a number of intra-molecular contacts, a topology can therefore be defined based on these contacts and their relations. Any continuous deformation (e.g. stretching) which is devoid of tearing or gluing, renders the contacts and the inter contact relations unchanged.

In structural biology contacts between two aminoacids are defined based on a distance threshold. From the network of aminoacid contacts, some topological features of proteins have been inferred such as node degree, clustering coefficients, betweenness and closeness centrality. How do these network topological ingredients affect protein folding? It is commonly believed that the topology of the native structure determines the sequence of folding events (David Baker, Nature 2000). This is largely rooted in the success of the coarse grained protein models and the insensitivity of the folding transition states to single point mutations [112, 113]. The average graph connectivity is shown to be highly correlated with folding probability of the transition state [114]. Other macroscopic measures of protein structural and energetic properties such as radius of gyration, rms distance, solvent-accessible surface area, and potential energy fail to serve as predictors of the probability of a given conformation to fold. For small proteins, however, range of interaction and number of contacts correlate with the folding rates (Fig 4.2.). For larger proteins, these considerations alone cannot explain the folding rate differences between proteins. These network topological considerations are statistical and global properties and therefore provide insight on the kinetics of the folding process and not much on the details of (un)folding pathways. When interaction strengths are considered, more detailed information about the intermediate conformations visited during folding and unfolding path are found [115].

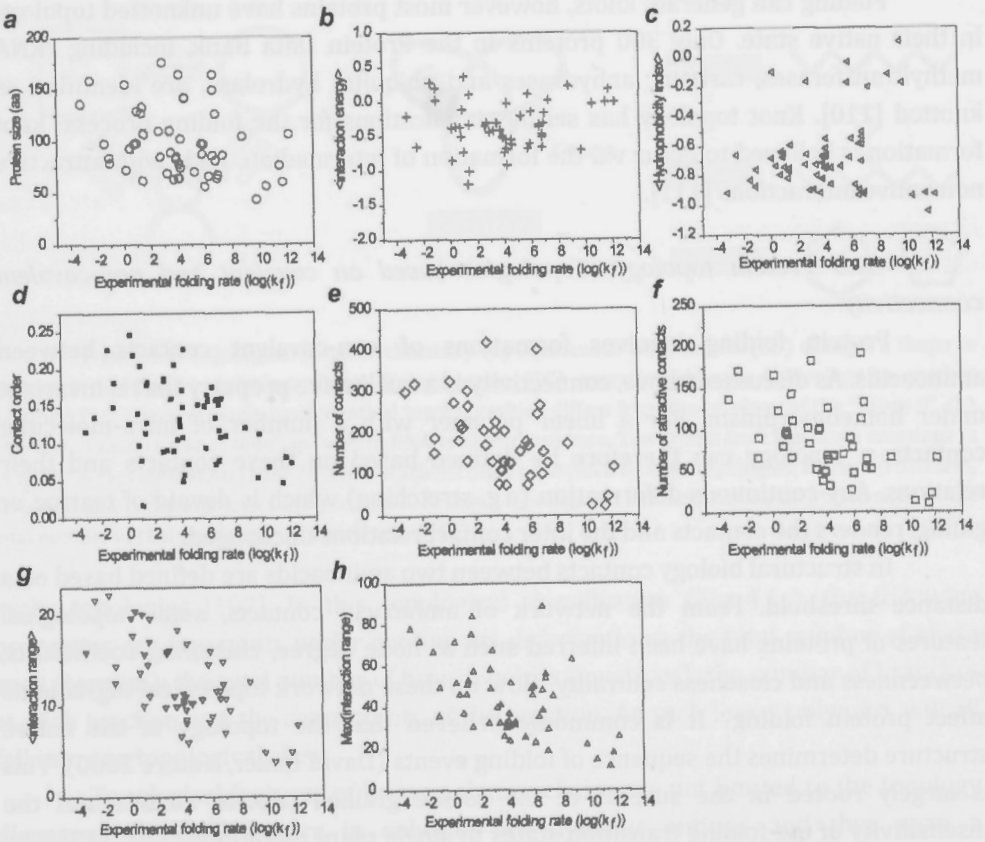


Figure 4.2: Folding rate determinants of small proteins. PDB of 59 proteins with a range of secondary structures (all alpha, all beta, mixed) were analyzed and compared to their experimental folding rates reported in the literature [116]. For energy estimation and hydrophobicity, Thomas Dill energy scale and Kyte-Doolittle scale were used respectively. Large proteins reported by Gromiha et al. [116] appear as outlier in these plots and are not shown.

4.2 Contact topology

Here we introduce another type of topological features of linear polymers with intra-molecular interactions. We aim to go beyond the statistical description of the contacts as provided by the network topology and characterize the inter-contact relations. We show that this notion can be used to probe the structure, folding and unfolding pathways of proteins. To show how, let's consider secondary structural motifs that are commonly present in proteins and let's stick to the commonly used definition of aminoacid-aminoacid contact based on distance threshold of ~ 7Å. Our approach is however general and can be applied to other molecules and with other

definitions for intra-molecular contact such as inter-aminoacid H-bond contacts and β-β contact in proteins (Fig S4.1).

For a feature to be topological, it has to be preserved under deformations that do not involve emergence or disappearance of contacts. A process that involves disruption of a contact or emergence of a contact will change the total number of contacts and therefore changes the topology. Instead when contacts emerge and disappear with preservation of the total number of contacts, topology of the system will only change, if the "arrangement" of the new contact set differs from the original arrangement.

Fig 4.3 provides a few examples of the secondary structural motifs that are common in proteins. Let's consider the example of a β-hairpin and choose two aminoacid pair contacts within the structure as shown in Fig 4.3a and disregard the rest. To understand the arrangement of these contacts, two representations are provided in Fig 4.3a, namely contact matrix (map) and bubble graph. It can be seen that the two contacts are arranged parallel to each other with one being contained in the second one. There are a few interesting features in this arrangement: (1) the relation between the two contacts is not symmetric. (2) the relation between the two contacts would not change if we replace the "turn" in the β-hairpin with any arbitrary structure. The second structure presented in Fig 4.3a has a similar bubble graph and contact map to those of the β-hairpin: the bubble graph of the former will only have a larger lower loop and can be converted to the bubble graph of the β-hairpin with continuous deformation (shrinking of the lower loop). (3) the arrangement of contacts will impose an order on the disruption of the contacts when we pull on the bubble graph from its termini (thought experiment): always the upper contact will break before the lower one is ripped apart.

For the chosen pairs of binary contacts in the structural motifs shown in Fig 4.3, several arrangements are distinguished: parallel (P), series (S) and cross (X). The displayed contact arrangement is a topological property because transition from a topology to another one (for example from P to S) would involve breaking the contacts between strands while transition between two conformations with identical topologies can happen smoothly with continuous deformation. Furthermore, the three topological relations introduced above form a complete set of relations. Note that, at this stage we only allow for mono-valent interactions: more complex cases will be discussed later. Inter-contact relations are formally defined in Box I based on the frameworks of the set theory and discrete mathematics.

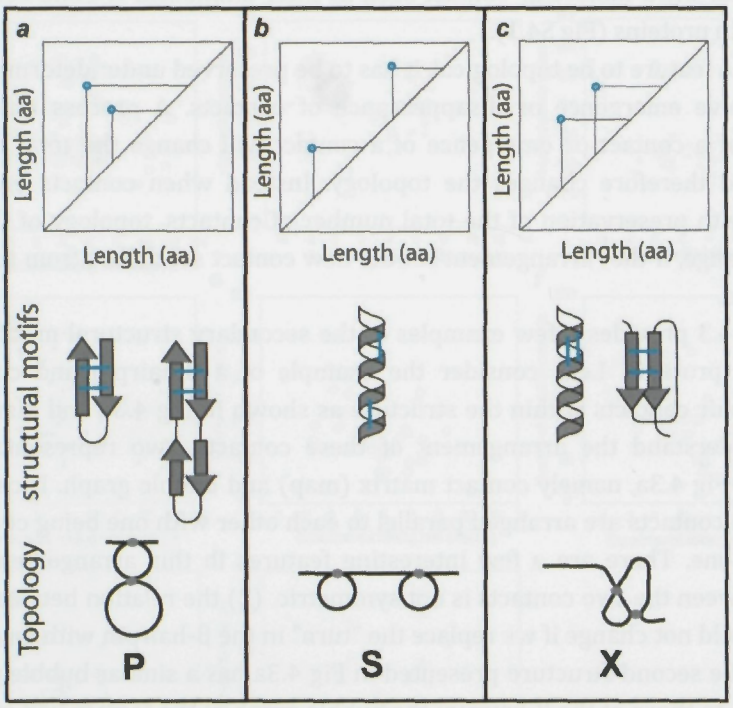


Figure 4.3: Arrangement of inter-residue H-bond contacts in basic structural motifs of proteins. Structures, contacts maps and loop diagrams are displayed for each motif. Three arrrangements are distinguished: Parallel (P), Seris (S) and Cross (X). (a) A β-hairpin and and generated by replacing the loop in the β-hairpin with a second β-hairpin. The motifs with the selected paralell contacts are topologically equivalent. (b) An α-helix with two selected contacts in series arrangement. (c) Two motifs with selected crossing contacts.

The inter-contact relation introduced above is a binary relation between contacts. How can we use this binary relation to describe the topology of a protein with many contacts? Consider a protein with multiple contacts (Fig 4.4a *i*). Any two contacts can be assigned to one of the following topological classes: parallel (P), series (S) and cross (X). To do the assignment, we keep the two contacts and remove the rest, the contact pair can be readily allocated to one the topological classes. This inter-contact relation has interesting mathematical properties that simplify the assignment. For example if we know the contact relation between contact A and B as well as B and C, we might already know the relation between A and C. P relation is for example transitive while S and X are not (Box I, Table I and Fig 4.4a).

Contact topology can be used to classify proteins to topological classes. For two proteins to be in one topological class there will have to contain the same number of contacts. Any two such proteins that have the same topology can be inter-converted

---------- BOX I: Protein topology ----------

Consider a linear chain (interval of $I = [1{:}n] \subset N$) made of monomers $i \in I$ (for example aminoacids or secondary structures) with some forming contacts: $C \subset \{(i,j)|i,j \in I \wedge 1 < (j-i)\}$. We only allow for binary interactions and define a contact as a binary interaction. Multivalent interactions will be considered later. Numbering is done from N- to C-terminal. Every contact will be associated with a length $L\big((i,j)\big) = |j-i|$. We define an adjacency matrix A as:

$$A(i,j) = 1 \quad if\ |i-j| = 1 \vee (i,j) \in C$$
$$A(i,j) = 0 \quad otherwise$$

$\forall C_1 \& C_2 \in C$ with $C_1 = (i,j)$ and $C_2 = (r,s)$, we define the following relations:

Parallel relation: $C_1 \mathbf{P} C_2 \Leftrightarrow [i,j] \subset [r,s]$

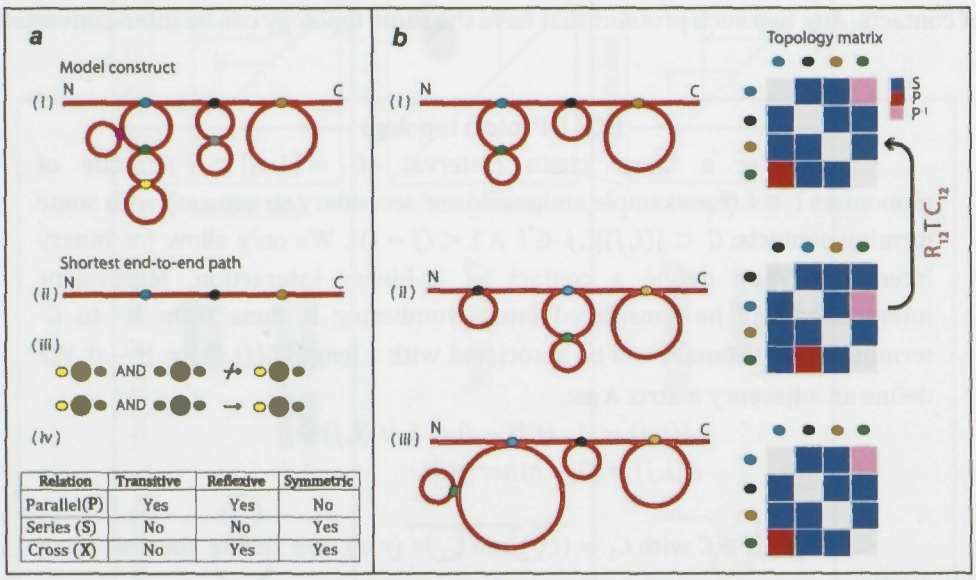
Series relation: $C_1 \mathbf{S} C_2 \Leftrightarrow [i,j] \cap [r,s] = \emptyset$

Cross relation: $C_1 \mathbf{X} C_2 \Leftrightarrow [i,j] \cap [r,s] \notin \{\emptyset, [i,j], [r,s]\}$

by continuous deformation. Some operations leave the topology of the protein unchanged (Fig 4.4b). An interesting operation is permutation of different parts of the protein that are in series (Fig 4.4b *i* to *ii*). Such operation has been recently performed on T4 Lysozyme, and it turned out that the topology critically determines the folding cooperativity in this case [117].

Topological relations between contacts can be listed in the form of a topology matrix (Box II). The matrix is not unique and its form depends on the labeling of the contacts. However all matrices belonging to one topology make a space which has a particular structure. The matrices can be inter converted by concerted row-column operations and will have identical spectral properties (Fig 4.4b). There are alternative ways to construct topology matrix other than the one introduced in Box II. From adjacency matrix A that contains both topological and length information, one can readily construct a reduced matrix lacking the length information and only carrying

the topological information. Removal of zero rows and columns of the adjacency matrix will result in a "topology matrix".



Relation	Transitive	Reflexive	Symmetric
Parallel(P)	Yes	Yes	No
Series (S)	No	No	Yes
Cross (X)	No	Yes	Yes

Figure 4.4: (a) Bubble graph of a hypothetical protein is illustrated in (i) and the corresponding shortest path is shown in (ii). (iii) shows that while P is a transitive relation, S is not. (b) Proteins within a topological class can be converted to each other by continuous deformation (i, ii, iii). One interesting case is permutation of subparts of a protein that are in series (i, ii). In this case the constructed topology matrices can be inter-converted by row/column operations described in the panel. Consequently the two topology matrices have the same spectral properties.

As mentioned before, the topological properties discussed above are general and can be generalized to various definitions for intra-molecular contacts. One choice of contacts are inter-aminoacid contacts that are energetically strong or highly conserved in evolution. One can for example ignore the contacts between restudies that repel each other as well as weakly attractive ones.

4.3 Contact topology and topological selection rules

In knot theory, pulling the termini of two knot and unknot proteins, allows us to distinguish them, as the later is disentangled while the former progressively tightens [118]. Similarly, pulling on a protein and breaking the contacts show the signature of the contact topology of the protein.

Topological constraints forbid certain orders of tearing contacts and therefore certain transitions between conformations cannot be visited in a pulling experiment (Fig 4.5). We call these constraints on transitions, topological selection rules in analogy with spectroscopy. In Fig 4.5 the transition from a state with end-to-end length of zero to length L, is forbidden when the contacts are parallel (in contrast to S and X relations). For a system with two contact sites in parallel, the inner contact can be removed only after the outer one is dissociated while, in series topology, both have a chance to open up.

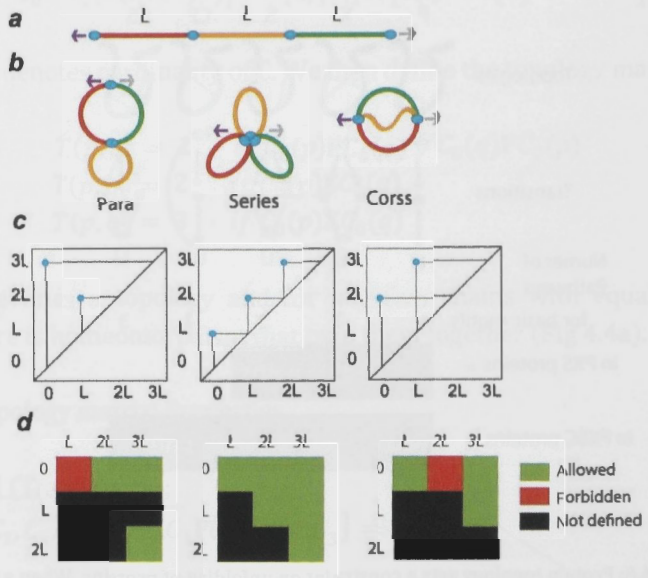


Figure 4.5: Topological selection rules. (a) Primary structure of the model polymer. (b) Three folded conformations of the polymer with three distinct contact topologies: Parallel, Series and Cross. (c) Corresponding contact maps of the three conformations in (b). (d) Topological selection rules for each topologies. (x,y) means a transition from x to y. Red denotes a forbidden transition and green illustrates allowed ones. Transitions in black are not meaningful because a transition has to lead to a length increase.

For a protein with contacts that are in parallel (P), series (S) or cross (X) relations with each other, the number of possible unfolding paths can be calculated easily. A parallel relation only allows one route while series and cross relations allow for two routes each. The number of possible pathways will therefore be $2^{(\text{number of S and X relations})}$ (see Fig 4.6).

The situation will be more complicated if we allow for bivalent contacts in which a contact site can form contact with two different contact sites. Such contacts may show cooperativity and dissociation of one may influence the stability of another (concerted "C" topologies). For proteins containing contacts that are in concerted

topological relations (Concerted parallel CP and concerted series CS), the number of possible paths can not be calculated simply by multiplying the number of possible routes for each contacts (Fig 4.6).

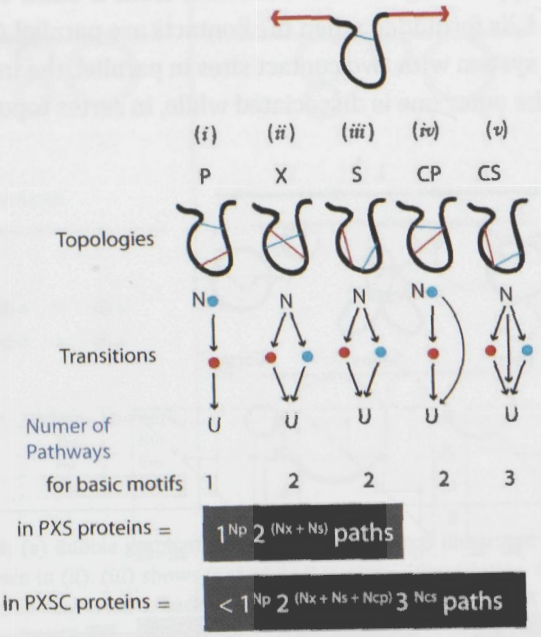


Figure 4.6: Protein topology sets a constraint on unfolding of proteins. When a protein is unfolded by mechanical force applied to its ends, the protein visits lengths defined by its contact position, contact energy and topological relation between them. For a system with two contact sites in parallel (*i*), the inner contact can be removed only after the outer one is dissociated. In series topology (*iii*), both have a chance to open up with a probability that depends on the contact energies. Contacts may also show cooperativity and dissociation of one may influence the stability of another (concerted "C" topologies: concerted parallel (*iv*) concerted series (*v*) topologies).

----------BOX II: topological constraints ----------

Topology matrix: Definition

We construct C_0, by putting contacts in order of the magnitude of their first element:

$$C_0 = \{(t(i),(i,j))\,|\, t \in [1,|C|] \wedge [t(i) < t(i') \Leftrightarrow i < i']\}$$

|C| denotes cardinality of C. We then define the topology matrix T as,

$$T(p,q) = 1 \quad if\ C_0(p)\mathbf{P}C_0(q) \vee C_0(q)\mathbf{P}C_0(p)$$
$$T(p,q) = 2 \quad if\ C_0(p)\mathbf{S}C_0(q)$$
$$T(p,q) = 3 \quad if\ C_0(p)\mathbf{X}C_0(q)$$

T defines a topology and for any two chains with equal topology matrix, there is homeomorphism that map them together (Fig 4.4a).

Topology matrix: Properties

Chain rule I (Transitivity):

$$\forall C_1, C_2, C_3 \in C:\ [C_1\mathbf{P}C_2 \wedge C_2\mathbf{P}C_3] \Longrightarrow C_1\mathbf{P}C_3$$

Similarly,

$$[C_1\mathbf{P}C_2 \wedge C_2\mathbf{P}C_3 \wedge \ldots \wedge C_v\mathbf{P}C_{v+1}] \Longrightarrow C_1\mathbf{P}C_{v+1}$$

For S and X relations, a chain relation will not put any constraint on the relation between the contacts at the chain ends.

Chain rule II: $\forall C_1, C_2, C_3 \in C:\ [C_1\mathbf{P}C_2 \wedge C_2\mathbf{S}C_3] \Longrightarrow C_1\mathbf{S}C_3$

Chain rule III: $\forall C_1, C_2, C_3 \in C:\ [C_1\mathbf{P}C_2 \wedge C_2\mathbf{X}C_3] \Longrightarrow C_1\mathbf{S}C_3 \vee C_1\mathbf{X}C_3$

Chain rule IV: $\forall C_1, C_2, C_3 \in C:\ [C_1\mathbf{S}C_2 \wedge C_2\mathbf{X}C_3] \not\Rightarrow C_3\mathbf{P}C_1$

All other relations (except $C_3\mathbf{P}C_1$) are possible.

4.4 Contact topology and pulling experiments

In the previous sections, we discussed how pulling experiment can reveal the signature of the underlying topology of the system. As far as the topology is concerned the chemical nature of the protein and the contact free energies can be ignored. However in a real pulling experiment, energies will show up as the probability of the topologically allowed transitions. For simplicity we assume that the largest contribution to the free energy of a transition comes from the interaction energy of the contact that breaks during the transition. This is referred to as "contact energy" in the following discussion. How do topology and contact energy guide the unfolding of a protein whose inter-contacts relations can be described by the three relations: P, S and X (the PXS proteins)?

Pulling molecules from their termini is often performed using optical tweezers or AFM where the conformation of the protein is described by one measured property, its end-to-end length. In pulling experiments, two chains with identical contact length vector$L_c(p) = L\big(C_0(p)\big), p = 1{:}\,|C|$, and topology matrix will have a similar unfolding length sets (Fig 4.4b). Chains with identical topology matrix (T) that differ in their L_c will have similar selection rules for transitions but the final measured unfolding intermediate length sets will differ in general (Fig 4.4b i, iii).

For a chain with n contacts, assuming one-by-one dissociation of the contacts, we will have n unfolding steps. At each unfolding step (e.g. step i), a shortest path can be defined for the system and the contacts belonging to the shortest path will form the unfolding contact set $U(i)$: $\forall C, C' \in U(i), CSC' \vee C\mathbf{X}C'$. Contact topology matrix can be used alternatively to find U(i). At unfolding step i, the contact energy and loading rate will define the probability of a contact to open up. For any contact that does not belong to $U(i)$, this probability is zero.

Two chains with similar topology, contact length vector and energy will provide a similar unfolding length histogram when we perform pulling experiment. Two chains with similar total length, and similar unfolding intermediate length set and transition rules, will have to have the same topology.

4.5 Contact topology defines the unfolding of a complex protein

To test the applicability of the topological analysis presented above, we performed mechanical unfolding of a single protein. Firefly luciferase (550 aa), is a relatively large protein that is regarded as a model for complex proteins (> 100 residues). Such complex proteins may populate intermediates while folding or

unfolding. Here we perform mechanical unfolding and refolding experiment on luciferase for the first time and compared our experimental observation with the predictions of the topological analysis of the protein's crystal structure (Fig 4.7a).

From the crystal structure, we can construct a contact map based on inter-residue distance. A common choice is the distance between C^{α} of the residues. The contact map of luciferase is shown for inter residue distance less than 7Å. The residues that are in contact might be attracting or repelling each other. Interestingly the distribution of the contact energies shows bimodality for luciferase and many other proteins (Fig 4.7b,c). A subset of contacts can thus be distinguished with significantly high contact energies (the red color peak in Fig 4.7c). Expectedly many of these strong contacts are in close proximity of each other in the 2D contact map, as they for instance belong to two interacting β-strands. This allows one to further simplify the system by grouping the residue contacts into several contact sites (Fig 4.7e). We call these clusters of inter residue contacts simply as "contacts" in the rest of our discussion and perform the topological analysis on these contacts. The topology matrix and the most probable unfolding path are presented in Fig 4.8b. The latter is estimated using the topological relations between contacts and the contact energies.

When we pulled on a single luciferase with optical tweezers, we observe different sequences of protein lengths. During the experiment we always start with fully compact protein having a nearly zero end-to-end distance. After stretching, the protein unfolds and end-to-end distance increases and finally it reaches ~ 198 nm which is the contour length of the protein. In a simplest unfolding path, we observe a single transition from 0 to 198 nm. In other cases one or more unfolding intermediate lengths were visited. Putting all different paths together, we find that luciferase visited eight unfolding intermediate lengths (Fig 4.8.a) each with a certain probability (Fig 4.8.b black vertical drop lines). We also studied the length-length transitions and found that some transitions are rare (or possibly forbidden) while others occur with significant frequencies. When we compared the most probable path as predicted theoretically with the observed lengths during experiment, we find a very good match (Fig 4.8b red dashed lines). This agreement was robust when slightly different distance cutoff or different clustering of the contacts was employed. We also do not expect that different choices for aminoacid-aminoacid interaction energy affect the outcome if the energy order of the contacts remains the same with different choices.

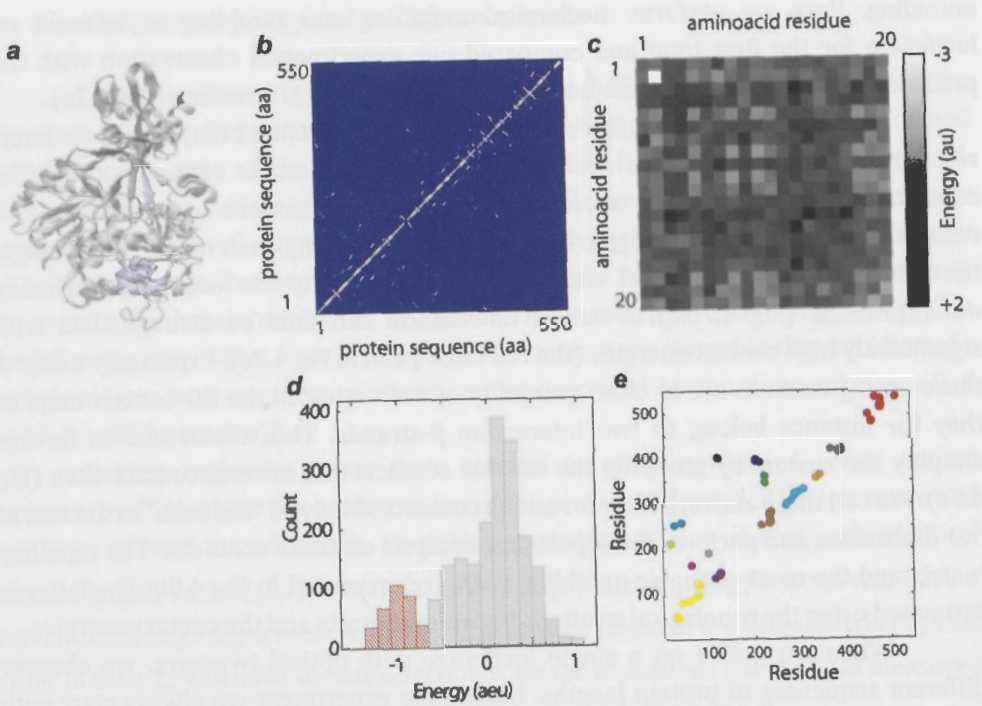


Figure 4.7: Coarse-graining protocol for contact analysis. Crystal structure of the protein (a) is used to construct the contact map (b) based on a distance threshold. An interaction energy matrix (e.g. Thomas Dill matrix) (c) is then used to filter out weak and repulsive contacts. Panel (d) shows that strong contacts can be readily identified due to a natural clustering of the contact energies in luciferase and many other proteins. The inset is the Thomas-Dill energy histogram of all the contacts within the protein. Two populations of attractive contacts (negative energies) can be recognized from the shape of the histogram. The peak on the left was used to construct the contact site map (e).

We note that the experimental data presented above is complex and such data must be interpreted with care. For example some lengths are transiently visited while others bear force on the order of 10s pN for seconds. Two different conformations might have similar length (degenerate length-states) and therefore might be left unresolved experimentally. It is hard to say if a transition is forbidden: experiments must be repeated many times until a rare event is likely to be seen.

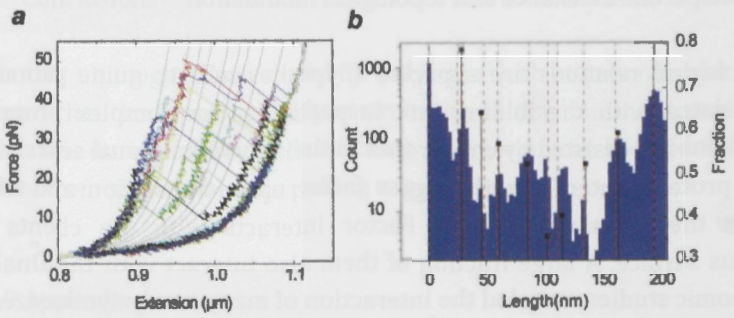


Figure 4.8: Unfolding pathway of luciferase. (a) Experimental results from single molecule unfolding with optical tweezers. The grey lines are the assigned worm-like chain models of the assigned intermediates. They were found by fitting a subset of experiments containing 40 unfolding traces and used for the whole set. (b) Histogram of the states were measured by two protocols: counting the states based on the assigned intermediates (black vertical drop lines) and subtracting the DNA worm-like chain from a selected subset of traces and counting and binning the protein lengths measured during the pulling cycle with 50 Hz sampling rate (dark blue histogram). The dashed red lines are the predicted unfolding intermediates based on the contact sites, their energy and topological relations.

4.6 Topological constraints on energy landscapes

Currently random energy landscape (REM) is commonly employed as a zeroth order approximation of free energylandscape for understanding the conformational dynamics of random biopolymers [119]. For proteins, the formation energies of non native interactions are often assumed to be distributed randomly, giving rise to a rough energy landscape. Because native contacts are often significantly stabilizing, there is a smooth overall slope in the energy landscape towards the native state of proteins (funnel like landscape). When a protein is under force, its free energy landscape will change: the unfolded conformation will have a reduced energy and the transition barriers will be suppressed. This leads to overall tilting of the landscape. The rate of transitions between states will therefore be altered under applied external force.

From our discussion on topology, we know that transitions are not only affected by the barrier energies, but also fundamental constrains will be introduced by topology. It is therefore conceivable that topology acts as a complementary notion to the notion of landscape in predicting the transitions.

4.7 Chaperone assistance and topological modulation

Topological relations are expected (hypothesized) to guide protein folding and be correlated with the folding rate in particular for complex proteins. Most complex proteins are assisted by chaperones in their conformational search. All newly synthesized proteins interact with trigger factor upon translation and likely after release from the ribosome. Trigger Factor interacts with the clients with its heterogeneous surface. A large fraction of them also interact with the DnaK system. Recent proteomic studies revealed the interaction of many newly synthesized nascent chains with DnaK.

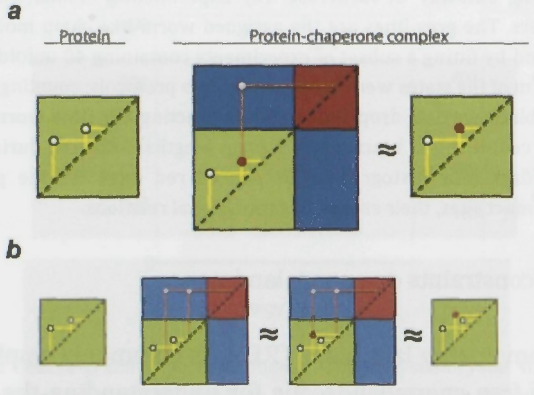


Figure 4.9: Contact map of bi-molecular complexes and the corresponding reduced forms (a) the effect of interaction can be reduced to an increase in the stability of an existing contact site. (b) The effect of chaperone can be reduced to a new contact in the protein changing its topology.

Here we show how an interacting protein (e.g. chaperone) can act as a topological modulator. A chaperone can interact with a client in three different ways: (a) simply enclosing it within its cavity without any tight interaction (b) Interacting with a large surface or segment of the chain and shielding it (c) interaction via making strong contact points. In the last case, at least two different scenarios are imaginable (Fig 4.9). In one case, shown in Fig 4.9.a, the contact with the chaperone increases the stability of an existing contact in the protein, leading to an effectively higher contact energy. In another case, chaperone interaction with multiple sites on the protein can be reduced to a protein model having additional contacts (Fig 4.9.b). The new contact sites will bring new topological constraints to the system influencing its unfolding and folding pathways.

4.8 Conclusions

Contact topology is introduced for polymers with intra-molecular interactions. It is proposed that the topology likely restrain the conformational search of proteins and nucleic acids during folding and unfolding. The notion can be used to interpret the complex data from single molecule pulling experiments.

4.9 Appendix

SI 4.1. β-β contact contact topology

Consider an protein consist of 4 beta strands and spacer loops (Fig S4.1a) and assume that in the folded state of the protein, pair wise interactions of beta strands are allowed to form beta sheets with two strands. Proteins tertiary fold can be partly described by its β-β contact map. For our example protein, a number of conformations are possible with different choice of interacting beta strand-pairs. Some of these conformations are schematically illustrated in Fig S4.1b & c. By looking at the contact map at different scales, we find a number of points: for example, the orientations of the strands give a clear signature in the contact map with parallel strands displaying lines parallel to the diagonal while anti-parallel strands display lines perpendicular to the diagonal. The second interesting feature is the relation between beta sheets (or simply contacts) as illustrated in the insets of Fig S4.1b & c. Similar relations can be distinguished for β-β contact as for inter-aminoacid contact in Fig 4.3.

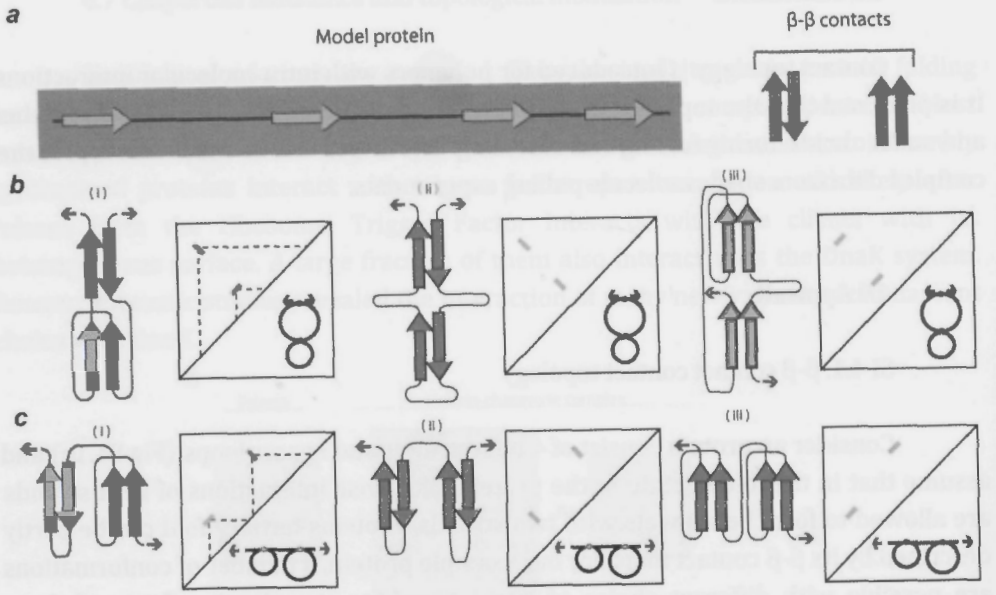


Figure S4.1: Arrangement of β-β contacts in a simple all β model protein. (a) structure of the protein with four beta strands. (b,c) A few tertiary conformations of the protein are depicted with their corresponding contact maps. Aminoacid contact maps are filtered to only contain β-β contacts. Contacts between beta strands will appear differently depending on their orientation. The insets display the topologies of the conformations. Conformations b and c belong to two different topological classes, P and S respectively.

CHAPTER 5

Chaperones Trigger factor and Heat-Shock Protein suppress entry into misfolded state

Off-pathway intermediates and kinetic traps can dramatically influence the final yield and folding rate of complex proteins. Despite being a commonly accepted notion, the literature on characterization of such intermediate states is scarce. Here we find an off-pathway intermediate in the folding pathway of model protein, luciferase which significantly reduces the yield of native protein. We show that chaperones Trigger factor and heat shock protein Hsp70 can help luciferase to survive the trapped state. This study provides a first single-molecule-level evidence for chaperone suppression of an off-pathway intermediate in the folding of a protein.

5.1 Introduction

The native states of structured proteins typically correspond to the structures with highest thermodynamic stability [120]. For proteins of small size (60-100 residues), fundamental folding mechanism involves the interaction of a relatively small number of residues to form a folding nucleus, about which the rest of the structure condenses rapidly [121]. Despite a high degree of disorder, these nuclei have the same overall topology as the native fold and will almost invariably be converted to their corresponding native states. If these key interactions are not formed, the protein cannot fold to a stable globular structure [122]. For larger proteins, the folding process may lead to misfolding intermediates. For proteins with more than about 100 residues, one or more intermediates are believed to be populated during the folding process. These intermediates may assist the protein to find its correct structure or they might act as traps that inhibit the folding process [123-125]. Regardless of the role of these intermediates in folding, the physical properties of the intermediates provide important insight into the folding of larger proteins [126].

In contrast to on-pathway folding intermediates for which the structural features have been extensively studied by a variety of techniques, including Chevron analysis (Fig 5.1.a), NMR spectroscopy and Φ-value analysis [127-133], the study of misfolded metastable states remains a challenge. Few examples exist where an off-pathway intermediate is described as being populated during a folding process [129, 134-136]. D1pPDZ is an example of a protein with a known off-pathway intermediate. The folding process of D1pPDZ involves a kinetic competition between folding to the native state and misfolding to a low-energy off-pathway intermediate. A kinetic partitioning is observed between two folding pathways, one leading to the more stable native state and the other to the less stable off-pathway intermediate [137]. The misfolding state, like the one observed in D1pPDZ (95 residues), is likely to be rare in small proteins, but it may become more common as the size of the protein increases and its topology becomes more complex.

To understand the intermediate folds (both on- and off- pathway ones) we studied Firefly luciferase (550 residues) with optical tweezers at the single molecule level for the first time (Fig 5.1.a). Firefly luciferase is a 61-kDa (550 aa) multi domain protein and the light-producing enzyme from the North American beetle *Photinus pyrali.* It serves as a model substrate for protein folding and aggregation studies. However its sensitivity to aggregation, made it difficult to address its folding pathway

in bulk assays. Due to the kinetic competition between an on-pathway reaction leading to the native state and detrimental pathways leading to irreversible

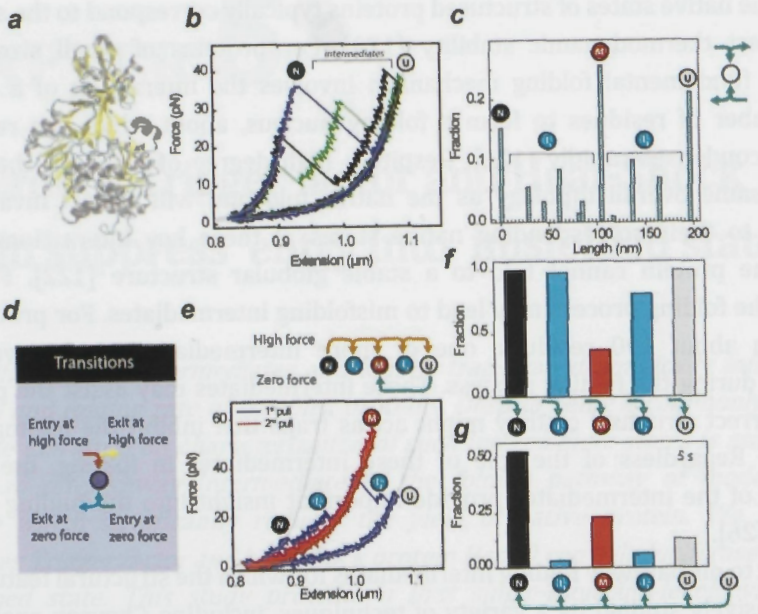


Figure 5.1: Unfolding and refolding of luciferase (a) Structure of Firefly luciferase (1LCI). (b) Unfolding from native state (N) to unfolded state (U). Force-extension curves of luciferase are recorded by optical tweezers. The unfolding occurs in single- or multi step. (c) Length statistics for states (N = 203) visited in the unfolding/refolding cycles. Three most populated states can be recognized: N, U and M. Other intermediate states are populated with low frequencies and are grouped into I_1 and I_2 categories. (d) Exit from and entry into a state may occur at high force or zero force. (e) Refolding from unfolded state to M-state. An example case is presented in which the protein refolded to M-state after being unfolded. The statistics presented in (c) is split for unfolding (in panel f) and refolding (in panel g).

aggregation, reactivation yields close to unity can only be achieved at low temperature and very low protein concentrations (Ruth Herbst et al., Biochemistry 37, 6586-6597, 1998). Chaperones are known to increase the folding yield of luciferase e.g. by suppressing aggregation both *in vivo* and *in vitro* [138, 139]. In the following we characterize folding pathway of luciferase and investigate the influence of chaperones on Luciferase folding process.

5.2 Single-molecule investigation of (un)folding intermediates of luciferase

Using optical tweezers, we induced the mechanical unfolding of luciferase by stretching it from its termini with forces up to 65 pN. The unfolded protein (U; contour length ~198 nm) was then relaxed to zero force and allowed to fold. Natively folded luciferase (N) occasionally unfolds in a single step with an unfolding force of about 40 pN. In most cases however, luciferase unfolding occurs through several unfolding intermediates (Fig 5.1.b, c), some visited transiently while others could withstand the applied force for up to seconds. When an unfolded protein was relaxed to zero force, a subsequent pull after 5s waiting time resulted in a number of states with different lengths (as identified by fitting worm-like chain model to the force-extension traces). By looking at the frequency of the states visited during many stretching/relaxation cycles (Fig 5.1c), we found a frequently populated medium-size state (M-state) at a protein length of ~ 105 nm (~ 258 aa fold size).

During stretching/relaxation cycles, the states can be visited during unfolding as well as refolding: in other words, "exit from" and "entry into" a state can both occur at high force and both occur at low force in these experiments (Fig 5.1d). Entry into M-state for example can occur via unfolding a protein state with shorter length than M-state; it can also occur via folding of a protein state with a longer length (Fig 5.1e).

When we looked into unfolding experiments that start with a natively folded luciferase (N) and end with fully unfolded protein (U), we find that the entry into M-state via unfolding happens with a much less probability than for any other state (Fig 5.1f). When unfolded protein was allowed to refold at zero force for 5s, the entry into M-state via refolding occurred with a higher probability than the entry into any other state except for the natively folded protein (Fig 5.1g). From these observations, we surmised that M-state could be a misfolded (trapped) state.

To test if the M-state is a kinetic trap, we monitored the time evolution of the folding process. First we fully unfolded the luciferase, and then relaxed the unfolded state to zero force. We then stretched the protein after 1 min waiting time. We found that almost half of the cases the protein had entered the M-state. Even after 5 min waiting time, the unfolded chain transited to and got trapped in this state in a significant fraction of the chains tested (Fig 5.2a). Second, we monitored the time evolution of the M-state. For this aim, we first generated the unfolded state via unfolding and then allowed the unfolded state to enter M-state via refolding at zero force. M-state was then characterized by stretching it and measuring its length. However we did not attempt to unfold it with force. Instead the protein was relaxed to zero force. When we stretched the protein after 5 s waiting time, in most cases, we

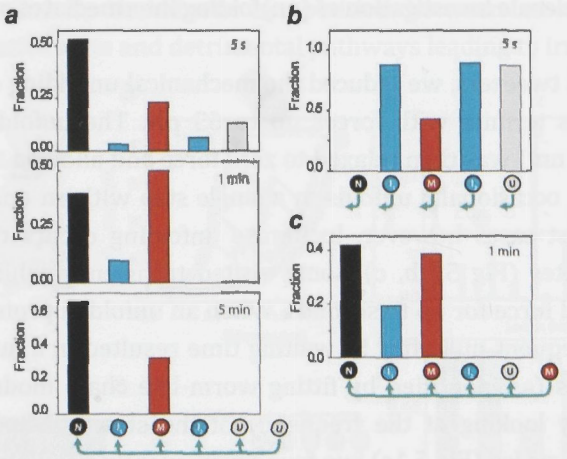


Figure 5.2: M-state is a kinetic trap. (a) Entry into M-state at zero force. luciferase has a densely populated trapped state in its folding pathway as measured in single molecule level. Transition probability of unfolded luciferase to different folded states is quantified as a function of time (b) Exit from M-state is relatively infrequent at zero force i.e. M-state is kinetically stable. (c) Exit from M-state after 1 minute waiting time at zero force. The protein is found to maintane its state in ~ 40% iof the cases after this long waiting time. Comparing panels (b) and (c), it can be concluded that most exit events happen already after 5 seconds.

found it again in the M-state (Fig 5.2b). We performed similar experiment on other intermediate states and found that the exit probability of M-state is significantly less than the exit probability of other states at zero force. We then performed a similar experiment on the M-state with 1 min waiting time and found that the exit probability after 1 min is only slightly higher than the exit probability after 5 s waiting time. Almost half of the proteins trapped in the M-state, did not exit it after a min waiting in the absence of mechanical force. This indicates that M-state is a kinetic trap.

To estimate the mechanical stability of the M-state of luciferase, we performed unfolding experiment on the M-state generated via refolding. We found that unfolding of M-state occurs at forces much higher (p-value < 0.0001) than the unfolding force of other intermediates (Fig 5.3). The mechanical stability of the M-state is comparable to the mechanical stability of the natively folded luciferase.

Our data indicate that luciferase has an intermediate trapped state in its folding pathway, which significantly slows down its folding even at the single molecule level when the possibility for inter-molecular aggregation is eliminated. The state shows a dramatic mechanical and kinetic stability.

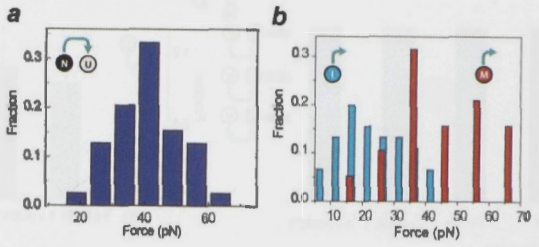


Figure 5.3: (a) Unfolding force histogram of nativeky folded luciferase. (b) Mechanical satability of M-state. Exit from M-state under force requires a higher force than exit from other intermediates.

5.3 Chaperones suppress misfolding

We then asked if the presence of Trigger factor and the DnaK system help the system to skip the trapped state (Fig 5.4). Performing similar experiments in the presence of Trigger factor (1 μM) led to a significantly reduced probability of entry into this state in the absence of force. For the protein states with this length, we observed a reduced residence time (increased exit probability at zero force) indicating that the state is destabilized. Additionally the state show a reduced mechanical stability (~ 10 pN reduction in required force) when it was generated via unfolding. Because the full conformational space of the protein is projected into one (or two) parameter, we cannot rule out the possibility of a degenerate length state. Trigger factor is expected to delay folding even in the absence of a misfolding state. We note that in our experiment, a large number chaperones are available to interact with one single luciferase molecule making the folding process relatively slow.

We observed similar influence for DnaKJ system particularly when we increased the ratio between DnaK and DnaJ. A reduced probability of entry into M-state together with an increased exit probability hints at a reduced thermodynamic stability of the M-state. The DnaK system, however, trapped the system in otherwise transient intermediates which is understandable considering the heat shock activity of the system (see chapter 6).

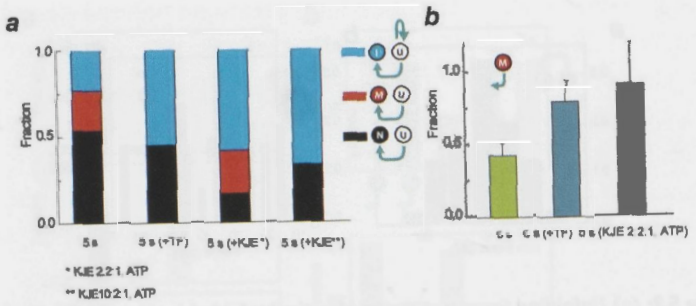


Figure 5.4: Exit from and entry into M-state under the influence of chaperones DnaK and Trigger factor (TF). KJE stands for DnaK, DnaJ, GrpE mixture. The concentrations of TF and DnaK were 1uM and 100 nM respectively. The co-chaperones were added with the specified molar ratios together with 1mM ATP. (a) Entry into M-state at zero force from U state is suppressed in the presence of chaperones (b) Exit from M-state is promoted in the presence of chaperones.

5.4 Conclusion

Model protein luciferase often fold to a non-native conformation which is mechanically and kinetically stable. We characterized this conformation by measuring its end-to-end length, unfolding force, and refolding probability. We found that chaperones TF and DnaK suppress entry into this non-native trapped state.

CHAPTER 6

Reshaping of a protein folding landscape by the chaperone Trigger Factor

Protein folding is often described as a search process, in which polypeptides explore different conformations to find their native structure. Molecular chaperones are known to improve folding yields by suppressing aggregation between polypeptides before this conformational search starts[140, 141], as well as rescue misfolds after it ends[140, 142]. While chaperones have long been speculated to also affect the conformational search itself, by reshaping the underlying folding landscape along the folding trajectory[126, 143], direct experimental evidence has so far been scarce. In Escherichia coli, the general chaperone Trigger Factor[144-146] (TF) could play such a role. TF has been shown to interact with nascent chains at the ribosome[147-149], delay their folding[8, 139, 150, 151],remain bound when the polypeptides are released into the cytosol[139], and to interact with fully folded proteins prior to their assembly into larger complexes[152]. To investigate the effect of TF on the conformational search, we measured the folding dynamics of Maltose Binding Protein (MBP) at the single-molecule level using optical tweezers. We show Trigger Factor binds folded structures smaller than one MBP domain, which are then stable for seconds and ultimately convert to the native state. Moreover, Trigger Factor binding stimulates native folding by suppressing misfolding interactions between domains in constructs of repeated MBP domains. These results indicate that Trigger Factor induces or deepens valleys in the energy landscape, which directs folding pathways along intermediates towards the native state and preempts competing misfolding trajectories. As Trigger Factor interacts with most newly-synthesized proteins in Escherichia coli, we expect these findings to be of general importance in understanding protein folding pathways.

6.1 Introduction

We used optical tweezers[153] to mechanically unfold single MBP molecules and subsequently allow them to refold under controlled conditions in the presence and absence of TF. This technique enables high resolution quantification of molecular lengths and forces in real time, which can reveal folding transitions between transiently stable protein structures as well as their resistance against forced unfolding. Optical tweezers and Atomic Force Microscopy have been used to characterize the folding dynamics of proteins in isolation[154] and of proteins that interact with ligands[155] and metal ions[156], but rarely to address how folding is affected by chaperones. We have previously shown that the chaperone SecB delays folding initiation by maintaining proteins in the unfolded state, rather than influencing the conformational dynamics along the folding pathway[18]. Here we investigate how TF changes the folding dynamics of two proteins: the single-domain MBP, which folds efficiently in isolation, and of an MBP repeat construct, which tends to misfold. This approach allows us to explore how TF affects the folding transitions within the single-domain MBP as well as misfolding between repeated MBP domains.

In the absence of any chaperone, stretching MBP molecules with optical tweezers (Fig 6.1a) displayed a simple unfolding pattern involving two transitions[18] (Fig 6.1b). At around 10 pN the measured protein length suddenly increased by 28 nm, indicating the unfolding of a polypeptide segment of that length. The remaining intermediate structure unfolded completely at around 22 pN, yielding the full MBP length of 120 nm. After relaxation to zero force, the polypeptide chain was allowed to refold during 5 s. Subsequent stretching typically showed either a fully folded or fully unfolded state, as evidenced by a curve similar to the first or a curve lacking unfolding features (refolding rate[18] ~ 0.7 s^{-1}), in line with a two-state folding process. Analysis of many stretching curves showed that a small fraction displayed additional protein states during unfolding (9% of observed states, N=215). In the distribution of protein lengths, these states are seen as small but significant peaks between 40 and 110 nm, which are consistent with previous AFM unfolding data[157] (Fig 6.1d, stars). The MBP native state thus contains a number of specific sub-structures that can be transiently stable during unfolding in the absence of chaperone.

6.2 Reshaping the landscape of a single domain protein

In the presence of TF (1 μM, method S6.2) the first stretching curves were unchanged, but subsequent stretching and relaxation data were quite different (Fig

6.1c). The curves now frequently displayed the intermediate protein lengths that were rare without chaperone, and consequently the small peaks in the length distribution were substantially higher (Fig 6.1d, stars). The peaks were distinct, suggesting that in different experiments the protein chains repeatedly adopted the same or similar structures. We note that the broadness of some peaks, and the small unfolding transitions observed occasionally, did indicate small structural variations. The correspondence in peak positions for experiments with or without TF (Fig 6.1d, stars) suggested that the structures were partial MBP folds defined predominantly by native interactions, rather than by misfolding interactions or by the existing interactions between MBP and TF.

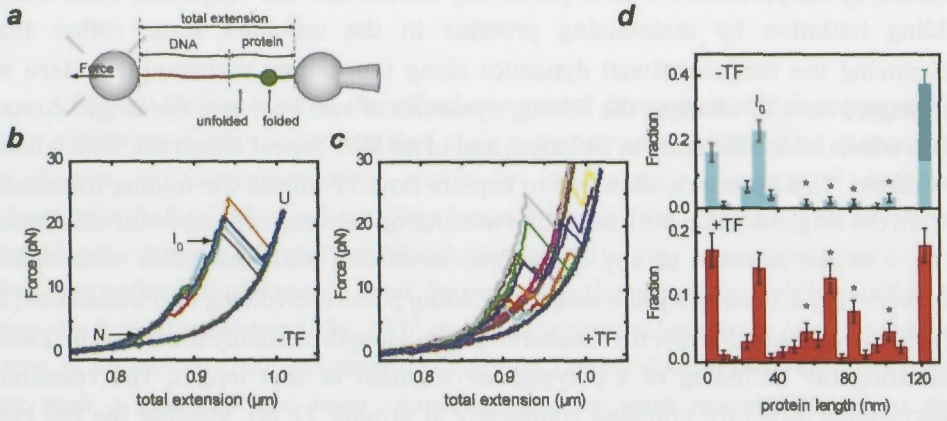


Figure 6.1: Trigger Factor promotes partial protein folds. (a) Schematic diagram of the set-up. Maltose Binding Protein (MBP) is tethered between two beads: one is held on a position controlled micropipette; the other is held by an optical trap that allows force detection. At the N-terminus MBP is attached to a bead by a 920 nm-long DNA tether via streptavidin-biotin linkages. The DNA is attached to the bead using antibody-digoxigenin linkages. At the C-terminus, MBP is attached to a bead via an antibody-myc-tag connection. (b) Stretching-relaxation experiments in the absence of Trigger Factor. The force-extension data shows one stable intermediate state I_0 in between the native state N and the unfolded state U, two unfolding transitions at about 10 pN and 22 pN, and refolding at zero force. (c) Stretching-relaxation experiments in the presence of Trigger Factor (1 μM) indicating additional stable states of intermediate protein length. (d) Protein length distribution of stable states with and without Trigger Factor, as derived from the force-extension data. Lengths are determined by fitting to the Worm-like chain model[158]. Error bars indicate the standard deviation. The native, I_0, and fully extended states correspond to lengths of about 0 nm, 28 nm, and 120 nm. The number of times a state with a certain length is visited is indicated as a fraction of the total number of visits to all states.

To probe the stability of complexes between TF and partial MBP folds we quantified the applied force at which they unfold. We found that the presence of TF

increased the required force by almost two-fold (from 15 to 27 pN on average, Fig 6.2a). Some unfolding forces exceeded even 40 pN; significantly higher than the 22±5 pN force at which the native state unfolds without chaperone. Note that the partial MBP folds are flanked by two (or one) non-folded polypeptide segments. TF could provide stabilization against applied force by binding to the folded MBP region, or by simultaneously binding both flanking polypeptides, just where they emerge from the folded structure. TF proteins that bind non-folded polypeptide away from the folded structure cannot increase the folding force. If the MBP chain would not be folded but sequestered on TF without tertiary structure, disruption of the complex would likely be gradual and at low-force as observed for SecB[18]. The TF-stabilized partial folds could unfold in different ways. The applied force could deform the complex as a whole and hence trigger MBP unfolding. Alternatively, (partial) TF dissociation may occur spontaneously, after which the unbound MBP structure unfolds by force. The latter scenario would be consistent with reported TF-substrate lifetimes[159-162], which are between 0.1 and 100 s, and thus of similar order as the stretching experiments that last several seconds.

The protection against forced unfolding suggests TF may also stabilize partial folds in the absence of force, and thus affect folding to the native state. To address this issue we analyzed folding transitions during the 5 s waiting periods at zero force in the presence of TF, by comparing protein lengths before and after this period (see Fig 6.2b-e). Sometimes, the protein length increased (14%, N=33) or remained constant (22%; Fig 6.2d, experiment 2, denoted as 2d-2), indicating the folded structure had unfolded or remained stable respectively. Most often however (64%) the protein length decreased, indicating that a larger part of the protein chain had become folded. We observed the full range of these refolding transitions: from unfolded to an intermediate length (Fig 6.2e-2), from one intermediate length to a shorter one (Fig 6.2c-2) and onwards to the native length (Fig 6.2c-4 and 2b-2). Given the 5 s waiting period, the average lifetime of TF-promoted partial folds is thus in the order of seconds[163]. These visits to partial folds decreased the overall folding rate, consistent with bulk studies[139, 151, 159]. The partially folded states were not observed during refolding in the absence of TF, suggesting they were either visited too briefly to be detected, or did not form at all. In the absence of applied force, TF thus transiently stabilized partial folds, resulting in a slow step-wise refolding *via* these partial folds. To test the generality of these central findings we performed experiments on a different substrate, namely the protein luciferase, which showed similar results.

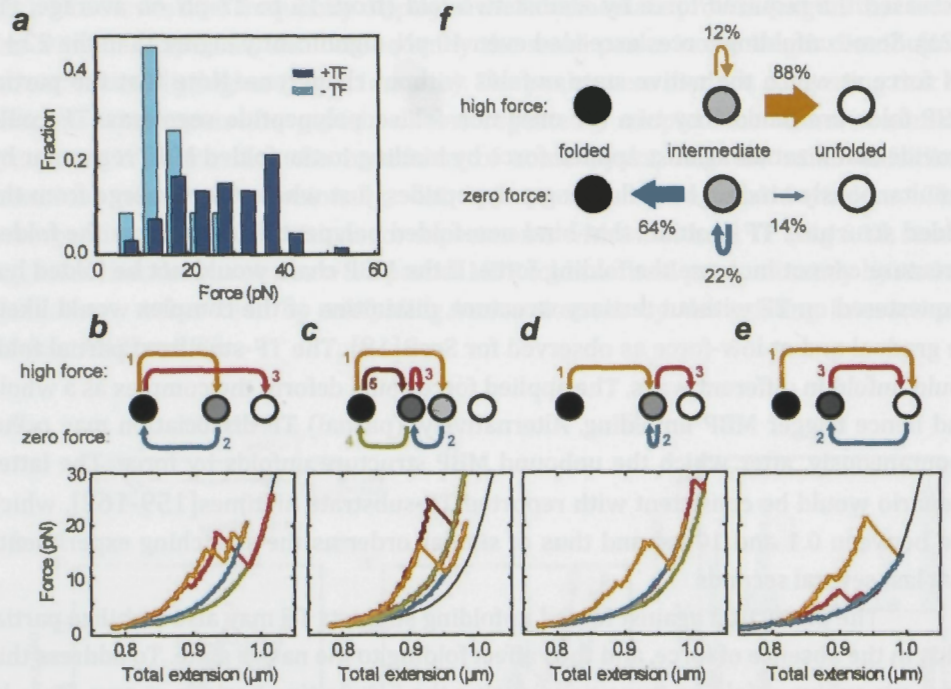


Figure 6.2: Trigger Factor stabilizes partial folds and mediates folding transitions. (a) Distribution of forces maximally sustained by partial MBP folds (protein lengths between 40 and 110 nm). In the presence (dark blue) of Trigger Factor (TF) the forces are higher than in its absence (light blue), indicating protection against forced unfolding. (b-e) Transitions between folded states with TF. In diagrams, circles represent protein states, with white the fully extended state and black the most compact state. Arrows represent transitions between these states during stretching (high force) or relaxation followed by waiting at zero force for 5 s. (zero force). Numbers indicate the order of experiments. At high force, proteins typically unfold but can also remain intact (panel c, experiment 3, denoted as c-3). At zero force the data shows the full spectrum of folding transitions: from unfolded to intermediate (e-2), from intermediate to another intermediate (c-2), and from intermediate to native (c-4 and b-2), as evidenced by subsequent characteristic unfolding pattern (b-3). The protein length can also remain unchanged (d-2). For clarity, the stretching-relaxation behaviour of fully unfolded protein with DNA tether is indicated by a black line. (f) Average probabilities of decreasing, maintaining, or increasing the protein length in the stretching (high force) and relaxation (zero force) experiments with TF, starting from any intermediate MBP fold (panel b-e).

Two main lines of evidence indicate that TF-mediated folding leads to the native structure. First, we found MBP function is restored after chemical denaturing and refolding in the presence of TF in a bulk assay, as tested by amylose binding (Fig 6.3a). Second, after unfolding by force and relaxation, subsequent stretching experiments displaying the characteristic two-step native unfolding pattern (Fig 6.1b)

indicated that MBP had refolded to the native state (Fig 6.2b-3). Once MBP is folded natively, we find that it binds TF only weakly or not at all. This can be seen in the optical tweezers experiments, where the first stretching curve showed the same unfolding pattern and unfolding force; with or without TF. A bulk column retardation assay further confirmed this lack of interaction (Fig 6.3c). As observed previously (Fig 6.2a), partial folds do interact with TF. To obtain further evidence that TF interacts with folding intermediates in bulk experiments, we purified MBP constructs that are truncated and thus cannot form the full MBP native state but can form partially folded structures (Fig 6.3b). In contrast with the full-length MBP, we found TF indeed binds these truncates in column retardation assays (Fig 6.3c). Binding of TF to an MBP truncate was further demonstrated in thermal unfolding experiments using FTIR, which showed that TF significantly suppressed heat-induced protein aggregation (Fig 6.3d). Overall, the data indicated that TF remains associated with MBP along the folding pathway until it reaches the native state.

6.3 Reshaping the landscape of a multi-domain protein

We hypothesized that TF association could be particularly beneficial for complex multi-domain proteins, which can suffer from misfolding interactions between domains. To test this hypothesis, we performed stretching experiments on a construct composed of four MBP repeats (MBP_4). When unfolded MBP_4 was relaxed in the absence of chaperone, subsequent stretching typically showed it had become similarly compact as native MBP_4 but then failed to unfold[18] (Fig 6.5a). This suggested that tight misfolding interactions had formed between domains. When unfolding did occur, the length increases were typically more than one full-length domain (370 residues). This large step-size again indicated interactions between domains, though less tight. Occasionally we observed signatures of native-like folds involving just one domain, with unfolding forces below 30 pN and length changes smaller than one MBP repeat. Overall, we found that the major part of the protein chain (73% on average, N=41 stretching curves) was involved in tight (35%) or weak (38%) misfolding interactions, and the rest (27%) in native interactions (Fig 6.5b). These data indicate a strong tendency for MBP_4 to misfold.

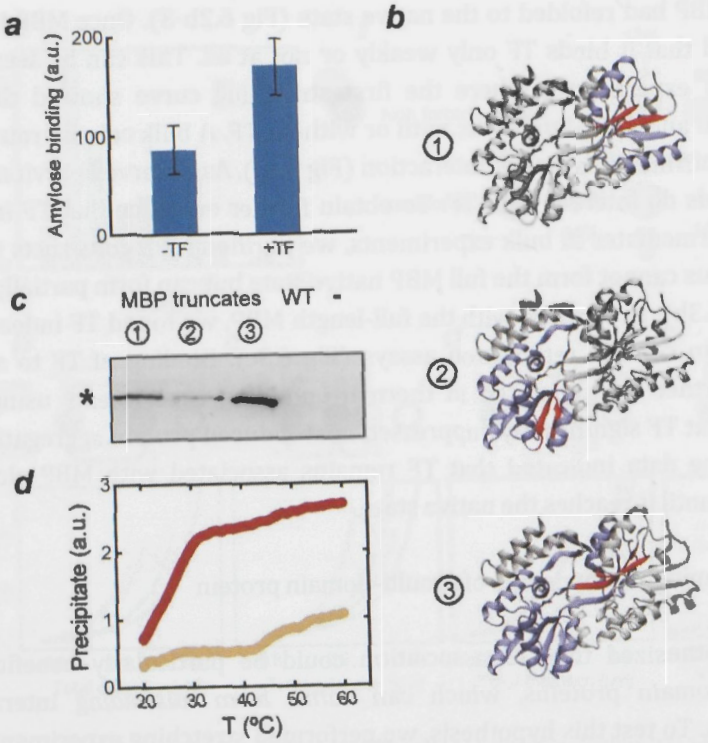


Figure 6.3: Trigger Factor binds partially folded states, not the native state (a) Bulk refolding and functional assay of MBP. Indicated is the amount of protein specifically bound to amylose, a mimic of the natural MBP substrate maltose, after chemical denaturation and subsequent refolding without and with Trigger Factor (TF). See Method S6.3. (b) *Constructed* MBP truncates in purple and red, with red indicating their C- and N-termini. See Method S6.1, Fig 6.4a, and Fig 6.4b. (c) TF binding to MBP truncates. All three MBP truncates bind TF, natively folded MBP (WT) does not. See Methods S4 and S5. (d) Thermal aggregation assay. Displayed is an FTIR signal that indicates the amount of precipitated MBP truncate nr. 3 (panel b) as a function of temperature, in the absence (red) and presence (orange) of TF. See Method S6.6.

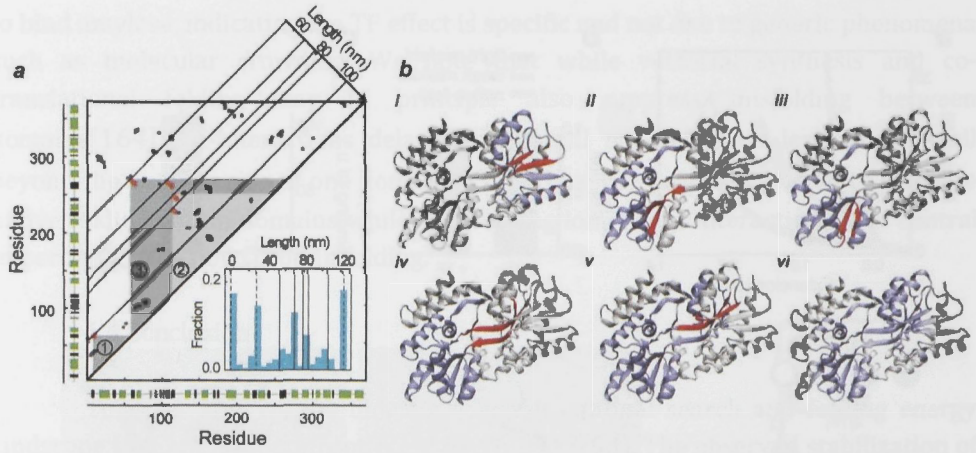


Figure 6.4: (a) Determination of MBP truncates that can fold only partially. Points are residue contact map of MBP (residue pairs within 7 Å of the crystal structure, see method S6.1). Residue contacts within 20 positions in sequence (along the diagonal) are omitted for clarity. Red points: Five most stable contact regions, as determined by adding energies of neighbouring contacts (method S6.1). For each contact region, one can consider the polypeptide segment in between the residues that form the contact region. The black lines indicate the lengths of these polypeptide segments. Each of these five polypeptide segments forms a sub-structure within the larger MBP crystal structure (Panel b). Three of these five polypeptide segments (truncates) were purified (indicated by the three shaded regions (method S6.4). Residue positions are for mature MBP lacking the signal sequence. Beta-sheet (black) and alpha-helices (green) are shown along axis. Inset: measured length distribution of folds (Fig 6.1d). Protein lengths of the five polypeptide segments as determined by the contact map analysis are indicated by the drawn black lines. Dashed lines: protein lengths for native, I_0, and fully unfolded states. (b) MBP sub-structures. In red are indicated the five MBP contact regions as determined by the contact map analysis (method S6.1). Purple indicates the polypeptide segment in between the residues forming the contact regions. Each panel shows one such sub-structure: *i*) residues 7-62 and calculated length, L ~ 105 nm *ii*) residues 112-228, L ~ 84 nm *iii*) residues 114-247, L ~ 78 nm *iv*) residues 104-266, L ~ 69 nm *v*) residues 57-267, L ~ 53 nm. *vi*) residues 1-279, L ~ 92 nm. Residue numbers are for mature MBP lacking the signal sequence.

In the presence of TF, the first stretching curve on fresh MBP_4 was similar to curves without TF. However, subsequent stretching of the same construct after relaxation showed marked differences (Fig 6.5a). Strikingly, with TF present, although weak misfolding interactions were still substantial (35%), the tight misfolds that failed to unfold were only rarely observed (less than 4%). The majority of the protein chain was now involved in native unfolding signatures (61%, N=55, Fig 6.5b). As a result, over 10 unfolding-refolding cycles on one construct could sometimes be followed, contrasting with maximum of 3 cycles due to tight misfolding in the absence

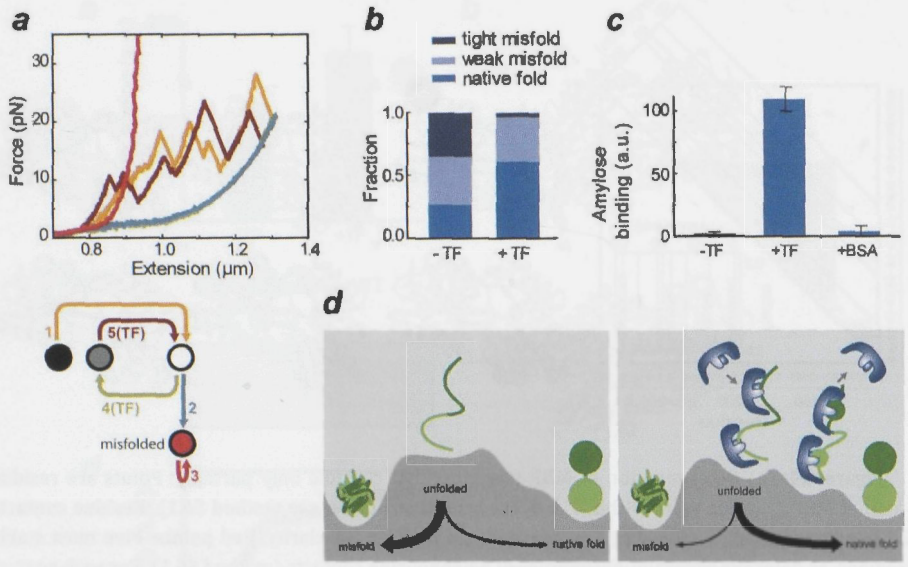


Figure 6.5: Trigger Factor suppresses misfolding between domains. (a) Force-extension curves for a four MBP repeat construct, with and without Trigger Factor (TF, 1 μM). See figure 6.2 caption for bottom diagram description. Without TF: after unfolding (experiment 1, orange), relaxation (2, blue), waiting at zero force for 5 seconds, stretching (3, bright red) shows no unfolding, indicating tight misfolding interactions between domains. With TF: after relaxation (4, green), waiting at zero force, stretching (5, dark red) showed extension changes characteristic of native interactions (e.g. at ~1.23 μm), and weak non-native interactions (at ~1.10 μm), but no tight ones. (b) Misfolded and folded fractions with and without Trigger Factor. Indicated are the fractions of the polypeptide length involved in native, weak, and tight misfolding interactions (see main text). Data is derived from multiple force-distance curves (N=41 without, N=55 with Trigger Factor) (c) Bulk refolding and functional assay of the multi-domain protein. Indicated is the amount of protein specifically bound to amylose, after chemical denaturation and subsequent refolding without and with Trigger Factor, and with the protein BSA (see Method S6.3, Fig 6.4b). (d) Cartoon of folding landscape reshaping by TF. Left: without chaperone, interactions between domains leads to efficient misfolding in multi-domain protein. Right: TF binding induces or deepens energy valleys, resulting in barriers to misfolding between domains in particular, and hence promoting native folding. This observed ability to distinguish native from non-native folding may be understood in kinetic terms: for residues within a domain to interact, partial unbinding of one TF may be sufficient, leading to a minor barrier. For more distant residues in different domains to interact, two TF monomers may need to unbind, leading to more significant kinetic barriers.

of TF. We further probed native refolding of MBP_4 in a bulk assay, which showed that in the absence of chaperone MBP_4 failed to refold to a functional state, while the presence of TF restored native folding, as shown by a more than 10-fold increase in amylose binding ability (Fig 6.5c, method S6.3). A control experiment where TF was exchanged with Bovine serum albumin (BSA) again led to misfolding and an inability

to bind amylose, indicating the TF effect is specific and not due to generic phenomena such as molecular crowding. We note that while vectorial synthesis and co-translational folding may in principle also suppress misfolding between domains[164], TF-interactions delay folding until nascent chain lengths are well beyond the average size of one domain[151]. Thus, the propensity of TF to promote native folding within domains while suppressing long range interactions is of central importance to de novo protein folding.

6.4 Conclusion

Together, our results indicate a conformational search and folding energy landscape that are profoundly affected by TF (Fig 6.5d). The observed stabilization of unfolded and partially folded states shows that TF induces energy valleys or deepens existing ones, which directs folding pathways along them to the lower-energy native state. TF is known to have a highly flexible structure and heterogeneous surface[144], which could allow TF to remain associated with the substrate as it undergoes sequential reorganizations during folding. Interestingly, the data suggest a simple mechanism by which TF resolves the generic chaperone challenge of distinguishing native from misfolding contacts: by interacting with a region of the polypeptide, TF effectively promotes local native contacts by shielding this region from misfolding interactions with distant regions, hence allowing domains to fold quasi-independently from the rest. Given the minimal requirements for this mechanism, it is intriguing to consider whether it is employed by other chaperone systems as well.

In summary, our study has shown that chaperones can directly influence the conformational search of polypeptides toward their native state. By reshaping the folding energy landscape, folding trajectories are channeled towards native folds and away from misfolds. The broad spectrum of newly synthesized protein chains that interact with TF in *E. coli* suggests that reshaping of the folding landscape by TF is of general importance in understanding the protein folding problem.

6.5 Appendix

Method S6.1: Contact map with high energy contact regions

The 3D conformation of MBP (PDB: 1jw4) can be represented in a symmetrical, square, boolean adjacency matrix (A_{ij}) of pairwise, inter-residue contacts. A_{ij} was constructed using a threshold of 7Å in spatial distance between the residue's C_α atom. Nearest neighbour contacts, with a distance $d(A_{ij}, A_{kl}) = ((i-k)^2+(j-l)^2)^{1/2}$ of less than 2, were considered to form a single contact region. The shortest range interactions, between residues spaced less than 20 residues along the polypeptide, were not considered. The stability of the contact regions was crudely estimated by summing the empirical Thomas-Dill energies [165] of all the residue-residue contacts within that region. The residues involved in the 5 most stable contact regions are indicated in red in Fig 6.4b. These contact residues define the ends of a sub-structure, as indicated in purple (Fig 6.4b). 4 of the 5 sub-structures are nested, such that for instance one large sub-structure (Fig 6.4b, panel v) encompasses 3 other sub-structures (Fig 6.4b, panels ii, iii, and iv). One of the stable sub-structures is independent of the others (Fig 6.4b, panel i). One may consider that two independent sub-structures are present simultaneously, and either interact and thus forming one larger fold (e.g. Fig 6.4b panels i and v, which together form the core state), or two distinct folds.

Method S6.2: Trigger Factor expression and purification

Trigger Factor with a Ulp1-cleavable N-terminal His6-SUMO tag was expressed from vector pCA528-tig in BL21 DE3Δ*tig*::Kan by induction with 1 mM isopropyl-1-thio-β-D-galactopyranoside at 30 °C. The protein was purified using Ni-IDA matrix (Protino; Macherey-Nagel, Düren, Germany); eluted material was supplemented with His6-Ulp1 protease, dialyzed overnight, and contaminants were removed by incubation with Ni-IDA matrix and subsequent anion-exchange chromatography (Resource Q, GE Healthcare).

Method S6.3: Bulk MBP and MBP_4 refolding assay

2.5 µM MBP_4 respective 10 µM MBP was denatured by incubation in RF buffer (50 mM HEPES-KOH, pH7.5, 100 mM KCl, 5 mM $MgCl_2$) containing 8 M urea at 37°C temperature for 12 hours and subsequently for 5 min at 60 °C. Samples were diluted 1:10 in RF buffer in the absence or presence of Trigger Factor (2 µM) in a final volume of 50 µl and incubated for 15 min at 30°C. 40 µl amylose magnetic beads (NEB) equilibrated in RF buffer was added and samples were incubated at 4 °C for 15 min.

Resin was washed four times in 1 ml RF buffer and bound MBP or MBP_4 was eluted by resuspending the resin in 40 µl RF buffer containing 20 mM maltose followed by centrifugation and collection of the supernatant. Samples were subjected to SDS-PAGE followed by silver staining. Protein bands corresponding to the MBP proteins were quantified using Image Gauge (FUJIFILM Science Lab).

Method S6.4: Expression and purification of N-terminally biotinylated MBP truncates

N-terminally biotinylated MBP fragments were expressed as hybrid proteins consisting of an Ulp1-cleavable N-terminal His_{10}-SUMO tag followed by a TEV protease-cleavage site, an AviTag (Avidity, LCC, Aurora, Colorado, USA), which is biotinylated *in vivo*, and the respective MBP fragment. Proteins were expressed in *E. coli* BL21 cells harboring pBirAcm, an IPTG inducible plasmid with an extra copy of the *birA* gene encoding biotin ligase (Avidity, LCC, Aurora, Colorado, USA) in LB medium supplemented with 20 mg/l Biotin, 20 mg/l Kanamycin, 10 mg/l Chloramphenicol, 0,2% glucose. Cells were lysed in denaturing buffer D containing 20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 8M urea, the lysate was incubated at RT for 1 hour, cleared from cell debris by centrifugation at 35.000 g for 30 min and incubated with Ni-IDA matrix (Protino; Macherey-Nagel, Düren, Germany) for 30 min. The matrix was washed extensively with buffer D followed by several washing steps in refolding buffer R (20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 1M urea) and finally buffer MBP (20 mM Tris-HCl pH 7.5, 0.2 M NaCl). Hybrid proteins were eluted in buffer MBP containing 500 mM imidazole. The eluate was supplemented with His6-Ulp1 protease, dialyzed overnight in 20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 10% glycerol, and remaining contaminants (His6-Ulp1 protease and His_{10}-SUMO) were removed by incubation with Ni-IDA matrix.

Method S6.5: Interaction studies of TF with MBP and MBP truncates

Saturating amounts of N-terminally biotinylated MBP or MBP truncates were immobilized by incubation with 150 µl Streptactin Sepharose (IBA GmBH, Göttingen, Germany) for 15 min at 4°C in MBP buffer (20 mM Tris-HCl pH 7.5, 0.2 M NaCl). Unbound protein was removed by repetitive washing steps using MBP buffer. TF interaction with immobilized MBP or MBP truncates was studied by a one-hour incubation with purified TF (5µM) at room temperature followed by 5 times washing for 5 min in MBP buffer supplemented with 0.1% NP40. Bound TF was eluted by boiling in SDS sample buffer (125 mM Tris-HCl pH 6.8, 3% SDS, 3% beta-mercaptoethanol, 0.01% bromophenol blue, 10% glycerol) for 5 min. Detection of TF

was done by standard western blotting procedure using TF-specific polyclonal antibodies from rabbit (lab collection).

Method S6.6: FTIR measurements

An MBP truncate (residues 57-267, see Fig 6.4b panel v) was purified as described above except that hybrid protein was directly digested while bound to the IDA matrix by over night incubation with the His6-Ulp1 protease in buffer MBP (20 mM Tris-HCl pH 7.5, 0.2 M NaCl). The eluate containing the biotinylated MBP truncate free of the SUMO-tag and the protease (both were His-tagged and remained bound to the affinity matrix) was concentrated using Vivaspin concentrators (Sartorius AG, Göttingen, Germany) to a concentration of 2 mg/ml and dialyzed overnight at 4°C in 20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 10% glycerol.

For FTIR measurements, purified TF and the MBP truncate were dialyzed in separate dialysis bags in the same buffer container to exclude any buffer differences. After dialysis, concentrations of TF and MBP fragment were determined using the Biorad Protein Assay and FTIR measurements were performed at 1:1 stoichiometry at protein concentrations of approximately 30 µM.

FTIR spectra were taken from 20 to 60 °C using Bruker Tensor27 instrumentation (Bruker GmBH., Karlsruhe, Germany) equipped with BioATRCell II. For each specta, 60 scans were conducted over a range from 3.000 to 1.000 cm^{-1}. Between each series of scans, temperature was raised by 1 °C followed by an equilibration time of 2 min. Spectra were corrected by subtraction of the background water spectrum, compensation for water vapor, and analysed by determining the area under the amide I band.

CHAPTER 7

Molecular mechanism of the DnaK chaperone system: insights from single molecule measurements

The DnaK system is the main heat-shock and ATP-depletion stress response system in E. coli. Previous studies have shown that DnaK and its co-chaperone DnaJ interact with unfolded proteins, protecting them from aggregation. Evidence exists for interactions of partially folded proteins with DnaK, but the binding sites are believed to be the unfolded stretches within the partially folded client. We show --using single-molecule pulling experiments-- that DnaK binds folded structures in partially folded proteins and protects them against unfolding. We find that DnaK requires nucleotide binding for this function, and its interaction with the folded structures involves the substrate binding groove known to be the interaction site for the unfolded stretches. Our findings allow us to propose a new model for the function of the DnaK system. The observed preservation of the protein folds against unfolding is beneficial for cells under external stresses because refolding of a fully unfolded protein is more difficult that refolding of a partially unfolded state with intact native tertiary elements.

7.1 Introduction

The70-kDa heat shock protein (Hsp70s, also called DnaK) is an ATP driven chaperone known to influence the folding of a wide range of proteins and to be a central actor in a number of cellular stress responses. It is over-expressed to 12-20% of the total cellular proteins in response to ATP depletion stress (hypoxia, metabolic stress) and thermal stress (heat and cold). DnaK also performs critical chaperone functions in the absence of environmental stress and has a significant basal expression level which amounts to ~ 2% of the total proteins. Depletion of DnaK in *Escherichia coli* mutants lacking the chaperone Trigger Factor leads to substantial aggregation of cytoplasmic proteins [8]. Co-immuno-precipitation showed a range of DnaK clients with varying association timescales (O(sec)-O(10 min)), some appeared to be full length proteins in this analysis and some appeared as newly translated nascent chains. The fraction of the newly synthesized proteins that interact transiently with DnaK is estimated to be 26-39% [166].

Exposure of some tertiary folded conformers of proteins to the DnaK system reportedly leads to unfolding of those proteins [136, 167] (Fig 7.1a). An example is the unfolding of helical segment of transcription factor sigma 32. The DnaK system also binds aggregates of Glucose-6-phosphate dehydrogenase (G6PDH) [168] and dissolves them leading to reactivation [169].

Based on crystallographic analysis and peptide library screening [170, 171], peptides were shown to bind through transient association of short hydrophobic segments with the C-terminal domain of DnaK, the substrate binding domain (SBD): SBD is composed of a binding groove and a lid that closes over the peptide upon ATP hydrolysis (Fig 7.1a). These observations led to the widely accepted notion that the substrate to DnaK is a short extended protein chain that is liberated from folded proteins by thermal excitation. This interpretation has recently been challenged based on geometric arguments: for a DnaK to bind a peptide stretch, the client protein has to unfold extensively to provide appropriate geometry for binding for the hydrophobic stretch (Fig 7.1.a). This is entropically an unlikely event [35]. Using electron paramagnetic resonance spectroscopy and spin labeled DnaK (labeled lid and β-sheet cavity) it has been shown that the lid may not close over bound proteins (Fig 7.1.b), and remains in a more open state that may accommodate regions with substantial tertiary structure [35]. There is some evidence for the interaction of the DnaK system with natively folded P1 RepA, Sigma Factor 32 (at high temperature) and Auxilin [172-175]. Furthermore, quantitative proteomics revealed that DnaK interacts with at least hundreds of newly synthesized or pre-existing cytosolic proteins [166].

Previous effort has provided important insights on the structure and functioning of the DnaK system. However, a number of fundamental questions have not yet been fully answered: What conformations are adopted by the clients and the chaperone when they are in a complex? What is the role of nucleotides and ATP cycle? What are the functional roles of the co-chaperones? How does the system affect protein folding and unfolding processes in healthy state and under heat and ATP-depletion stresses?

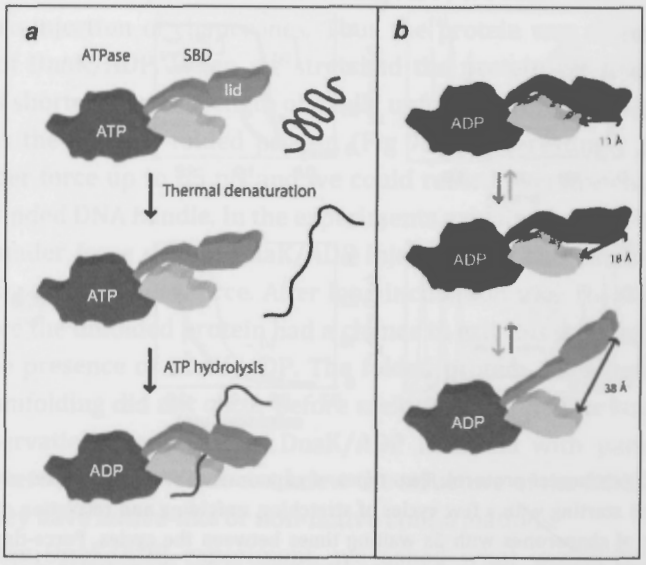


Figure 7.1: Schematic illustration of DnaK conformational states (a) DnaK is composed of an ATPase domain and a substrate binding domain (SBD) with a lid that closes over the peptide substrate. In this model, DnaK/ATP adopts an open conformation and can bind and accommodate unfolded proteins. This is achieved by ATP hydrolysis that results in closing of the binding cavity. Folded protein does not interact with the chaperone unless its is partly unfolded. In this picture DnaK can be regarded as an unfoldase [136]. (b) DnaK in the ADP state is reported to be a structurally degenerate system with three conformations being in equilibrium.

Optical tweezers and Atomic Force Microscopy have been used to characterize the folding dynamics of proteins in isolation and of proteins that interact with ligands and metal ions, but rarely to address how folding is affected by chaperones. Here we investigate how the DnaK system changes the folding dynamics of two proteins: the single-domain maltose binding protein (MBP), which folds efficiently in isolation, and of an MBP repeat construct, which tends to misfold. This

approach allows us to explore how the DnaK system affects the folding transitions within the single-domain MBP as well as misfolding between repeated MBP domains.

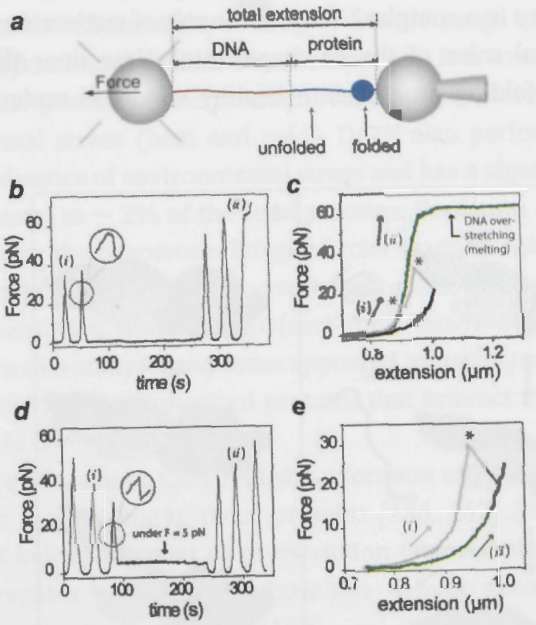


Figure 7.2: Experimental protocol. Two types of experiments were performed using an optical tweezers setup (a), both starting with a few cycles of stretching, unfolding and relaxation of the unfolded protein in the absence of chaperones with 5s waiting times between the cycles. Force-time plots (b, d) illustrate the sequence of events in our experiments. After these cycles, the unfolded chain was then (d) kept under force to avoid refolding or (b) was brought to zero force. The chaperones were infused and the cycles were repeated with 5 s waiting times. Panels (c) and (e) show the force extension plots recorded for DnaK/ADP experiments performed following the described protocol. The stars indicate the unfolding transitions. Note that DNA handle melts and displays apparent elongation at ~ 65 pN, a process which is known as overstretching transition (c).

7.2 Single-molecule experiment

The experiments described in this chapter follow a common protocol: after a tether is established, it is repeatedly stretched and released in the absence of chaperone with 5 s waiting times between unfolding/relaxation cycles (Fig 7.2). The protein unfolds during stretching in one (at ~ 20 pN) or two steps (at ~10pN and ~20 pN) (stars in Fig 7.2 c and d). Refolding typically occurs after 5 s waiting time. After proper unfolding and refolding is established, chaperones are injected at a specific phase of the stretching/relaxation cycle.

First we injected DnaK (1μM)/ADP (1mM) solution while keeping the unfolded protein under force to avoid refolding. When the chaperone was present we relaxed the tether to zero force and stretched it again. We observed that the recorded force extension curve followed the expected one of an unfolded protein (Fig 7.3c). This means that DnaK/ADP interacts with the unfolded chain and thereby prevents its refolding. This is in agreement with the previous observations that DnaK/ADP interacts with peptide stretches with a low off-rate.

In a second experiment with DnaK/ADP, we relaxed the tether to zero force right before injection of chaperones. Thus the protein was allowed to refold in the presence of DnaK/ADP. When we stretched the protein we found that the protein length was shorter than the length of a fully unfolded protein and in some cases it was as short as the natively folded protein (Fig 7.3c). Interestingly the protein did not unfold under force up to 65 pN and we could record overstretching (melting) of the double stranded DNA handle. In the experiments explained above in which the protein was kept under force during DnaK/ADP injection, we had similar observation with long waiting times at zero force. After long incubation time (of the order of minutes) at zero force the unfolded protein had a chance to exit this state and populate a folded state in the presence of DnaK/ADP. The folded protein displayed resistance against force and unfolding did not occur before melting of the double stranded DNA handle. These observations suggest that DnaK/ADP interacts with parts of proteins with significant tertiary fold. We do not know the structure of the folded protein parts and whether they have native-like or non-native conformations.

After DnaK/ADP experiment, we asked what would happen if GrpE (the nucleotide exchange factor) was present. As discussed above DnaK/ADP prevents folding of unfolded protein. Upon addition of GrpE (in the absence of ATP and in the presence of ADP), unfolded protein could fold and an intermediate protein length was recorded with a length between the native state and the unfolded protein. When the partially folded protein intermediate was relaxed to zero force, it

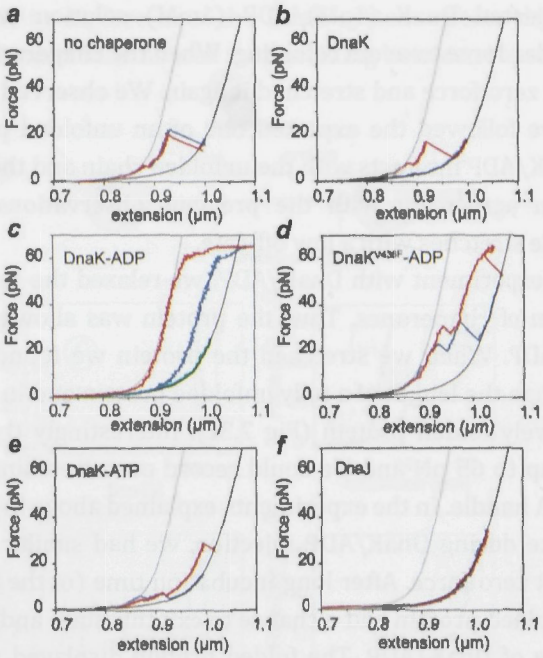


Figure 7.3: Overview of the effect of the DnaK system on single MBP refolding and unfolding. After being unfolded in the absence of chaperones (a), MBP was pulled in the presence of chaperones with 5 s waiting times between consecutive pulls. Two representative pulls are shown in red and blue for each experimental condition (b-f). Grey and black lines are worm-like chain (WLC) models[158] in correspondence with DNA handle and DNA/unfolded MBP construct respectively. During the injection of the chaperone, the tether was relaxed to zero force. In one case (green trace in panel c, the DnaK/ADP experiment) however, the unfolded protein was kept under 5pN force during the injection of the chaperone.

did not exit the state for many cycles and no further shortening was observed. When we stretched it to high forces, we sometimes observed protection against force while in other cases it unfolded at force below 20 pN. The observed increased probability of folding of the fully unfolded state is consistent with the competitive binding of GrpE to DnaK. Apparently MBP and GrpE compete for binding to DnaK, resulting in reduced probability of DnaK induced suppression of MBP refolding. Moreover GrpE increases the off-rate of client bound DnaK in the ADP state, which can be an alternative explanation for our observation. Interestingly, our data suggests that the off-rate of DnaK/ADP does not increase in the presence of GrpE when the substrate has tertiary structure.

The observed interaction between tertiary folded MBP structures with DnaK/ADP raises the question about the location of this interaction on DnaK (Fig 7.4c

1-3). In particular we ask whether this interaction depends on the substrate binding domain of DnaK and on the known peptide binding groove. To test this, we did a similar experiment with a peptide binding defective mutant of DnaK (V436F) in which a valine residing at the border of peptide binding groove is replaced with phenylalanine, an aminoacid with bulky side chain (Fig 7.4c-4). The size and orientation of the phenylalanine are thought to lead to the occupancy of the significant interior space of the peptide binding cleft, abolishing its interaction with client peptides (Mayer, NSMB 2000).

For experiments with DnaK mutant and ADP, we first unfolded MBP in the absence of the chaperone and then the chaperone was injected while the tether was under zero force. We observed that unfolded MBP refolds with a refolding rate (folding probability ~ 90%) comparable to the refolding rate in the absence of chaperone. This result indicates that DnaK mutant in the ADP state cannot interact with unfolded MBP, which is in agreement with previous reports that the chaperone cannot interact with peptide stretches. When the folded MBP with a zero end-to-end length was stretched, we observed a partial unfolding at a force comparable to the unfolding force of natively folded MBP (Fig 7.3d). This is in contrast to wild type DnaK/ADP that protects this state against force (Fig 7.3c). However, in both cases, we observe proteins with intermediate length (shorter than fully unfolded and longer than fully folded) that are stable against mechanical unfolding. These observations suggest that similar to DnaK/ADP, the mutant DnaK/ADP interacts with parts of protein with significant tertiary fold. The substrate binding site of DnaK is found to be affected by the mutation which indicates that it is likely located in the same region as the peptide binding groove. In the experiments with mutant DnaK, we observed that some of the emerged intermediate folds were present transiently during stretching and unfolding and were less stable against force. This could be explained by possible higher off-rate of mutant DnaK as compared to the wild type DnaK. Upon dissociation of mutant DnaK/folded substrate, the tertiary folded part cannot bear the force any longer and will be unfolded mechanically (Fig 7.4a).

An important question is the role of nucleotide. Therefore we performed similar experiments with Apo-DnaK (i.e. DnaK in the absence of nucleotide) and DnaK/ATP. When we incubated the unfolded protein with Apo DnaK at zero force, the next stretching resulted in only two states: the protein was either in the fully unfolded state or in the fully compact state with zero end-to-end length. We observed no intermediate lengths either during refolding or the subsequent unfolding. This is qualitatively similar to the behavior of the protein in the absence of the chaperone. However, the probability of refolding was dramatically reduced from ~ 80% to ~ 50%,

which indicates that Apo-DnaK interacts with the unfolded protein while we have no evidence for its interaction with tertiary folded parts (Fig 7.3b).

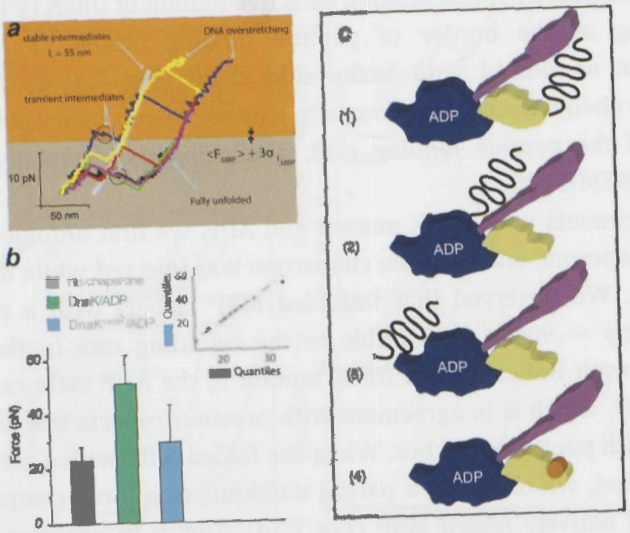


Figure 7.4: Small folds are populated and stabilized by mutant DnaK/ADP (a) a force extension plot demonstrating transient and mechanically-stable long-living intermediates populated in the presence of DnaK mutant and ADP. The region in orange highlights a force regime well above unfolding force of the protein without chaperone. (b) The unfolding force of the native length protein, demonstrating that occupying the cavity does not permit the stabilization of these large folds. The inset is a quantile-quantile plot of the force distributions in the absence of chaperone and DnaK^{V436F}/ADP showing a linear relationship which indicates the similarity of the two distributions. (c) Hypothetical binding modes (1-3) for interaction of a folded substrate with DnaK/ADP. (4) illustrates the DnaK mutant used to demonstrate that mode (1) is likely the correct model and to rule out the two other possibilities.

When DnaK/ATP was injected, we typically observed transient population of states with intermediate length (Fig 7.3e). In some cases we also observed the protein in fully compact or partly folded state with significant stability against force. In the absence of any nucleotide, MBP populates only two states in the stretching/relaxation experiment: the unfolded state and the fully folded state. No intermediate folds are populated and the folded states show the unfolding force and pattern of a native MBP in the absence of chaperone. We conclude that Apo DnaK does not interact with fully folded MBP or folded parts of MBP.

Next we asked whether DnaK interacts with unfolded MBP and how this interaction is affected in the presence of different nucleotides. If a chaperone interacts and protects the extended state and partially folded states of MBP, its effect on the folding rate will be a dramatic slowing down of the folding process. With 5s waiting

between the pulling cycles, we expect to see refolding in ~ 80% of the unfolded chains in the absence of chaperone. With DnaK/ADP this is reduced several fold (Fig 7.5, first bar on the left). For V436F-DnaK/ADP, we did not observe a significant reduction in refolding probability of the extended chain (folding probability = 0.9). Mixture of DnaK with ATP shows a significant fold growth within 5s waiting time from an unfolded state and an enhanced population of intermediate states. DnaK without any nucleotide significantly interacts with the extended polypeptide. Consequently the refolding rate is reduced in the presence of DnaK without nucleotide (Fig 7.5).

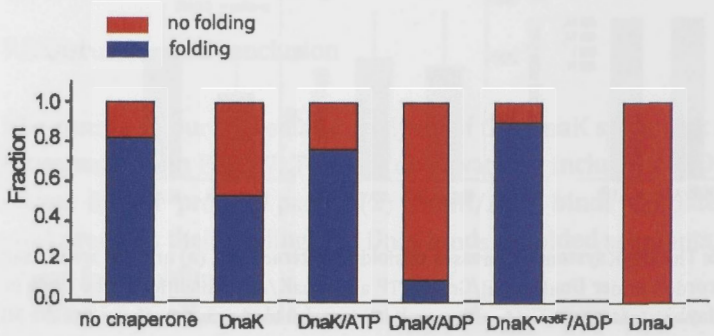


Figure 7.5: DnaK effect depends on the nucleotide state. The bar plot shows the fraction of unfolded proteins that fold partially or completely in the presence of DnaK (in blue) and different nucleotides. In red we see the fraction of the unfolded proteins that remain in the unfolded state after 5 s waiting in the absence or presence of chaperones.

In vivo, DnaK works together with co-chaperone DnaJ and the nucleotide exchange factor GrpE. To understand the role of co-chaperone, we allowed extended MBP to fold in the presence of DnaJ. As displayed in Fig 7.3f, DnaJ binds the unfolded protein and prevents refolding. This is consistent with the observation that DnaJ is able to prevent aggregation of denatured proteins. In this respect DnaJ is similar to SecB (a chaperone involved in protein traslocation [176]) and DnaK/ADP.

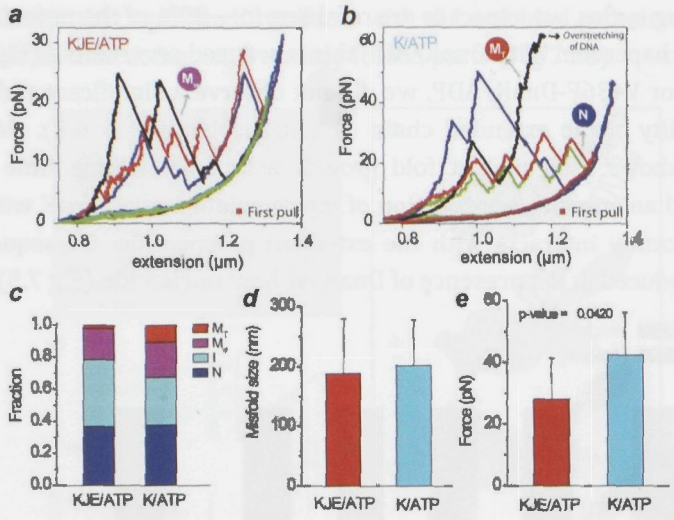


Figure 7.6: The DnaK system suppresses misfolding interactions. (a) and (b) are the sample force-extension traces recorded under DnaK/DnaJ/GrpE/ATP and DnaK/ATP conditions. The trace in red is the first pull and it displays unfolding steps of ~ 90 nm which is equal to the contour length of the MBP core. In panel a and b, during the consecutive pulls steps much larger than the length of one MBP are observed, which indicates the formation of weak misfolds (M_W). In panel b, the trace in black shows a case where the protein resisted forced unfolding. This behavior is interpreted as an indication for tight misfold formation (M_T). Steps smaller than 90 nm are denoted as intermediates (I). In panel a, the green trace shows a case where the protein did not form any tertiary folded structure within the 5 s waiting time. The green trace in panel (b) displays steps with equal length as the first pull (N) which indicates the native refolding of the protein. Similar steps can be seen in panel (a) as well. (c) Statistics of native-like and misfolding events. (d) and (e) compare the unfolding forces and sizes of the weak misfolds (M_W) under two different chaperone conditions.

A natural next step is to have all components together (DnaK, DnaJ, GrpE and nucleotides) and investigate the effect of the chaperone mixture on MBP folding. MBP is a protein that can fold efficiently in the absence of chaperone. Given the efficient folding of MBP and given our observations with individual chaperones, we do not expect to see any beneficial effect of the full system with this substrate. Likely a more complicated protein with inefficient folding will reveal the positive influence of the chaperone system. As such we performed experiments with a 4-tandem MBP construct which is prone to aggregation. When 4MBP is unfolded in the absence of chaperone, it typically misfolds to a state which is stable against forces up to 65 pN, much higher than the unfolding force of natively folded MBP. In many cases the misfolded state has a near zero end-to-end distance and the force extension curve

recorded during stretching will resemble that of double stranded DNA handle alone (Fig 7.5a).

Our experiment with DnaK/ATP and 4MBP as a client shows a significant improvement in 4MBP refolding, as evidenced by appearance of native cores in subsequent pulls. In Fig 7.6b a case is presented in which refolding of all 4 cores succeeded. We still observe tight aggregates and "weak misfolds" signified by steps that are larger than one MBP. Upon addition of DnaJ and GrpE, tight aggregations were only rarely observed and the weak misfolds showed a reduced stability (Fig 7.6 c, e).

7.3 Discussion and conclusion

The results of our systematic analysis of the DnaK system at single molecule level are summarized in Fig 7.7. The key observations include: (1) DnaK/ADP binds and stabilizes folded protein parts; (2) DnaK/ADP binds unfolded segments of proteins and prevents their folding; (3) DnaJ binds unfolded segments of proteins and prevents them from folding.

Binding of DnaK to folded parts depends on its nucleotide state. Apo state cannot bind folded regions while ADP state binds with significant affinity. Both Apo and ADP states bind the unfolded chain, with the latter having a higher affinity.

An important question concerns the conformational states of the DnaK. Our data shows that the client binding involves the peptide binding groove, for binding both to the unfolded stretches and to the folded regions of the client. The folded part occupies a space that goes beyond the inner space of the groove. These findings suggest that the DnaK/ADP adopts an open conformation when it is bound to the folded regions. The binding of the DnaK/ADP to unfolded stretches is suppressed when the inner space of the groove is occupied, suggesting that the DnaK/ADP adopts a closed conformation when it is bound to the unfolded regions. Apo-DnaK binds the unfolded stretches but does not bind folded regions. This suggests that the conformational sampling in the Apo-DnaK state is biased towards the closed conformation. The DnaK in the ATP state does interact with folded regions but the complex is less stable than the complex in the ADP state. This suggests that the DnaK/ATP also adopts open formations that differ from the open conformations corresponding to the ADP state.

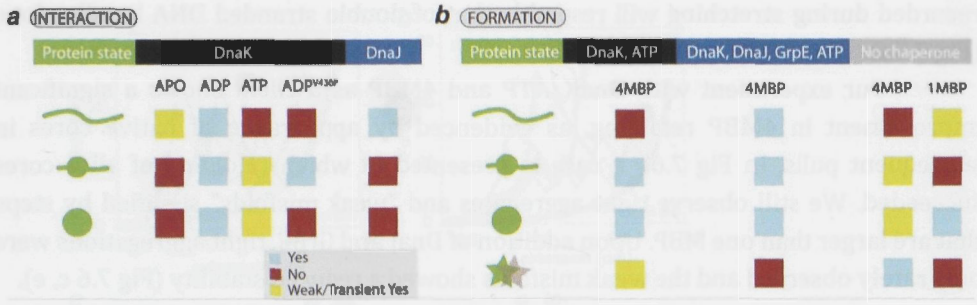


Figure 7.7: Formation and interaction of protein states with the DnaK system (a) interaction between protein state and the chaperones. (b) formation of a protein state in the presence and absence of chaperones. The green circle represents natively folded state (or a compact state with nearly zero end-to-end distance), while the stars represent misfolded conformations and inter-domain contacts. The green line represents the completely unfolded protein and the green line with the circle indicates partially folded proteins or proteins with some degree of native structure.

7.4 Appendix

7.4.1 Current model for the functioning of the DnaK system

The existing model for the functioning of the DnaK system has been subject to little changes within the last two decades and has been used to explain both stress response and assisted folding of nascent chains [6, 177, 178]. In this model the substrates of the DnaK system are the extended parts of partially (mis)folded proteins that interact and get stabilized by DnaK/DnaJ complex in a process driven by ATP hydrolysis. After the formation of the DnaK/ADP/DnaJ complex, DnaJ is assumed to leave the complex, which is followed by nucleotide exchange (mediated by GrpE) and substrate release. A fast folding substrate will then fold and a slow one will be grabbed again by DnaJ and DnaK and enters a new cycle. This means that as long as the stress is present, a slow folding protein will keep binding and unbinding the DnaK system [6], in a process that consumes energy, but without any progress in the folding process. In this picture protein aggregation is suppressed and under an excess of chaperones the system reaches a steady state in which a significant fraction of the proteome is frozen in the unfolded state until the stress is gone (Fig 7.8.d). This is because the presence of chaperones with a high affinity for unfolded stretches leads to a reduction in the free energy of the unfolded state and thereby the unfolded state is promoted (i.e. DnaK acts as an unfoldase). Thermal stress also reduces the free energy of the unfolded chain due to promotion of entropic effects and therefore unfolding is promoted under thermal stress.

The current model raises a few questions: what is the energy burden of this model? How efficient is this model? What is the use of functional redundancy of components in this model? Some parts of the model are not supported by strong evidences: One question of debate concerning the DnaK-chaperone cycle was the dissociation of DnaJ. On one side DnaJ/DnaK-ADP/client complex is considered to be stable and on the other side DnaJ is said to leave the client before nucleotide exchange. Based on the fact that DnaJ is only 1/10th to 1/30th as abundant as DnaK in vivo and can act sub-stoichiometrically, it was assumed that DnaJ leaves the cycle just after the transfer of the substrate onto DnaK and before GrpE binds to the complex [179]. The consequence of this process is a functional redundancy for DnaJ and DnaK-ADP in this model (see panel b; some non-overlapping function for DnaK is also speculated: DnaK is said to do the same function as DnaJ but at a later stage when the protein is unfolded enough. DnaK also enclose the chain while DnaJ does not need to interact with the backbone).

The energetic burden of this model will sound even more serious when we consider the concentrations of the components: cellular concentration of ATP is 1-10 mM under healthy conditions which can go down to 15-25% under ATP depletion stress. The total cellular proteins amount to 10 mM of which 2% is DnaK (12-20% under stress). DnaK/DnaJ/client substrate hydrolyzes ATP at a rate of $\sim 1.8s^{-1}$ per complex. This means that the existing model is imposing a significant energy load on the system. Heat stress itself also leads to the hydrolysis of ATP to free nucleotides leading to a drop in ATP concentration to half its normal value within minutes in eukaryotic cells [180, 181]. For *E. coli,* it has been suggested that the cell responds to a temperature up-shift by transient increase of the concentration of the sum of adenosine nucleotides (AXP) followed (after several minutes) by the ATP drop [182]. Chaperone over expression will therefore coincide with the drop in the ATP level. The transient increase has been attributed to RNA degradation (total RNA concentration per cell drops to 50%) and an increase in respiration and glucose uptake. ADP/ATP ratio increases (approaching one) and the high concentration of ADP may lead to population of the ADP state of many chaperones. The production of heat shock proteins is also costly and a significant part of the cellular energy budget will be allocated to their synthesis [183, 184]. Furthermore, the function of stress proteins may require --based on the current model-- considerable ATP turnover; refolding of a protein may consume in excess of 100 ATP molecules [184, 185].

7.4.2 Proposed model for the functioning of the DnaK system

Our data allow us to build a hypothetical model for DnaK activity in heat stress and ATP-depletion stress responses (Fig 7.8.a). The model is based upon (Fig 7.7) and a few additional geometric and logical arguments and assumptions: (1) when a small part of a protein is unfolded, DnaJ can approach the segment with less steric hindrance than DnaK due to its shape, size and flexibility (2) DnaJ/DnaK/ADP in complex with the substrate is stable and DnaJ does not leave the complex under the condition of heat stress. This is because hydrophobic interactions are entropic in nature and therefore are promoted under heat stress. (3) GrpE is a thermosensor and adopts a different state under thermal stress as compared to normal conditions and cannot efficiently exchange nucleotides.

Based on these, we propose a new model as described below. In the proposed model, the proteins are targeted at early stages of the thermal stress when the structure of the proteins is only slightly perturbed. The targeting is performed by DnaJ binding to short hydrophobic segments that are liberated from the proteins. DnaK is then recruited and by ATP hydrolysis the DnaK ADP accommodates the nearby folded structures and protects them from further unfolding ("break function").

The energy burden of this model is less that the current model and the model works under energy depletion stress too. Protection of the folded substrates keeps the proteome near its native state which saves lots of energy that would otherwise be required for protection of large amounts of unfolded structures. This way the system can protect the proteome in the absence of ATP (ATP depletion stress). In the proposed model the functional redundancy is minimal (Fig 7.8b,c) which in turn leads to a significant reduction of energy burden. For instance a peptide stretch can be protected by both DnaK/DnaJ and DnaJ alone. The former would use energy while the later can protect the peptide with no cost. In the proposed model the redundancy is reduced because the proteome is frozen at early stage after stress and therefore protection of short stretches is likely carried out by DnaJ and folded parts are protected by DnaK. The system's steady state is now close to the native proteome rendering the recovery of the system efficient and fast upon releave of stress (Fig 7.8.d). The proposed model is also consistent with the reports suggesting that GrpE unfolds reversibly at elevated temperatures. This ensures that nucleotide exchange is slowed down at heat shock conditions and, since ADP release is rate-limiting for substrate release, substrate dissociation is also slowed down [186]. We note that the current model would be energetically costly even with this reduced activity of GrpE.

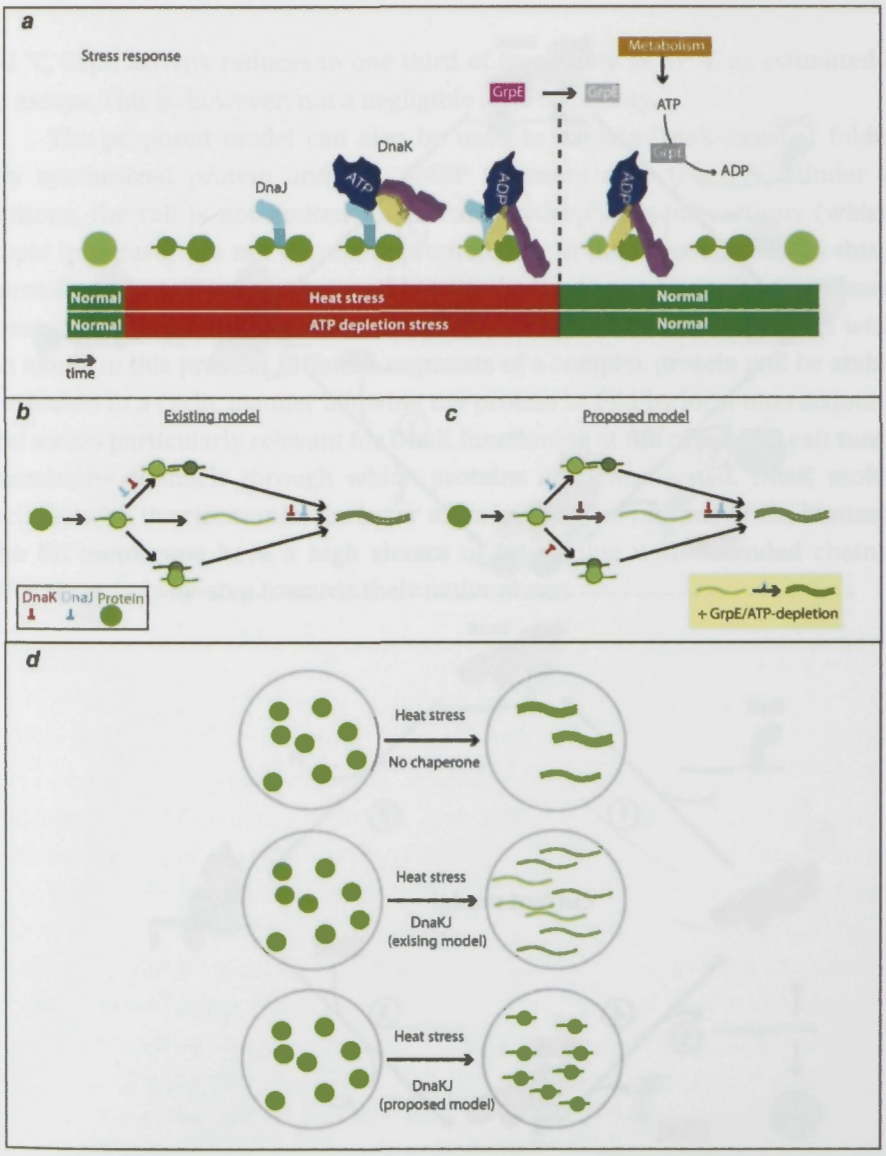


Figure 7.8: Proposed model for DnaK mediated stress response (a) schematic illustration of the model. GrpE is colored differently under stress condition to highlight that the molecule may be in a different conformational and functional mode under heat and ATP depletion stresses. The functional role of DnaK and DnaJ are depicted based on current models (b) and the new model (c). The evident redundancy in (b) is absent in (c). In (d) the steady state of the system is shown for different scenarios. In the new model the proteome is frozen in near native conformation. The DnaK system in the new model antagonizes both detrimental effects of thermal stress namely boosting unfolding and aggregation. In the current model only the later detrimental effect of the stress is suppressed.

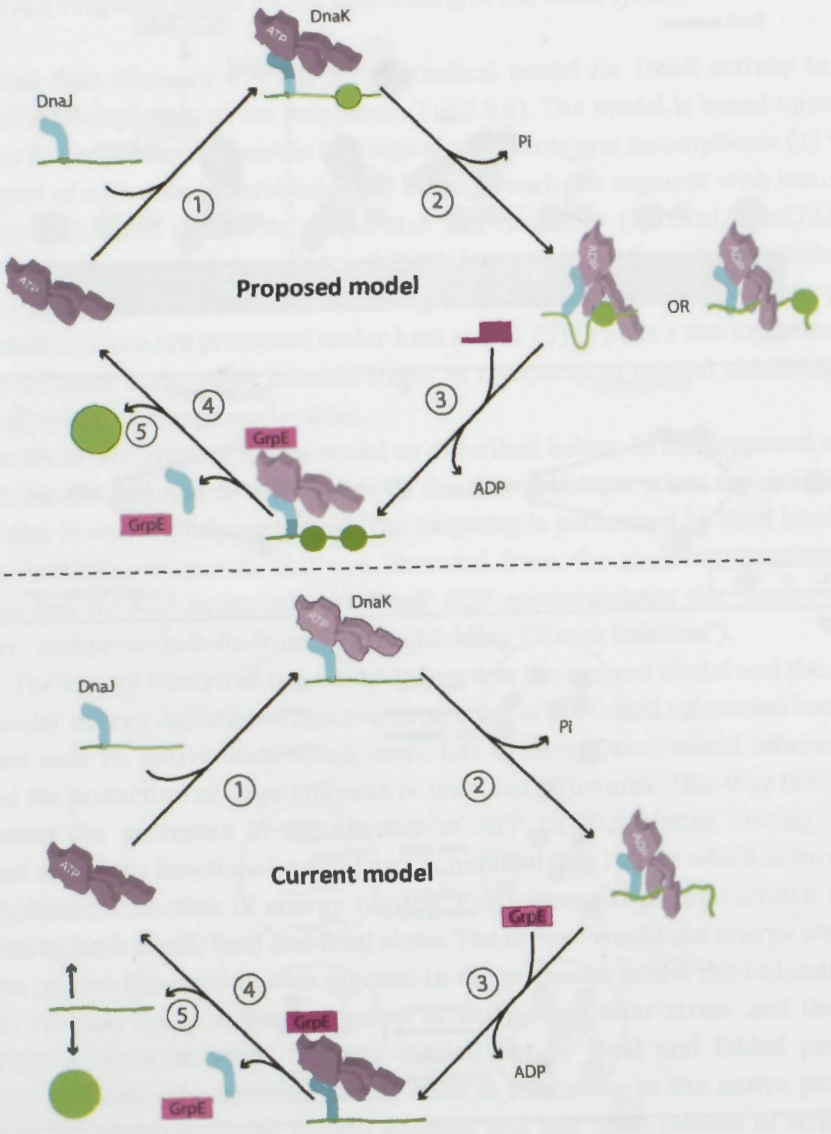


Figure 7.9: DnaK reaction cycle and de novo folding. Top panel presents the proposed model: (1) DnaJ-mediated delivery of substrate to ATP-bound DnaK. (2) Hydrolysis of ATP to ADP, accelerated by DnaJ, results in tight binding of substrate by DnaK. (3) Dissociation of ADP catalyzed by GrpE. Apo-DnaK does not show affinity to folded parts and has a reduced affinity (compared to ADP-DnaK) to unfolded regions. (4) DnaJ dissociates from its complex with DnaK and the substrate. (5) Released substrate either folds to its native state or is transferred to downstream chaperones. A similar cartoons can be drawn to explain folding of multi domain proteins, the models fits well with the 4MBP data presented in Fig 7.6. the lower panel presents the current model (see section 7.4.1)

At 48 °C, GrpE activity reduces to one third of its activity at 37 °C as estimated by in vitro assays. This is, however, not a negligible level of activity.

The proposed model can also be used to explain DnaK-assisted folding of newly synthesized protein and our 4MBP tweezers data (Fig 7.9). Under these conditions, the cell is not limited in ATP and hydrophobic interactions (which are entropic in nature) are not thermally promoted as in the stressed cells. In this case, the protein chain will interact reversibly with the DnaK system via DnaJ assistance or without it. Due to its concentration, DnaJ will not have a chance to interact with the chain alone. In this process, different segments of a complex protein will be stabilized and released in a cyclic manner allowing the protein to fold by local interactions. This model seems particularly relevant for DnaK functioning at the ribosomal exit tunnel or at membrane channels through which proteins are translocated. DnaK molecules associated with the ribosomes, the inner mitochondrial membrane or the luminal side of the ER membrane have a high chance of interacting with extended chains and guiding them step-by-step towards their native states.

CHAPTER 8

Protein folding under the influence of HtpG

Hsp90 is involved in protein quality control, assists degradation of mutated proteins, and has become recognized as an important anti-cancer drug target. HtpG, the bacterial homolog of the Hsp90, is an ATP driven chaperone that, in contrast to Hsp90, functions without any co-chaperones. It forms a homodimer with open and closed conformations. Here we investigate the mechanism that underlies chaperone function of HtpG, by studying its influence on the folding and unfolding pathways of the model protein Luciferase at single molecule level. HtpG was found to interact with the unfolded chain in an ATP independent manner. We found steps during refolding and unfolding that depend on HtpG and its ATPase function. The emergence of steps suggests interaction of HtpG with the folded regions of proteins or loop induction in polypeptide chains.

8.1 Introduction

The 90 kDa heat shock (Hsp90) family of proteins are to ATP-driven homodimeric multi-domain molecular chaperones that interact with a wide range of client proteins in higher eukaryotes [187]. The molecular mechanism of the chaperone function remains poorly understood. The Hsp90 family of proteins are believed to also act as hubs of oncogenic signaling networks and are identified as promising pharmacological targets for the treatment of malignancies [188].

The structure of the Hsp90 homolog in *E. coli* (HtpG) has been investigated with a multitude of experimental and computational techniques (Fig 8.1). Three domains have been recognized for HtpG: N-terminal, Middle and C-terminal domains. These studies have revealed dramatic conformational changes in HtpG that accompany its ATP cycle (K_m = 250 μM, ATPase rate = 0.011 s^{-1}) [189]. Crystallographic and SAXS approaches have found two conformers for the apo-HtpG dimer: the open (V-shaped), and the extended states [190, 191]. These two states interconvert with a bias towards the open conformation. In the presence of ADP, a highly compact semi-closed dimer structure is observed. Studies based on mutational analysis, cross-linking and electron microscopy (EM) have revealed that the open and closed conformations may coexist in a dynamic equilibrium for the apo, ADP, and AMPPNP states [191]. Conformational dynamics of HtpG in solution has been also studied by using amide hydrogen exchange mass spectrometry (HX-MS) and fluorescence spectroscopy, indicating that HtpG continuously fluctuates between closed and open conformations irrespective of its nucleotide state [192].

Computer simulations have revealed that nucleotide binding profoundly affects the internal dynamics of HtpG both locally and globally, in which the quasi-rigid domains move relative to each other [59, 60]. Two hinges are responsible for these movements: one located at the boundary between the N-terminal and Middle-domains and a second one is located at the end of a three-helix bundle in the Middle-domain. The latter unfolds upon ATP hydrolysis. The presence of anti-correlated motions between the NTD and CTD in the open apo-HtpG form, and a highly mobile ADP-bound form are the two fundamental characteristics of the HtpG chaperone dynamics.

Client protein binding affects HtpG conformation [189]. The nature of client-chaperone interaction and the conformations of the chaperone and the clients in the complex are not yet well understood. The reported values for the stoichiometry of the interaction between the Hsp90 and its (protein/peptide) substrate vary depending on the Hsp90 homologs and the conditions in the chaperone assay, and range from 0.4 to

more than 16 [193]. Structural analysis of Hsp90 in complex with a substrate showed an asymmetric binding of the substrate to the N-terminal and middle domains of one of the two monomers of the chaperone dimer [194]. Considering the symmetry of the dimer, it is debated if HtpG has one or two binding sites for its substrates in solution and whether 1:1 or 2:1 stoichiometry (peptide-chaperone) is the dominant stoichiometry [195, 196].

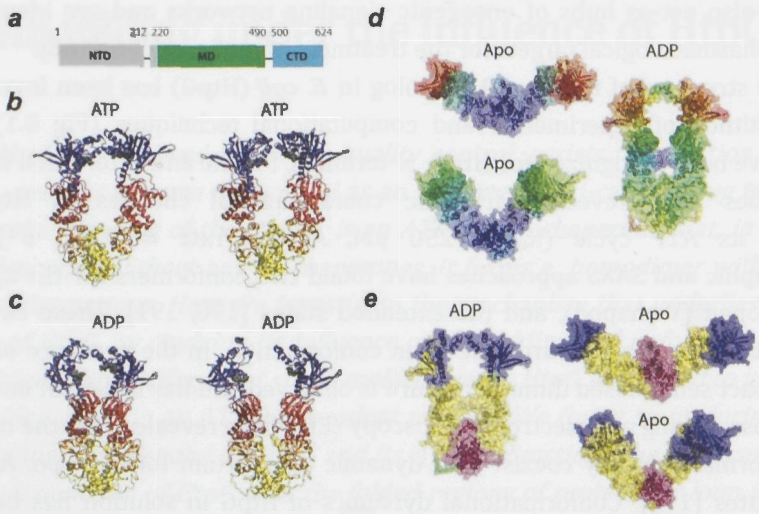


Figure 8.1: Structure of HtpG (a) HtpG has three domains: N-terminal, Middle and C-terminal domains. (b) shows two dominant conformations of the ATP state as obtained by simulations. The three domains are shown in blue, red and yellow respectively. (c) presents the simulated conformations of the ADP state. (d) Structures of the Apo state, which are extended and V-shaped and the structure of the ADP state, which is closed. The surface is color coded according to the mobility of the residues (from more rigid blue regions to more flexible-red regions). (e) X-ray structure of ADP and Apo states (lower right) and SAXS based structure of Apo state (upper right). The figures are adapted from Dixit A, Verkhivker GM. PLoS One. 2012;7(5):e37605 and Morra G, Potestio R, Micheletti C, Colombo G. PLoS Comput Biol. 2012;8(3):e1002433.

The *in vivo* characterized substrates vary in size (14-290 kD), sequence and structural motif [197, 198]. Constitutively unfolded protein Δ131Δ is commonly used as a substrate for HtpG. Δ131Δ is a fragment of staphylococcal nuclease which is globally unfolded (with significant residual structure), yet does not aggregate even at high concentrations. Binding of this substrate stimulates ATP hydrolysis in HtpG. In the Apo state, HtpG partially closes around Δ131Δ. Furthermore the fully closed AMPPNP state shows a high affinity to the substrate [199]. NMR studies have revealed that the chaperone binds to the structured region of Δ131Δ. Based on this observation, it has been suggested that the chaperone preferentially binds a locally structured

region in a globally unfolded protein. Others reported that HtpG binding to substrate commences after substrate unfolding [200], and HtpG has been found to bind unfolded proteins [201].

Key questions concern (1) whether substrate contacts are within a single monomer or across HtpG monomers, and whether one or multiple substrate contacts are utilized, (2) determining how substrate and nucleotide binding influence HtpG conformational dynamics, (3) how HtpG mediates client protein folding. [202, 203]

8.2 Single-molecule folding and unfolding experiment on a model protein

Here we investigate the folding and unfolding of the model protein luciferase in the presence of HtpG using optical tweezers. Luciferase is a 61 kDa protein with 550 residues which is commonly used as a model multi-domain protein in folding studies. The protein termini were linked to two polystyrene beads via molecular linkers and the protein was stretched and relaxed using an optical tweezers setup [204]. While one of the beads was kept by the laser trap, the other bead was moved at a constant speed of 50 nm s^{-1}.

Firstly we stretched the protein from its ends until it fully unfolded and then exposed the unfolded luciferase to HtpG (1μM), in the presence of 1mM ATP and an ATP regenerating system (3 mM phosphoenolpyruvate and 20 ng/ul Pyruvate kinase). The folding probability was measured by incubating the unfolded protein with the chaperone solution for 5 s at zero force followed by stretching the protein to characterize its length. A reduction in the protein length as compared to the length of the fully unfolded protein, indicated the protein had folded, either to the native state, orto a partially folded state. This data was used to compute the folding probability. We observe a decreased folding probability of the unfolded protein in the presence of wild type HtpG (Fig 8.2). The folding probability was further decreased when ATPase deficient $HtpG^{E34A}$ was used. This finding indicates that the unfolded chain interacts with the chaperone and the interaction does not depend on the ATP hydrolysis per se, but its affinity and kinetics depend on the nucleotide state.

In the absence of chaperone the unfolded protein typically relaxed to zero force with a force-extension relation identical to the force-extension curve of the worm-like chain (WLC) model with the contour length of the unfolded luciferase. With chaperone, two types of deviations from the worm-like chain model were observed: gradual (Fig 8.3.a black star) and stepwise compactions (Fig 8.3.a green star) of the chain were detected. During stepwise compactions, folding occurred despite applied force and the reduction in extension was often accompanied by an increase in force.

Intriguingly, during relaxation, some regions of the force extension curve were aligned with the worm-like chain fits to the luciferase unfolding intermediates, while other regions did not show this behavior (Fig 8.3b). This indicates that the same states are sometimes observed during refolding and unfolding and that this process is HtpG dependent.

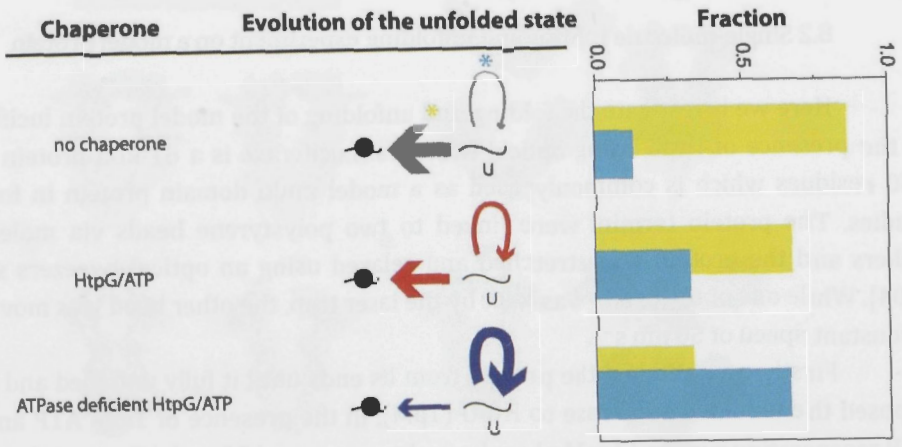


Figure 8.2: HtpG interacts with unfolded luciferase in an ATPase independent manner. In this experiment, unfolded luciferase was allowed to fold in the absence and presence of wild type and ATPase deficient mutant HtpG. The thickness of the arrows indicates the probability of finding the protein in the unfolded state (loop arrow) and in a partially or completely folded state after 5 s waiting time. Interaction with the unfolded chain leads to an increase in the fraction of finding the protein in the unfolded state after the waiting time. The probabilities are also presented as a bar plot.

We characterized HtpG induced refolding steps with their step length change and the force at the refolding event. For this aim, we converted the recorded force-extension curves for the protein-DNA tether to force-protein length plots by fitting the worm-like chain model of a DNA-polypeptide model with the polypeptide length as the fitting parameter. Steps were manually detected from the resulting force-protein length plots. The refolding step length distribution (Fig 8.4b) showed a distinct peak with a characteristic step size of 11.85 $\pm$ 1.04 nm. The steps occurred at an average force of 14 $\pm$ 7 pN. Interestingly, the observed stepsize is similar to the reported distance change between the N-domains during open-close transition of HtpG (~13 nm) [190].

HtpG also profoundly altered the pattern observed when stretching luciferase (Fig 8.4.c). Increased occurrences of small step sizes (< 14 nm) was observed as compared to the frequency of small stepsizes in the absence of the chaperone. When the unfolding forces of these small steps were analyzed, two populations were distinguished, one unfolding at forces below 25 pN and another population with a higher unfolding force (Fig 8.4c,e). The former group has a similar force/stepsize statistics as the steps observed during refolding (Fig 8.4.c, d).

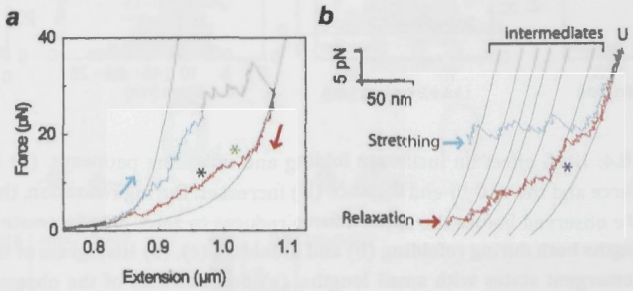


Figure 8.3: A representative force-extension curve recorded during unfolding and subsequent refolding of luciferase in the presence of HtpG/ATP. (a) Force extension plot during unfolding (blue) and refolding (red). During refolding marked deviation from the worm-like chain (gray curve on the right) is seen. Two types of events can be distinguished that underlie the deviation: step-wise contraction (green star) and gradual contraction (black star). (b) Alignment of the visited lengths during refolding and unfolding. In the presence of HtpG, relaxation trace sometimes follow (purple star) the WLC models corresponding to the unfolding intermediates in the absence of the chaperone.

The emergence of these steps might be an indication of a protein-bound closed form of HtpG. These steps might be a signature of HtpG clamping the protein chain and forming loops. For this process to take place, ATP hydrolysis is essential. Thus to test the hypothesis we performed a similar type of experiment with $HtpG^{E34A}$, the ATPase deficient mutant of HtpG.

In the presence of $HtpG^{E34A}$/ATP, the relaxation of the unfolded luciferase did not result in any significant deviation from the force-extension curve of the unfolded protein. However when we estimated the refolding probability, we observed a reduced refolding probability indicating that $HtpG^{E34A}$/ATP does interact with the unfolded chain. Interestingly the unfolding length histogram was rather close to the no chaperone case (Fig 8.5b) in the presence of $HtpG^{E34A}$/ATP, with some increased prevalence of small steps. A closer look at these steps showed that they are not overlapping with the characteristic steps observed in the presence of HtpG (Fig 8.5c) and are larger.

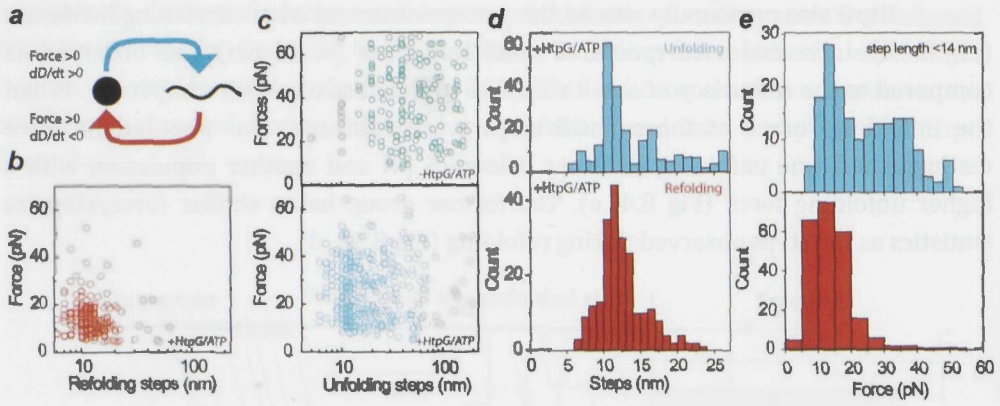


Figure 8.4: HtpG effect on luciferase folding and unfolding pathways. (a) During unfolding the protein is under force and the end-to-end distance (D) increases. During relaxation, the distance decreases. Refolding steps are observed before the applied force reduces to zero. The force-step plots show a dense region at small lengths both during refolding (b) and unfolding (c). (d) Histogram of the lengths zoomed to characterize the emergent states with small lengths. (e) For statistics of the observed small steps. Two peaks can be distinguished in the unfolding data, a narrow peak at low force and a broad peak at higher forces. Luciferase in the absence of chaperone have a unfolding force distribution that overlaps with the observed broad peak in the presence of chaperone.

In the presence of wild type HtpG, luciferase visited the characteristic unfolding intermediate lengths of luciferase (described in chapter 3) with higher frequencies per unfolding trace as compared to unfolding in the absence of the chaperone. When the chaperone is absent, luciferase visits various intermediate lengths during unfolding. Sometimes multiple contacts can break in one go and the intermediate lengths will not be visited. Therefore not all intermediates are observed in one unfolding experiment. The fact the we have a higher chance of visiting the intermediates in the presence of the chaperone suggests that the chaperone may reduce the probability of these concerted unfolding events.

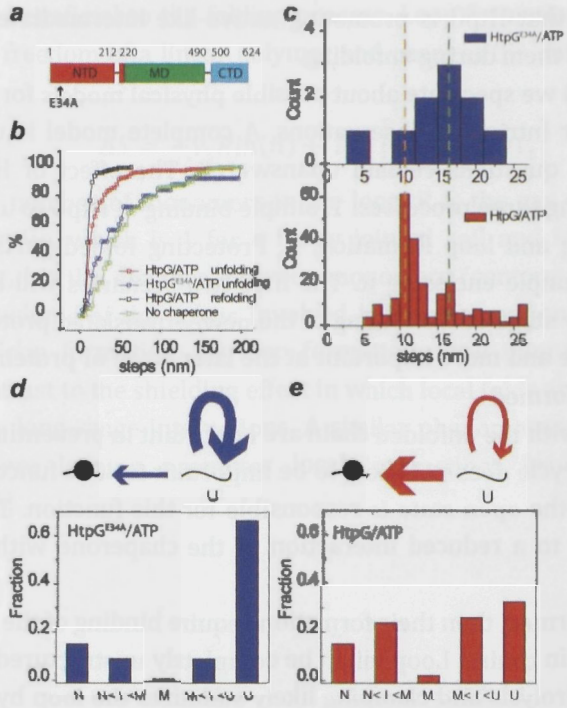


Figure 8.5: HtpG-induced small steps depend on ATP hydrolysis. (a) Structure of the HtpG^{E34A} with the mutation site highlighted. (b) and (c) show the length statistics in the presence of HtpG^{E34A}. (d) and (e) Refolding of luciferase in the presence of mutant HtpG (purple) and wild type HtpG (red). Starting from the unfolded state, the fractions of the final states visited after 5 s waiting time are plotted.

In the absence of chaperones, luciferase populates a non productive middle state (M-state) during refolding (chapter 5). Intriguingly, this state is avoided in the presence of HtpG/ATP. Similarly when HtpG^{E34A}/ATP was present, we rarely observed the M-state being populated after 5 s waiting time. Thus suppression of the M-state does not seem to be dependent on ATPase activity of the chaperone (Fig 8.5.d, e).

8.3 Discussion and conclusion

In summary we observed: (1) the emergence of small steps during refolding and unfolding that depends on ATP hydrolysis (2) interaction of HtpG with the unfolded chain in an ATP independent manner (3) suppression of a misfolded state (4) higher frequency of visiting native unfolding intermediate lengths (5) partial alignment of refolding intermediate lengths with unfolding intermediate lengths. Our

observations suggest that HtpG is promoting native-like intermediate states during refolding and protects them during unfolding.

Here (Fig 8.6) we speculate about possible physical models for functioning of HtpG that explain our intriguing observations. A complete model is unclear at this point and still many questions remain unanswered. The effect of HtpG might be understood considering three processes: 1. Simple binding of HtpG to a chain (similar to SecB); 2. Clamping and loop formation; 3. Protecting folded parts (TF like, see chapter 5) by for example enclosing it. The first two scenarios will be particularly important at the early stage of the folding of the newly translated proteins, while the third one will be more and more important at the later stage of protein folding when folded structures are formed.

Interactions with the unfolded chain are important in preventing aggregation. The opening/closing cycle does not seem to be important for this function and in our experiment probably the open state is responsible for this function. This is because ATP hydrolysis leads to a reduced interaction of the chaperone with the unfolded chain.

If loops are formed, then their formation require binding of the HtpG dimer to two sites on the protein chains. Loop might be completely unstructured or have some folded parts. ATP hydrolysis and clamping likely stabilizes the loop by tightening its endings. Clamping events can be observed as reductions in the end-to-end distance of the unfolded protein under moderate force. We note that the contour length of the loop is not related to the length of the steps observed in our unfolding experiment. During unfolding experiments multiple events might occur which results in heterogeneity of the recorded length changes: the closed form of HtpG can be forced to open up; the chaperone might be ripped off from one of the loop's ends resulting in complete or partial unfolding of the loop; the HtpG generated loops might have various topological arrangements with for example one being formed within another one. The length changes observed in our experiments can therefore be classified to chaperone-specific, protein specific and random steps. Finally, to explore whether HtpG indeed forms polypeptide loops and compacts linear polypeptides complementary single-molecule or bulk experiments are needed.

8.4 Appendix

It is conceivable that protection of folded parts or unfolded stretches of a protein by a chaperone helps in protein folding and suppresses the aggregate formation. It is however not very clear how the formation of loops with unspecified

lengths can be beneficial to the folding process. Loop formation is known to reduce the degrees of freedom of a linear polymer and changes the entropy of the extended chain by:

$$\Delta S = -\nu_3 R ln(n) + Rln\left[\left(\frac{3}{2\pi C_n l^2}\right)^{\nu_3} V_t\right]$$

where n is the number of monomers in the loop, R is the gas constant, C_n is Flory's characteristic ratio, which is 1 for a freely jointed coil and increases as the chain becomes stiffer, l is the distance between monomers (contour length/residue), V_t is the approach volume of the atoms involved in loop formation and ν_3 is the scaling exponent for loop formation. By loop formation, local interactions are promoted directly, in contrast to the shielding effect in which local folds are promoted indirectly by suppressing long-range interactions. A similar phenomenon is observed for DNA molecules where looping promotes local interactions leading to transcription activation [205].

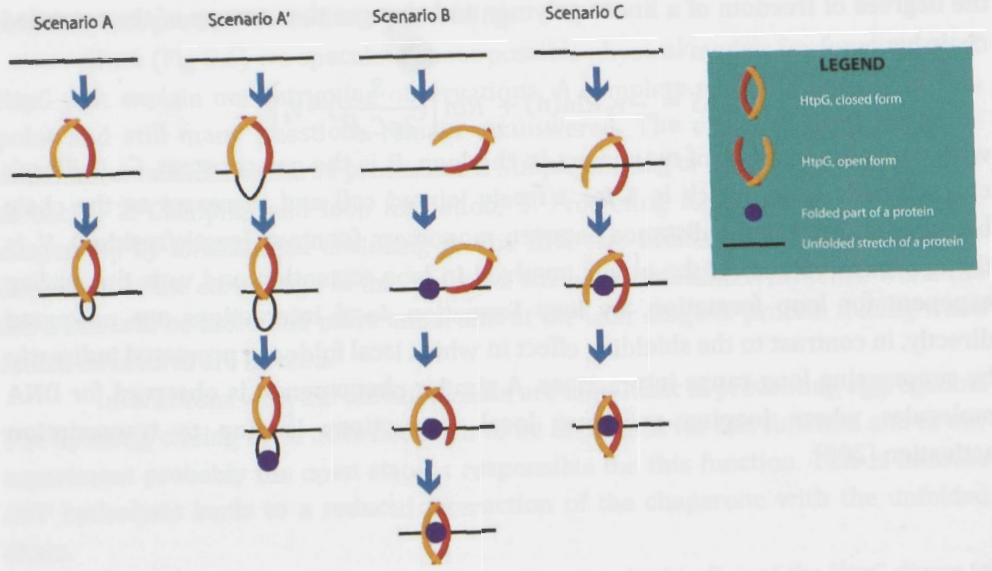


Figure 8.6: Speculations for possible mechanisms of HtpG chaperone activity. Four scenarios are presented. Low force transitions between protein-chaperone complex states are shown. In each case the reverse process happens at high force. In scenario A, the chaperone binds the chain at two sites and a polypeptide loop is formed. This model provides rationale for ~13 nm during unfolding and refolding. Alternatively, in scenario A', the loop with unspecified length is formed and is grabbed from its two ends by the chaperone. A transition from open to closed chaperone state will result in ~ 13 nm step irrespective of the loop size (the length change is chaperone specific). Loop formation however changes the entropy of the chain and promotes local interactions resulting in local folds. In scenario B the chaperone binds the chain via one of the monomers and hence suppresses refolding (like SecB [18]). The protein folds partially and spontaneously after which the chaperone stabilizes (like TF or destabilizes) the fold by binding it. Unfolding this structure will likely result in length changes that are protein specific. The unfolding forces will be however changed when chaperone is bound. In scenario C, the chaperone acts like a folding cavity. It encloses part of the chain and the folding occurs between its arms.

CHAPTER 9

Cellular response to heat and other external perturbations

Chaperone systems participate in stress response processes, both as stress sensors as well as stress buffers. However, other cellular components and biomolecules are also active in sensing and stress response and cooperate with, complement or can replace chaperones. Here we discuss briefly the role of lipids and sugars in heat stress response. We also briefly discuss pharmacological stress and shear stress, two important and common perturbations, biological systems are exposed to. We show how proteome respond to pharmacological perturbations and how medical interventions can be managed to minimize the unwanted stress.

9.1. Membranes and heat stress response: phase transitions, rate of thermal energy transfer and proton hopping

In a cell, the energy scale is comparable to the natural thermal fluctuations in the system. Consequently, protective measures have been evolved to combat externally imposed thermal stresses. Nearly half of the energy that is derived from the oxidation of glucose or fatty acids is released by cells as heat. Cells are also exposed to heat sources in their environment. In humans, the temperature of the cellular environment rises in febrile states (global rise in body temperature), or locally at inflammation sites and sites of malignancy with high cellular metabolic rate or due to medical interventions. In ectothermic animals and other organisms like unicellular organisms, cells might be exposed frequently to aggressive thermal fluctuations, for example by exposure to sun light.

Cells sense and respond to thermal stresses and thermal gradients. Lipid membranes, associated temperature gated ion channels and membrane associated chaperones are at the forefront of sensing and responding to environmental stresses. These various systems are active in this process at different timescales and stress levels.

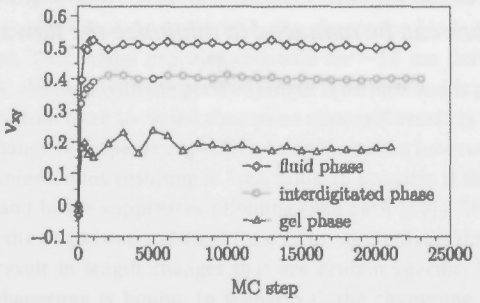


Figure 9.1: Relaxation of $-\Delta Lx/\eta\ \Delta Ly$ in fluid, interdigitated and gel phases after a rescaling of the inter-bead distances in y-direction by a factor $(1+\eta)$ and fixing the membrane size in y-direction. The plateau value reached after about 5000 Monte Carlo steps, independent of the value of η, yields the Poisson's ratio ν.

Lipid bilayer membranes are known to respond to temperature showing a rich phase behavior: fluid-, gel- ripple and interdiggitated phases are visited at different temperature and in some cases multiple phases co-exist at a given temperature. Membrane phases differ in their mechanical properties that are in turn known to play an important role in a number of membrane processes, as, for example, membrane fusion and modulations of membrane channel activities. Cell membrane mechanics at small scales (< 300 nm) is dominated by the lipid part and the effect of

cytoskeleton is minimal. Membrane mechanics can be characterized by two measures: Young's modulus and Poisson's ratio. Efforts has been made to measure these properties but for Poisson's ratio there has been no success due to technical difficulties. Poisson's ratio is defined as the ratio of the size change in the direction perpendicular to the applied force versus the expanded length in the direction of the force [206]. Experimentally it is difficult to measure Poisson's ratio of thin films because of the small thickness of membranes in the nanometer range. Estimation of Poisson's ratio by simulation of the membranes is not trivial because these simulations are typically done under periodic boundary conditions. For the coarse grained lipid model introduced by Lenz and Schmid [207], we successfully calculated the Poisson's ratio in the gel, fluid, and interdigitated phases (Fig 9.1) under periodic boundary conditions. This is a step towards accurate characterization of the mechanical properties of cellular membrane and the response of the system to perturbations.

One of the important features of cellular membranes is their asymmetric lipid distribution between the leaflets. To establish and maintain asymmetric membrane lipid composition, transbilayer movement of lipids from one leaflet to the other must be regulated. The transbilayer movement occurs by either passive flip-flop or via an enzyme-catalyzed process and its probability is expected to vary depending on the temperature of the system and the membrane phase. For example, using the Jarzynski equality [14], we calculated the free energy of the transition from the work associated with non-equilibrium movements of single lipids across a symmetric phosphatidylcholine bilayer. We found that the associated free energy difference is a factor of 5 larger in the gel phase as compared to the fluid phase (estimated from 50 sample paths). The method can be applied to estimate the free energy of the lipid flip flop across any arbitrary bilayer.

For natural lipid bilayers that are composed of various lipids, phase segregation may also occur in response to a change in temperature. Cells also modulate the fraction of the desaturated lipids in the membrane in response to thermal fluctuations. Cholesterols are also known to have critical role in thermal adaptations of cell membranes.

In many cases, however, thermal stresses have to be sensed and avoided well before phase transition temperature of the membrane is reached which might be in turn lethal to the cell due to a large number of functional molecules associated with the membranes. Membrane is associated with a number of thermo sensors and temperature gated ion channels like TRP ion-channel subfamilies, molecular Rhodopsin and membrane associated heat shock proteins. Vibrational dynamics of

membrane might be coupled to the vibrational dynamics of associated protein and thereby affects its binding affinity and allosteric regulation [208].

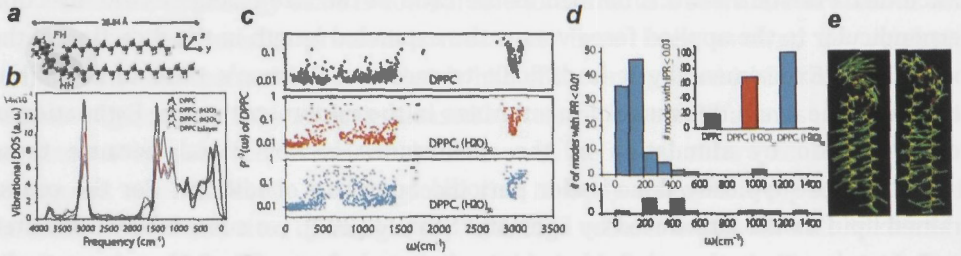


Figure 9.2: Molecular mechanism of heat transfer along membrane lipids. (a) Molecular structure of DPPC. The linear extension reduces to 28.24 Å at full hydration. (b) Vibrational density of states of DPPC as well as partially and fully hydrated DPPC. We find that water induces the emergence of low-frequency vibrational modes. (c) Inverse participation ratios for DPPC in the nonhydrated (black), partially hydrated (red), and fully hydrated (blue) DPPC. (d) Distribution of modes with participation of more than 100 DPPC atoms. No such modes exist for $\omega > 1500$ cm^{-1}. The color coding is the same as in panel a. (e) On the left, the lipid part of the eigenmode at 33.02 cm^{-1} is shown, which is important in energy transfer along fully hydrated DPPC. On the right, a breathing mode at 79.69 cm^{-1} is displayed, which carries thermal energy efficiently within the plane of the membrane.

To understand these biophysical processes, we decided to characterize the thermal and vibrational properties of the lipid bilayer as a function of hydration level and composition (Fig 9.2). The thermal conductivity of the lipid membrane is not known and is expected to be very different from that of a random mixture of lipids. Molecular alignment and the interfacial water are known to critically affect the resulting thermal conductivity of molecular systems. The energy transfer dynamics of a lipid monolayer interfacing aqueous environment have been studied with time-resolved surface sum-frequency generation [209]. Energy transfer along the lipid bilayer remains to be investigated. Another challenge in the field is the molecular-scale assignment of the mechanism behind the vibrational (thermal) energy transfer. New developments in density functional theory allow one to perform ab initio calculation of vibrational properties of large biomolecular systems. Using this approach, we have shown that membrane can act as an efficient medium for the energy transfer due to water induced extended vibrational modes of the lipid-water system [210].

Another interesting process which can be thermally modulated is the electrical conduction in the plane of the membrane. Enhanced proton conduction along phospholipid-water interfaces was first observed in the mid 1980s and was

recently confirmed by more sophisticated means such as scanning tunneling microscopy (STM). This conductivity is believed to be of functional importance because lateral proton diffusion along membrane surfaces represents the most efficient pathway for H+ transport between protein pumps. The molecular mechanism underlying the high lateral proton conductivity has not yet been resolved. We addressed this problem by studying hydrated phosphatidylcholine and phosphatidylethanolamine membranes. We showed that the electric field inhomogeneity at the lipid interface is such that auto-ionization of water is promoted and inter oxygen distances are reduced leading to astonishingly increased probability of water ionization even at low temperature which is further increased as the temperature rises to body temperature.

In summary, membrane sense and respond to heat and mechanical stress. Statics and dynamics of membrane is profoundly affected in this process and in some cases upon stress, membrane sends signals to downstream stress response systems in the cytoplasm.

9.2. Protein networks: pharmacological stress and the influence of thermal variations

A central tenet of modern biology is that the function of a protein depends critically on its three-dimensional structure. For instance enzymes recognize their substrate via shape matching as described by Emil Fischer's lock-and-key model. In many cases the substrate is itself a protein, which means that to have proper functionality proteins need to find each other and interact. Today we know that in many cases, a well defined three dimensional structure is not essential for functionality, while inter-protein contacts are. Only a small fraction of proteins are completely isolated in protein-protein interaction networks [211] and do not have any known interactions to other proteins, while many are not structured [212].

As discussed in previous chapters, thermal stress can influence intra-chain interactions leading to unfolding of proteins. It can also influence inter protein interactions and affect binding constants and reaction rates. Protein networks are also exposed to other "bad" stresses such as toxins as well as "good" stresses such as pharmacological manipulations. Variations in body temperature set points and pH can also influence the magnitude and nature of network response to medications.

We have systematically analyzed the dose-response relation of basic structural motifs of signaling pathways (protein networks) and found that fundamental constraints are imposed by the topology and the dynamics of the motifs

on the selectiveness and toxicity of any drug independent of its chemical nature and selectivity.

An example of a 5-node network topology is presented in Fig 9.3 with two inputs S1 and S2 and two outputs X1 and X2. This network both has a shared node (V) and an incoherent feed-forward loop and -- for specific parameters -- this network can multiplex two binary signals. A closer look at this topology reveals a non-monotonic relation between S1 and X2 and between V and X2, while monotonic relation exists between S1 and X1, and S2 and X2. We studied the effects of these non-monotonic relations on the effectiveness and toxicity of drugs. Further, we studied the effectiveness and toxicity of the drug with respect to extrinsic drug concentration fluctuations (Fig 9.3).

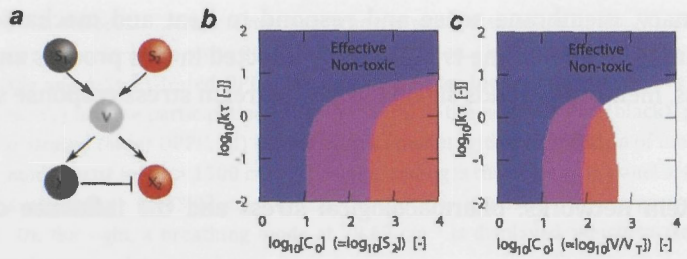


Figure 9.3: Contour plots for the effectiveness and toxicity as function of average dose C_0 and administration parameter $K = k\tau$ for the 5-note system (shown in panel a) that is (b) initially in state S1 = ON and S2 = OFF with an agonist for S2 or (c) initially in state S1 = OFF and S2 = OFF with an agonist for V. k denotes the dilution rate of the drug after administration and τ is the dose-interval. Monod-Wyman-Changeux model [213] was used to describe the network dynamics.

For a typical clinical treatment, we imagine that the average agonist concentration equals the required effective concentration C0. However, for a non-continuous uptake and some type of degradation/dilution of agonist over time, this implies that the actual agonist concentration varies as a function of time. However, it is not the average dose C0, but the actual dose C (t) that determines the toxicity and effectiveness of the drug. Indeed, for such a delivery process we need to specify toxicity and effectiviness in more detail. A drug is effective, if both for the minimal and maximal dose of the drug (thus for the full dose-range) the required effectiveness is observed (e.g. X2 from OFF to ON). A toxic drug indicates that within the dose-interval (at least at the minimal or at the maximal drug dose) a toxic effect is observed.

In Fig 9.3 we show the dependence of toxicity and effectiviness on the required dose C0, the time between administrations and the dilution rate k. These results clearly show that it is possible that a drug for which the required dose C0 is non-toxic, does have intervals of toxicity, depending on the administration procedure.

9.3. von Willebrand Factor under shear stress: the influence of metal ion

In response to vascular lesions, platelets are stimulated to secure bleeding arrest by formation of a platelet plug. In the high-shear environment found in the microvasculature or stenosed arteries, the adhesion of platelets critically depends on interactions with the plasma glycoprotein von Willebrand factor (VWF). This factor recruits platelets to sites of vascular damage by interacting with collagen and the platelet receptor GPIbα.

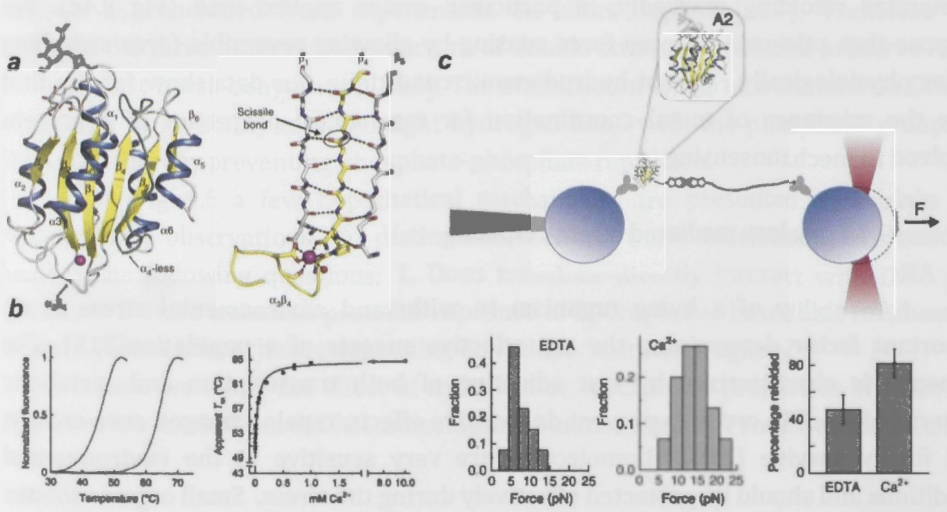


Figure 9.4: Calcium stabilizes the von Willebrand factor A2 domain and promotes its refolding. (a) Crystal structures of calcium-bound wt-A2 (left) and backbone representation (left) of the β4-scissile strand (yellow) and neighbouring strands illustrating how calcium interlocks β4 with β1 via coordination by Asp1498 and Asn1602. (b) Thermofluor stability assays were performed with monomeric A2 in buffer without (blue) or with (red) 1 mM CaCl2. Curves were normalized to maximum fluorescence signal. Data show representative curves of triplicate experiments. Calcium binds A2 with high affinity and stabilizes A2 in a concentration-dependent manner. Apparent melting temperatures were extracted from the mid-point of the unfolding transitions. Error bars represent s.d. (c) Single A2 domains (enlarged inset) are tethered between polystyrene beads that are held by a piezo-controlled micropipette and the optical trap. The tether includes a 2,500-bp DNA spacer to prevent unspecific bead–bead interactions. Presented below are the force distribution histograms for unfolding events with and without calcium obtained at a loading rate of 5 pN s−1. Mean unfolding forces are 7±3 pN (n=13) and 14±4 pN (n=15), respectively. Probability to refold to the native state at zero load significantly increases in the presence of calcium ($P<0.03$). T-bars represent s.d. from counting statistics.

von Willebrand factor (VWF) multimers mediate primary adhesion and aggregation of platelets. VWF potency critically depends on multimer size, which is

regulated by a feedback mechanism involving shear-induced unfolding of the VWF-A2 domain and cleavage by the metalloprotease ADAMTS-13. We performed crystallographic and single-molecule optical tweezers data studies on VWF-A2 providing mechanistic insight into calcium-mediated stabilization of the native conformation that protects A2 from cleavage by ADAMTS-13 [214]. The structure of A2 is presented in Fig 9.4a and residues involved in calcium binding are shown. We found that calcium binding markedly increases its thermodynamic stability (Fig 9.4b). Unfolding of A2 requires higher forces when calcium is present. Calcium further accelerates refolding markedly, in particular, under applied load (Fig 9.4c). We propose that calcium improves force sensing by allowing reversible force switching under physiologically relevant hydrodynamic conditions. Our data show for the first time the relevance of metal coordination for mechanical properties of a protein involved in mechanosensing.

9.4 Trehalose mediated stress-response

An ability of a living organism to withstand environmental stress is an important factor determining the reproductive success of a population[215]. The response is characterized by fast adjusting of both transcription and metabolic patterns of a cell in order to prevent destructive effects, repair damages, save energy, and finally survive [216]. Biomolecules are very sensitive to the environmental conditions and should be protected effectively during the stress. Small organic solutes such as sugars, amino acids etc. are shown to stabilize biomolecules effectively both *in vitro* and *in vivo* experiments under certain conditions [217, 218].

The disaccharide trehalose is a metabolite found in a wide range of organisms representing different domains of cellular life. It is known to be synthesized in many of those upon various stresses [219] and it is thought to protect proteins and cellular membranes from loss of activity caused by heat, cold, desiccation, dehydration, and oxidation[220].

It is shown that trehalose prevents denaturation of proteins and stabilizes proteins at various pH values (from 2.5 to 7). Melting temperature and melting free energy increase linearly with the increasing trehalose concentration (1 to 2 M)[221]. Trehalose is shown to stabilize compact intermediates thereby decelerating protein refolding, this effect increases with concentration (0.25M to 1M)[222]. Destabilization effect is also reported for trehalose. Bromelain is shown to be destabilized by the presence 1 M of trehalose[223].

Trehalose has been shown to decelerate polyglutamine-mediated protein aggregation and the resultant pathogenesis by stabilizing an aggregation-prone model protein[224]. Trehalose (0.5 M) prevents aggregation and inhibits reactivation of denatured luciferase *in vitro*[225].

In addition to proteins, trehalose influences the physico-chemical properties of nucleic acids. Destabilization of DNA by trehalose is reported in solution based on reduced melting temperature of double stranded DNA[226]. The extent of DNA destabilization due to trehalose depends on DNA length and depends on percent GC content as concluded from experiments on short duplexes[227]. Trehalose can stabilize the three dimensional structure of double stranded DNA and preserve its B-form against harsh dehydration[228]. The stabilization effect of trehalose on DNA is attributed to its ability to make tight hydrogen bond with the phosphate groups of DNA, and thereby preventing phosphate-phosphate repulsion.

In Fig 9.5 a few hypothetical mechanisms are presented to explain the experimental observations. To distinguish between these machnisms, we need to answer the following questions: 1. Does trehalose directly interact with DNA and proteins? 2. Does trehalose promote exposure of hydrophobic bases and residues? 3. Does trehalose effect change depending on the concentration regime?

Here we study the effect of trehalose on mechanical properties of DNA and Proteins. We use 920 nm dsDNA molecule and Maltose Binding Protein as models. In

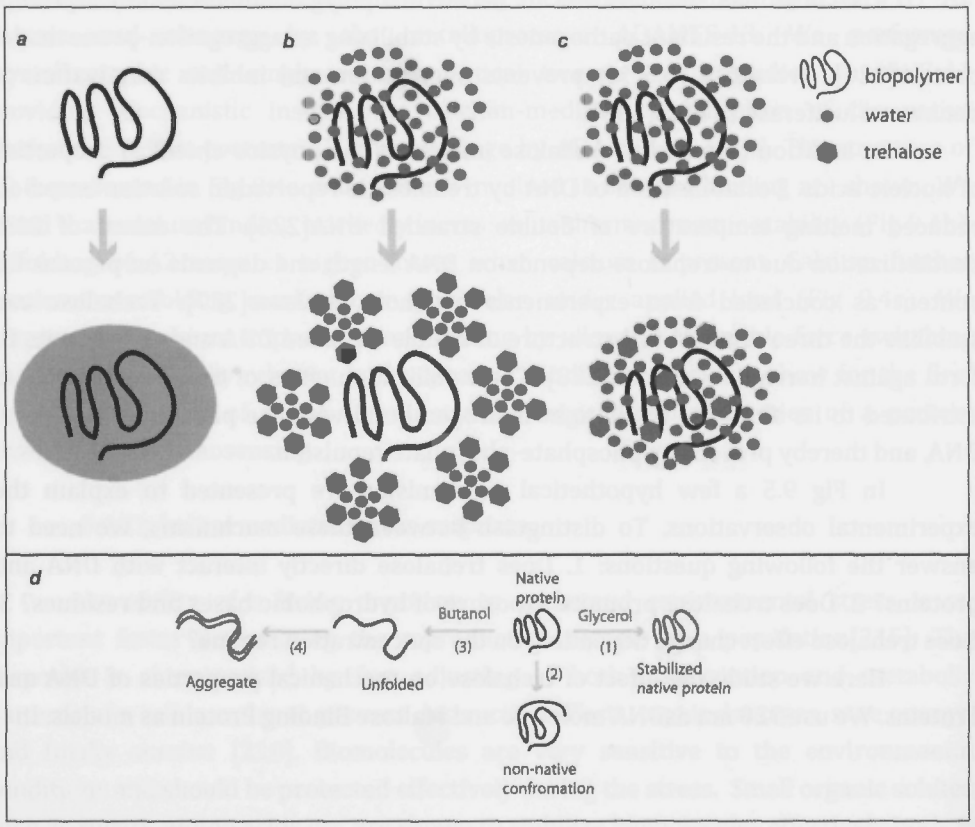


Figure 9.5: Possible mechanisms for trehalose induced stabilization and destabilization of proteins and nucleic acids (a) Vitrification theory assumes that trehalose forms a glassy matrix that acts as a cocoon and presumably physically shields the protein or indeed cells from abiotic stresses [219]. (b) Preferential exclusion theory proposes that there is no direct interaction between trehalose and protein (or biomolecule). Instead, as can be seen, addition of trehalose to bulk water sequesters water molecules away from the protein, decreasing its hydrated radius and increasing its compactness and consequently stability. (c) Water replacement theory talks of substitution of water molecules by trehalose-forming hydrogen bonds, maintaining the three-dimensional structure and stabilizing biomolecules[219]. (d) Possible conformational reactions that can be influenced by addition of trehalose to aqueous solution of proteins. While addition of alcohols like butanol leads to denaturation, glycerol, ethylene glycol and formamide stabilize the protein [229]. The effect expectedly depends on the concentration of the alcohol.

In the following section, we discuss DNA stretching experiment and the overstretching transition in the absence of trehalose. In the next section, similar experiments on DNA in the presence of trehalose are discussed.

Here we study the effect of trehalose on mechanical properties of DNA. We use 920 nm dsDNA as a model DNA molecule. In the following section, we discuss DNA stretching experiment and the overstretching transition in the absence of trehalose. In the next section, similar experiments on DNA in the presence of trehalose are discussed.

9.4.1 Single molecule mechanics of DNA

We used the optical tweezers setup (see chapter 2) to measure the response of double stranded DNA (2687 base pairs) to applied forces. In these experiments, the DNA was grabbed from its ends by immobilizing the end of one of the strands on each side, thus leaving the DNA torsionally unconstrained.

A typical force-extension curve is presented in Fig 9.6a, a typical waiting time between cycles is 5 s. In the stretching curve, when the force on the DNA increases, overstretching occurred at 65 pN with a narrow force distribution. These experiments on DNA provide us with additional findings: Firstly already at about 58-63 pN a stepwise overstretching behavior can be observed. Secondly, at 70% overstretching (about 1555-1580 nm) the force increased beyond 65 pN, marking the end of the overstretching transition. Thirdly, the relaxation curves, when the beads were brought close together again, exhibited significant hysteresis. This means that the reversal of overstretching is a non-equilibrium process. Forth, we observed that overstretching did not reverse completely; on average a residual of about 5 nm was present at forces of 5 pN. Until recently the mechanism of the overstretching transition and also the origin of the hysteresis are not completely understood.

The observed residual overstretching has, to our knowledge, not been previously reported. Closer inspection reveals that this residual overstretching is reversible. In measurements where we performed consecutive DNA overstretching experiments with intermediate waiting times on the order of minutes, consecutive rising curves overlapped also if the relaxation curves showed residual overstretching. A clear explanation for the observed residual overstretching is still missing.

9.4.2 DNA stretching in the presence of trehalose

The experiments were performed in three steps. First the DNA was stretched and relaxed multiple times in the absence of trehalose. Second, trehalose was infused. Third, DNA was stretched and relaxed multiple times. We found that the

overstretching force significantly reduces with the concentration of trehalose, and shows a significant difference already at 10 nM concentration (Fig 9.7a).

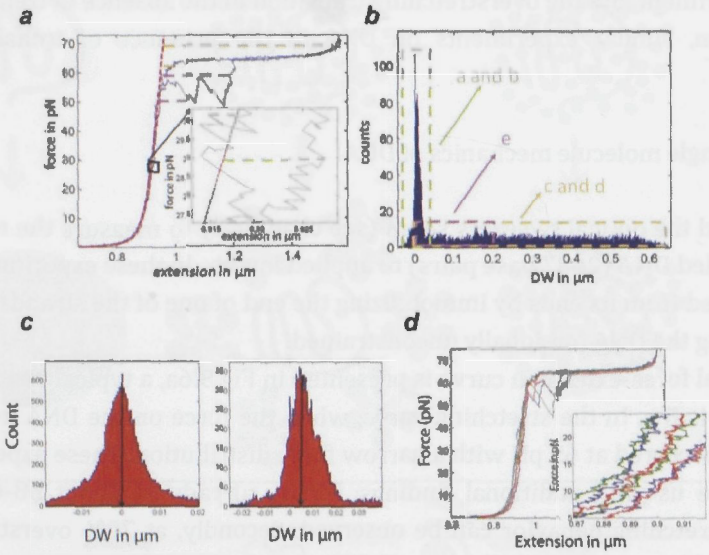


Figure 9.6: Single-molecule mechanics of DNA (a) Stretching and overstretching of a double stranded DNA. (b) distribution of DW length. Central peak (green box) are DW values close to the WLC model. The brown box marks the DNA overstretching transition. The purple box marks an area in which typically hysteresis occurs during relaxation.(c) The central peak in panel (b) is split for stretching (left) and relaxation (right). The shift in the peak position indicates the residual overstretching. (d) Three consecutive DNA overstretching experiments with intermediate waiting time (minutes) at zero force. In the insets it can be seen that residual overstretching reversed during the waiting time.

We propose a model for the observed destabilization of dsDNA by trehalose (Fig 9.7b). In this model addition of trehalose increases the hydrophobicity of the environment[230] by either sequestering water molecules or due to the interaction of its hydrocarbon backbone with the hydrophobic bases. Consequently melting of dsDNA will be facilitated. Because the hydrophobicity of guanine is much higher than adenine and pyrimidine bases (G>>A>T~C)[231], this model predicts that the effect of trehalose is more prominent for GC rich DNA molecules. We planned additional experiments and simulations to test this hypothesis and to reveal the role of direct nucleotide-trehalose interactions (e.g. H-bonds or van der waals interactions) in this process.

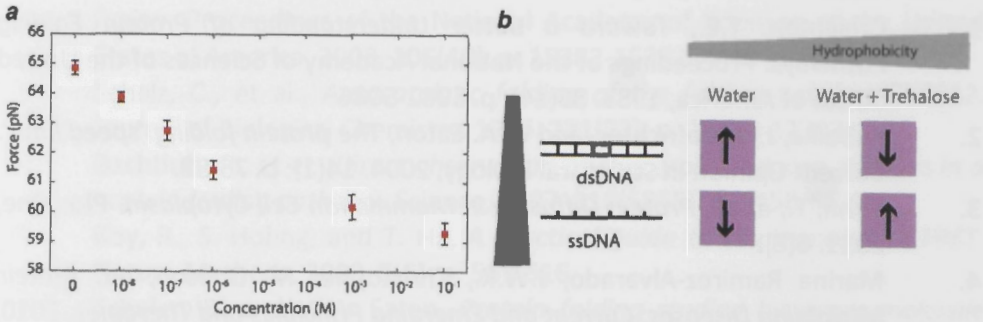


Figure 9.7: Single-molecule mechanics of DNA in the presence of trehalose. (a) Overstretching transition force as a function of trehalose concentration (b) Proposed model for trehalose induced destabilization of dsDNA. Single stranded DNA has exposed hydrophobic bases and has therefore tendency to form double stranded DNA in aqueous environment to shield the hydrophobic bases. Trehalose is hypothesized to promote exposure of hydrophobic bases by either sequestering water or by direct can de waals interaction with the bases. The latter may lead to a change in the structure of water hydrogen bond network in the vicinity of the single stranded DNA and change the entropic panlthy associated with hydration of hydrophobic bases. Note that the strength of a single hydrogen bond between a hydroxyl group and a given H-bond acceptor is probably higher for water O-H than trehalose O-H group due to the larger magnitude of the positive charge on the H atom in the former.

9.4.3 Outlook

Our preliminary data on DNA mechanics, shows a concentration dependent influence of trehalose. Further experiments are needed to shed light on the underlying molecular mechanism. One such experiment would involve comparison of the effect of 1-propanol, 1-hexanol and glycerol on DNA overstretching. While 1-hexanol and 1-propanol, have the same number of hydroxyl groups (H-bond formation site), they differ in their hydrophobicity. On the other side, the hydrocarbon backbones of 1-propanol and glycerol are identical, while glycerol has three times more hydroxyl groups. Moreover it is interesting to study the effect of trehalose on folding pathways of model proteins, their folding rates and stability. Finally it is interesting to investigate the interplay between trehalose and cellular chaperones.

References

1. Creighton, T.E., *Toward a Better Understanding of Protein Folding Pathways.* Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(14): p. 5082-5086.
2. Kubelka, J., J. Hofrichter, and W.A. Eaton, *The protein folding 'speed limit'.* Current Opinion in Structural Biology, 2004. **14**(1): p. 76-88.
3. Kuhn, T., et al., *Protein Diffusion in Mammalian Cell Cytoplasm.* Plos One, 2011. **6**(8).
4. Marina Ramirez-Alvarado, J.W.K., Christopher M. Dobson, ed. *Protein Misfolding Diseases: Current and Emerging Principles and Therapies.* 2010, Wiley. 1069.
5. Hartl, F.U. and M. Hayer-Hartl, *Protein folding - Molecular chaperones in the cytosol: from nascent chain to folded protein.* Science, 2002. **295**(5561): p. 1852-1858.
6. Hartl, F.U., A. Bracher, and M. Hayer-Hartl, *Molecular chaperones in protein folding and proteostasis.* Nature, 2011. **475**(7356): p. 324-332.
7. Krishna, M.M.G., Y. Lin, and S.W. Englander, *Protein misfolding: Optional barriers, misfolded intermediates, and pathway heterogeneity.* Journal of Molecular Biology, 2004. **343**(4): p. 1095-1109.
8. Deuerling, E., et al., *Trigger factor and DnaK cooperate in folding of newly synthesized proteins.* Nature, 1999. **400**(6745): p. 693-696.
9. Liu, S., et al., *Binding of the human Prp31 nop domain to a composite RNA-protein platform in U4 snRNP.* Science, 2007. **316**(5821): p. 115-120.
10. Feng, D. and G. Jin, *Introduction to condensed matter physics*. 2005, Hackensack, NJ: World Scientific. v. <1 >.
11. Floyd, D.L., S.C. Harrison, and A.M. van Oijen, *Analysis of Kinetic Intermediates in Single-Particle Dwell-Time Distributions.* Biophysical Journal, 2010. **99**(2): p. 360-366.
12. Schotte, F., et al., *Watching a protein as it functions with 150-ps time-resolved X-ray crystallography.* Science, 2003. **300**(5627): p. 1944-1947.
13. Kolano, C., et al., *Watching hydrogen-bond dynamics in a beta-turn by transient two-dimensional infrared spectroscopy.* Nature, 2006. **444**(7118): p. 469-472.
14. Jarzynski, C., *Nonequilibrium equality for free energy differences.* Physical Review Letters, 1997. **78**(14): p. 2690-2693.
15. Hummer, G. and A. Szabo, *Free energy reconstruction from nonequilibrium single-molecule pulling experiments.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(7): p. 3658-3661.

16. Floyd, D.L., et al., *Single-particle kinetics of influenza virus membrane fusion.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(40): p. 15382-15387.
17. Scholz, C., et al., *Autocatalytic folding of the folding catalyst FKBP12.* Journal of Biological Chemistry, 1996. **271**(22): p. 12703-12707.
18. Bechtluft, P., et al., *Direct observation of chaperone-induced changes in a protein folding pathway.* Science, 2007. **318**(5855): p. 1458-61.
19. Roy, R., S. Hohng, and T. Ha, *A practical guide to single-molecule FRET.* Nature Methods, 2008. **5**(6): p. 507-516.
20. Schuler, B. and W.A. Eaton, *Protein folding studied by single-molecule FRET.* Current Opinion in Structural Biology, 2008. **18**(1): p. 16-26.
21. Kellermayer, M.S.Z., et al., *Folding-unfolding transitions in single titin molecules characterized with laser tweezers.* Science, 1997. **276**(5315): p. 1112-1116.
22. Tskhovrebova, L., et al., *Elasticity and unfolding of single molecules of the giant muscle protein titin.* Nature, 1997. **387**(6630): p. 308-312.
23. Rief, M., et al., *Reversible unfolding of individual titin immunoglobulin domains by AFM.* Science, 1997. **276**(5315): p. 1109-1112.
24. Deniz, A.A., et al., *Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(10): p. 5179-5184.
25. Talaga, D.S., et al., *Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(24): p. 13021-13026.
26. Ferreon, A.C.M. and A.A. Deniz, *Protein folding at single-molecule resolution.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2011. **1814**(8): p. 1021-1029.
27. Jacob, E. and R. Unger, *A tale of two tails: why are terminal residues of proteins exposed?* Bioinformatics, 2007. **23**(2): p. E225-E230.
28. Sharma, S., et al., *Monitoring protein conformation along the pathway of chaperonin-assisted folding.* Cell, 2008. **133**(1): p. 142-153.
29. Chakraborty, K., et al., *Chaperonin-Catalyzed Rescue of Kinetically Trapped States in Protein Folding.* Cell, 2010. **142**(1): p. 112-122.
30. Hofmann, H., et al., *Single-molecule spectroscopy of protein folding in a chaperonin cage.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(26): p. 11793-11798.

31. Jewett, A.I. and J.E. Shea, *Reconciling theories of chaperonin accelerated folding with experimental evidence.* Cellular and Molecular Life Sciences, 2010. **67**(2): p. 255-276.
32. Little, W.C., et al., *Stretched Extracellular Matrix Proteins Turn Fouling and Are Functionally Rescued by the Chaperones Albumin and Casein.* Nano Letters, 2009. **9**(12): p. 4158-4167.
33. Mapa, K., et al., *The Conformational Dynamics of the Mitochondrial Hsp70 Chaperone.* Molecular Cell, 2010. **38**(1): p. 89-100.
34. Marcinowski, M., et al., *Substrate discrimination of the chaperone BiP by autonomous and cochaperone-regulated conformational transitions.* Nature Structural & Molecular Biology, 2011. **18**(2): p. 150-U210.
35. Schlecht, R., et al., *Mechanics of Hsp70 chaperones enables differential interaction with client proteins.* Nature Structural & Molecular Biology, 2011. **18**(3): p. 345-U135.
36. Ratzke, C., et al., *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, 2010. **107**(37): p. 16101-16106.
37. Wennmalm, S., L. Edman, and R. Rigler, *Conformational fluctuations in single DNA molecules.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(20): p. 10641-10646.
38. Edman, L., et al., *The fluctuating enzyme: a single molecule approach.* Chemical Physics, 1999. **247**(1): p. 11-22.
39. Neuweiler, H., C.M. Johnson, and A.R. Fersht, *Direct observation of ultrafast folding and denatured state dynamics in single protein molecules.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(44): p. 18569-18574.
40. Bocking, T., et al., *Single-molecule analysis of a molecular disassemblase reveals the mechanism of Hsc70-driven clathrin uncoating.* Nature Structural & Molecular Biology, 2011. **18**(3): p. 295-U78.
41. Horst, R., et al., *Direct NMR observation of a substrate protein bound to the chaperonin GroEL.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(36): p. 12748-12753.
42. Stan, G., et al., *Coupling between allosteric transitions in GroEL and assisted folding of a substrate protein.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(21): p. 8803-8808.

43. Aubin-Tam, M.E., et al., *Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine.* Cell, 2011. **145**(2): p. 257-267.
44. Maillard, R.A., et al., *ClpX(P) Generates Mechanical Force to Unfold and Translocate Its Protein Substrates.* Cell, 2011. **145**(3): p. 459-469.
45. Valle, F., et al., *AFM structural study of the molecular chaperone GroEL and its two-dimensional crystals: an ideal "living" calibration sample.* Ultramicroscopy, 2002. **93**(1): p. 83-89.
46. Mou, J.X., et al., *High resolution surface structure of E-coli GroES oligomer by atomic force microscopy.* Febs Letters, 1996. **381**(1-2): p. 161-164.
47. Mou, J.X., et al., *Chaperonins GroEL and GroES: Views from atomic force microscopy.* Biophysical Journal, 1996. **71**(4): p. 2213-2221.
48. Vinckier, A., et al., *Atomic force microscopy detects changes in the interaction forces between GroEL and substrate proteins.* Biophysical Journal, 1998. **74**(6): p. 3256-3263.
49. Pfister, G., et al., *Detection of HSP60 on the membrane surface of stressed human endothelial cells by atomic force and confocal microscopy.* Journal of Cell Science, 2005. **118**(8): p. 1587-1594.
50. Zhu, Y., et al., *Single Molecule Force Spectroscopy of the Cardiac Titin N2B Element EFFECTS OF THE MOLECULAR CHAPERONE alpha B-CRYSTALLIN WITH DISEASE-CAUSING MUTATIONS.* Journal of Biological Chemistry, 2009. **284**(20): p. 13914-13923.
51. Ahn, M., et al., *Chaperone-like activities of alpha-synuclein: alpha-Synuclein assists enzyme activities of esterases.* Biochemical and Biophysical Research Communications, 2006. **346**(4): p. 1142-1149.
52. Elcock, A.H., *Molecular Simulations of cotranslational protein folding: Fragment stabilities, folding cooperativity, and trapping in the ribosome.* Plos Computational Biology, 2006. **2**(7): p. 824-841.
53. Sliozberg, Y. and C.F. Abrams, *Spontaneous conformational changes in the e. coli GroEL subunit from all-atom molecular dynamics Simulations.* Biophysical Journal, 2007. **93**(6): p. 1906-1916.
54. Sliozberg, Y. and C.F. Abrams, *A simulation study of the effects of nucleotide binding on the vibrational character of the GroEL subunit protein.* Biophysical Journal, 2007: p. 380a-380a.
55. Sliozberg, Y.R., et al., *Computational and experimental investigation of morphology in thermoplastic elastomer gels composed of AB/ABA blends in B-selective solvent.* Soft Matter, 2011. **7**(16): p. 7539-7551.

56. Fan, H. and A.E. Mark, *Mimicking the action of GroEL in molecular dynamics simulations: Application to the refinement of protein structures.* Protein Science, 2006. **15**(3): p. 441-448.
57. Ahmad, A., et al., *Heat shock protein 70 kDa chaperone/DnaJ cochaperone complex employs an unusual dynamic interface.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(47): p. 18966-18971.
58. Golas, E., et al., *Simulation of the Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics.* Journal of Chemical Theory and Computation, 2012. **8**(5): p. 1750-1764.
59. Morra, G., et al., *Corresponding Functional Dynamics across the Hsp90 Chaperone Family: Insights from a Multiscale Analysis of MD Simulations.* Plos Computational Biology, 2012. **8**(3).
60. Dixit, A. and G.M. Verkhivker, *Probing Molecular Mechanisms of the Hsp90 Chaperone: Biophysical Modeling Identifies Key Regulators of Functional Dynamics.* Plos One, 2012. **7**(5).
61. McClellan, A.J., et al., *Protein quality control: chaperones culling corrupt conformations.* Nature Cell Biology, 2005. **7**(8): p. 736-741.
62. Xie, Y., T. Maxson, and Y. Tor, *Fluorescent Ribonucleoside as a FRET Acceptor for Tryptophan in Native Proteins.* Journal of the American Chemical Society, 2010. **132**(34): p. 11896-11897.
63. Fitter, J., et al., *Single molecule fluorescence spectroscopy: a tool for protein studies approaching cellular environmental conditions.* Soft Matter, 2011. **7**(4): p. 1254-1259.
64. Altieri, D.C., *Mitochondrial Compartmentalized Protein Folding and Tumor Cell Survival.* Oncotarget, 2011. **2**(4): p. 347-351.
65. Huang, B., et al., *Counting low-copy number proteins in a single cell.* Science, 2007. **315**(5808): p. 81-84.
66. Betzig, E., et al., *Imaging intracellular fluorescent proteins at nanometer resolution.* Science, 2006. **313**(5793): p. 1642-1645.
67. Cox, S., et al., *Bayesian localization microscopy reveals nanoscale podosome dynamics.* Nature Methods, 2012. **9**(2): p. 195-200.
68. Burz, D.S. and A. Shekhtman, *Structural Biology: Inside the Living Cell.* Nature, 2009. **458**(7234): p. 37-38.
69. Sakakibara, D., et al., *Protein structure determination in living cells by in-cell NMR spectroscopy.* Nature, 2009. **458**(7234): p. 102-U10.
70. Svoboda, K. and S.M. Block, *Biological Applications of Optical Forces.* Annual Review of Biophysics and Biomolecular Structure, 1994. **23**: p. 247-285.

71. Tolic-Norrelykke, I.M., K. Berg-Sorensen, and H. Flyvbjerg, *MatLab program for precision calibration of optical tweezers.* Computer Physics Communications, 2004. **159**(3): p. 225-240.
72. Tolic-Norrelykke, S.F., et al., *Calibration of optical tweezers with positional detection in the back focal plane.* Review of Scientific Instruments, 2006. **77**(10).
73. Smith, S.B., Y.J. Cui, and C. Bustamante, *Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules.* Science, 1996. **271**(5250): p. 795-799.
74. Wang, M.D., et al., *Stretching DNA with optical tweezers.* Biophysical Journal, 1997. **72**(3): p. 1335-1346.
75. Czerwinski, F., A.C. Richardson, and L.B. Oddershede, *Quantifying Noise in Optical Tweezers by Allan Variance.* Optics Express, 2009. **17**(15): p. 13255-13269.
76. J. Freudenberger, S.S., and B. Venditti, in *17th European Signal Processing Conference (EUSIPCO).* 2009.
77. Mantooth, B.A., et al., *Cross-correlation image tracking for drift correction and adsorbate analysis.* Review of Scientific Instruments, 2002. **73**(2): p. 313-317.
78. King, G.M., et al., *Ultrastable Atomic Force Microscopy: Atomic-Scale Stability and Registration in Ambient Conditions.* Nano Letters, 2009. **9**(4): p. 1451-1456.
79. Artursson, T., et al., *Drift correction for gas sensors using multivariate methods.* Journal of Chemometrics, 2000. **14**(5-6): p. 711-723.
80. Holmberg, M., et al., *Drift counteraction for an electronic nose.* Sensors and Actuators B-Chemical, 1996. **36**(1-3): p. 528-535.
81. Barhoumi, A., et al., *Light-induced release of DNA from plasmon-resonant nanoparticles: Towards light-controlled gene therapy.* Chemical Physics Letters, 2009. **482**(4-6): p. 171-179.
82. Wong, L.S., F. Khan, and J. Micklefield, *Selective Covalent Protein Immobilization: Strategies and Applications.* Chemical Reviews, 2009. **109**(9): p. 4025-4053.
83. Gunnarsson, A., et al., *Kinetic and thermodynamic characterization of single-mismatch discrimination using single-molecule imaging.* Nucleic Acids Research, 2009. **37**(14).
84. Lin, C.X., et al., *Functional DNA nanotube arrays: Bottom-up meets top-down.* Angewandte Chemie-International Edition, 2007. **46**(32): p. 6089-6092.

85. Kam, N.W.S., et al., *Nanotube molecular transporters: Internalization of carbon nanotube-protein conjugates into mammalian cells.* Journal of the American Chemical Society, 2004. **126**(22): p. 6850-6851.
86. Leunissen, M.E., et al., *Switchable self-protected attractions in DNA-functionalized colloids.* Nature Materials, 2009. **8**(7): p. 590-595.
87. Valignat, M.P., et al., *Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(12): p. 4225-4229.
88. Maye, M.M., et al., *Switching binary states of nanoparticle superlattices and dimer clusters by DNA strands.* Nature Nanotechnology, 2010. **5**(2): p. 116-120.
89. Braun, E., et al., *DNA-templated assembly and electrode attachment of a conducting silver wire.* Nature, 1998. **391**(6669): p. 775-778.
90. Patolsky, F., Y. Weizmann, and I. Willner, *Actin-based metallic nanowires as bio-nanotransporters.* Nature Materials, 2004. **3**(10): p. 692-695.
91. van den Broek, B., M.C. Noom, and G.J.L. Wuite, *DNA-tension dependence of restriction enzyme activity reveals mechanochemical properties of the reaction pathway.* Nucleic Acids Research, 2005. **33**(8): p. 2676-2684.
92. van Mameren, J., E.J.G. Peterman, and G.J.L. Wuite, *See me, feel me: methods to concurrently visualize and manipulate single DNA molecules and associated proteins.* Nucleic Acids Research, 2008. **36**(13): p. 4381-4389.
93. Cecconi, C., et al., *Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers.* European Biophysics Journal with Biophysics Letters, 2008. **37**(6): p. 729-738.
94. Chivers, C.E., et al., *A streptavidin variant with slower biotin dissociation and increased mechanostability.* Nature Methods, 2010. **7**(5): p. 391-U76.
95. Holmberg, A., et al., *The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures.* Electrophoresis, 2005. **26**(3): p. 501-510.
96. Waner, M.J., et al., *Thermal and sodium dodecylsulfate induced transitions of streptavidin.* Biophysical Journal, 2004. **87**(4): p. 2701-2713.
97. Marttila, A.T., et al., *Recombinant NeutraLite Avidin: a non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties.* Febs Letters, 2000. **467**(1): p. 31-36.
98. Schmidt, T.G.M. and A. Skerra, *The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins.* Nature Protocols, 2007. **2**(6): p. 1528-1535.

99. Schmidt, T.G.M., et al., *Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin.* Journal of Molecular Biology, 1996. **255**(5): p. 753-766.
100. Busby, M., et al., *Optimisation of a multivalent Strep tag for protein detection.* Biophysical Chemistry, 2010. **152**(1-3): p. 170-177.
101. Kim, M., et al., *A Nanoscale Force Probe for Gauging Intermolecular Interactions.* Angewandte Chemie-International Edition, 2012. **51**(8): p. 1903-1906.
102. Tang, J.L., et al., *Recognition Imaging and Highly Ordered Molecular Templating of Bacterial S-Layer Nanoarrays Containing Affinity-Tags.* Nano Letters, 2008. **8**(12): p. 4312-4319.
103. Afonina, I., et al., *Efficient priming of PCR with short oligonucleotides conjugated to a minor groove binder.* Nucleic Acids Research, 1997. **25**(13): p. 2657-2660.
104. Medintz, I.L. and J.R. Deschamps, *Maltose-binding protein: a versatile platform for prototyping biosensing.* Current Opinion in Biotechnology, 2006. **17**(1): p. 17-27.
105. Wickner, W., A.J.M. Driessen, and F.U. Hartl, *The Enzymology of Protein Translocation across the Escherichia-Coli Plasma-Membrane.* Annual Review of Biochemistry, 1991. **60**: p. 101-124.
106. Taylor, W.R., *Protein knots and fold complexity: Some new twists.* Computational Biology and Chemistry, 2007. **31**(3): p. 151-162.
107. Tezuka, Y. and H. Oike, *Topological polymer chemistry: Systematic classification of nonlinear polymer topologies.* Journal of the American Chemical Society, 2001. **123**(47): p. 11570-11576.
108. Rosa, A., *Topological jamming of spontaneously knotted polyelectrolyte chains driven through a nanopore.* Physical Review Letters, 2012.
109. Sumners, D.W. and S.G. Whittington, *Knots in Self-Avoiding Walks.* Journal of Physics a-Mathematical and General, 1988. **21**(7): p. 1689-1694.
110. Virnau, P., L.A. Mirny, and M. Kardar, *Intricate knots in proteins: Function and evolution.* Plos Computational Biology, 2006. **2**(9): p. 1074-1079.
111. Wallin, S., K.B. Zeldovich, and E.I. Shakhnovich, *The folding mechanics of a knotted protein.* Journal of Molecular Biology, 2007. **368**(3): p. 884-893.
112. Onuchic, J.N. and P.G. Wolynes, *Theory of protein folding.* Current Opinion in Structural Biology, 2004. **14**(1): p. 70-75.
113. Sali, A., E. Shakhnovich, and M. Karplus, *How Does a Protein Fold.* Nature, 1994. **369**(6477): p. 248-251.

114. Dokholyan, N.V., et al., *Topological determinants of protein folding.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(13): p. 8637-8641.
115. Ghosh, A., K.V. Brinda, and S. Vishveshwara, *Dynamics of lysozyme structure network: Probing the process of unfolding.* Biophysical Journal, 2007. **92**(7): p. 2523-2535.
116. Gromiha, M.M., A.M. Thangakani, and S. Selvaraj, *FOLD-RATE: prediction of protein folding rates from amino acid sequence.* Nucleic Acids Research, 2006. **34**: p. W70-W74.
117. Shank, E.A., et al., *The folding cooperativity of a protein is controlled by its chain topology.* Nature, 2010. **465**(7298): p. 637-U134.
118. Taylor, W.R. and K. Lin, *Protein knots - A tangled problem.* Nature, 2003. **421**(6918): p. 25-25.
119. Onuchic, J.N., Z. LutheySchulten, and P.G. Wolynes, *Theory of protein folding: The energy landscape perspective.* Annual Review of Physical Chemistry, 1997. **48**: p. 545-600.
120. Dobson, C.M., A. Sali, and M. Karplus, *Protein folding: A perspective from theory and experiment.* Angewandte Chemie-International Edition, 1998. **37**(7): p. 868-893.
121. Fersht, A.R., *Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(4): p. 1525-1529.
122. Vendruscolo, M., et al., *Three key residues form a critical contact network in a protein folding transition state.* Nature, 2001. **409**(6820): p. 641-645.
123. Roder, H. and W. Colon, *Kinetic role of early intermediates in protein folding.* Current Opinion in Structural Biology, 1997. **7**(1): p. 15-28.
124. Sanchez, I.E. and T. Kiefhaber, *Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding.* Journal of Molecular Biology, 2003. **325**(2): p. 367-376.
125. Khan, F., et al., *The kinetic pathway of folding of barnase.* Journal of Molecular Biology, 2003. **333**(1): p. 169-186.
126. Dobson, C.M., *Protein folding and misfolding.* Nature, 2003. **426**(6968): p. 884-90.
127. Brockwell, D.J. and S.E. Radford, *Intermediates: ubiquitous species on folding energy landscapes?* Current Opinion in Structural Biology, 2007. **17**(1): p. 30-37.
128. Matouschek, A., et al., *Transient Folding Intermediates Characterized by Protein Engineering.* Nature, 1990. **346**(6283): p. 440-445.

129. Capaldi, A.P., C. Kleanthous, and S.E. Radford, *Im7 folding mechanism: misfolding on a path to the native state.* Nature Structural Biology, 2002. **9**(3): p. 209-216.
130. Friel, C.T., et al., *The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints.* Nature Structural & Molecular Biology, 2009. **16**(3): p. 318-324.
131. Ivarsson, Y., et al., *Engineered symmetric connectivity of secondary structure elements highlights malleability of protein folding pathways.* Journal of the American Chemical Society, 2009. **131**(33): p. 11727-33.
132. Korzhnev, D.M., et al., *A transient and low-populated protein-folding intermediate at atomic resolution.* Science, 2010. **329**(5997): p. 1312-6.
133. Religa, T.L., et al., *Solution structure of a protein denatured state and folding intermediate.* Nature, 2005. **437**(7061): p. 1053-6.
134. Ivarsson, Y., et al., *The folding pathway of an engineered circularly permuted PDZ domain.* Protein Eng Des Sel, 2008. **21**(3): p. 155-60.
135. Bollen, Y.J.M., M.B. Kamphuis, and C.P.M. van Mierlo, *The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(11): p. 4095-4100.
136. Sharma, S.K., et al., *The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase.* Nature Chemical Biology, 2010. **6**(12): p. 914-920.
137. Gianni, S., et al., *Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain.* Nature Structural & Molecular Biology, 2010. **17**(12): p. 1431-U57.
138. Tanaka, N., et al., *The substrate binding domain of DnaK facilitates slow protein refolding.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(24): p. 15398-15403.
139. Agashe, V.R., et al., *Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed.* Cell, 2004. **117**(2): p. 199-209.
140. Hartl, F.U., A. Bracher, and M. Hayer-Hartl, *Molecular chaperones in protein folding and proteostasis.* Nature, 2011. **475**(7356): p. 324-32.
141. Kramer, G., et al., *The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins.* Nature Structural & Molecular Biology, 2009. **16**(6): p. 589-597.

142. Sharma, S.K., et al., *The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase.* Nat Chem Biol, 2010. **6**(12): p. 914-20.
143. Bartlett, A.I. and S.E. Radford, *An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms.* Nat Struct Mol Biol, 2009. **16**(6): p. 582-8.
144. Hoffmann, A., B. Bukau, and G. Kramer, *Structure and function of the molecular chaperone Trigger Factor.* Biochim Biophys Acta, 2010. **1803**(6): p. 650-61.
145. Wegrzyn, R.D. and E. Deuerling, *Molecular guardians for newborn proteins: ribosome-associated chaperones and their role in protein folding.* Cell Mol Life Sci, 2005. **62**(23): p. 2727-38.
146. Merz, F., et al., *Molecular mechanism and structure of Trigger Factor bound to the translating ribosome.* Embo Journal, 2008. **27**(11): p. 1622-1632.
147. Ferbitz, L., et al., *Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins.* Nature, 2004. **431**(7008): p. 590-6.
148. Kramer, G., et al., *L23 protein functions as a chaperone docking site on the ribosome.* Nature, 2002. **419**(6903): p. 171-174.
149. Hoffmann, A., et al., *Trigger factor forms a protective shield for nascent polypeptides at the ribosome.* J Biol Chem, 2006. **281**(10): p. 6539-45.
150. Teter, S.A., et al., *Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains.* Cell, 1999. **97**(6): p. 755-765.
151. Hoffmann, A., et al., *Concerted Action of the Ribosome and the Associated Chaperone Trigger Factor Confines Nascent Polypeptide Folding.* Mol Cell, 2012.
152. Martinez-Hackert, E. and W.A. Hendrickson, *Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone.* Cell, 2009. **138**(5): p. 923-34.
153. Borgia, A., P.M. Williams, and J. Clarke, *Single-molecule studies of protein folding.* Annu Rev Biochem, 2008. **77**: p. 101-25.
154. Fernandez, J.M. and H. Li, *Force-clamp spectroscopy monitors the folding trajectory of a single protein.* Science, 2004. **303**(5664): p. 1674-8.
155. Junker, J.P., F. Ziegler, and M. Rief, *Ligand-dependent equilibrium fluctuations of single calmodulin molecules.* Science, 2009. **323**(5914): p. 633-7.
156. Jakobi, A.J., et al., *Calcium modulates force sensing by the von Willebrand factor A2 domain.* Nat Commun, 2011. **2**: p. 385.

157. Bertz, M. and M. Rief, *Mechanical unfoldons as building blocks of maltose-binding protein.* J Mol Biol, 2008. **378**(2): p. 447-58.
158. Marko, J.F. and E.D. Siggia, *Stretching DNA.* Macromolecules, 1995. **28**(26): p. 8759-8770.
159. Maier, R., C. Scholz, and F.X. Schmid, *Dynamic association of trigger factor with protein substrates.* J Mol Biol, 2001. **314**(5): p. 1181-90.
160. Rutkowska, A., et al., *Dynamics of trigger factor interaction with translating ribosomes.* J Biol Chem, 2008. **283**(7): p. 4124-32.
161. Kaiser, C.M., et al., *Real-time observation of trigger factor function on translating ribosomes.* Nature, 2006. **444**(7118): p. 455-60.
162. Lakshmipathy, S.K., et al., *Versatility of trigger factor interactions with ribosome-nascent chain complexes.* J Biol Chem, 2010. **285**(36): p. 27911-23.
163. Park, S., T.E. Klein, and V.S. Pande, *Folding and misfolding of the collagen triple helix: Markov analysis of molecular dynamics simulations.* Biophysical Journal, 2007. **93**(12): p. 4108-15.
164. Jaenicke, R., *Stability and folding of domain proteins.* Prog Biophys Mol Biol, 1999. **71**(2): p. 155-241.
165. Thomas, P.D. and K.A. Dill, *An iterative method for extracting energy-like quantities from protein structures.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(21): p. 11628-11633.
166. Calloni, G., et al., *DnaK Functions as a Central Hub in the E.coli Chaperone Network.* Cell reports, 2012. **1**(3): p. 251-64.
167. Rodriguez, F., et al., *Molecular basis for regulation of the heat shock transcription factor sigma32 by the DnaK and DnaJ chaperones.* Molecular Cell, 2008. **32**(3): p. 347-58.
168. Acebron, S.P., et al., *DnaJ recruits DnaK to protein aggregates.* Journal of Biological Chemistry, 2008. **283**(3): p. 1381-90.
169. Ben-Zvi, A., et al., *Active solubilization and refolding of stable protein aggregates by cooperative unfolding action of individual hsp70 chaperones.* Journal of Biological Chemistry, 2004. **279**(36): p. 37298-303.
170. Rudiger, S., et al., *Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries.* Embo Journal, 1997. **16**(7): p. 1501-1507.
171. Zhu, X.T., et al., *Structural analysis of substrate binding by the molecular chaperone DnaK.* Science, 1996. **272**(5268): p. 1606-1614.
172. Kim, S.Y., et al., *Interaction of the DnaK and DnaJ chaperone system with a native substrate, p1 RepA.* Journal of Biological Chemistry, 2002. **277**(47): p. 44778-44783.

173. Wawrzynow, A., et al., *Atp Hydrolysis Is Required for the Dnaj-Dependent Activation of Dnak Chaperone for Binding to Both Native and Denatured Protein Substrates.* Journal of Biological Chemistry, 1995. **270**(33): p. 19307-19311.
174. Wawrzynow, A. and M. Zylicz, *Divergent Effects of Atp on the Binding of the Dnak and Dnaj Chaperones to Each Other, or to Their Various Native and Denatured Protein Substrates.* Journal of Biological Chemistry, 1995. **270**(33): p. 19300-19306.
175. Pishvaee, B., et al., *A yeast DNA J protein required for uncoating of clathrin-coated vesicles in vivo.* Nature Cell Biology, 2000. **2**(12): p. 958-963.
176. Bechtluft, P., et al., *SecB-A chaperone dedicated to protein translocation.* Molecular Biosystems, 2010. **6**(4): p. 620-627.
177. Szabo, A., et al., *The Atp Hydrolysis-Dependent Reaction Cycle of the Escherichia-Coli Hsp70 System - Dnak, Dnaj, and Grpe.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(22): p. 10345-10349.
178. Mayer, M.P. and B. Bukau, *Hsp70 chaperones: Cellular functions and molecular mechanism.* Cellular and Molecular Life Sciences, 2005. **62**(6): p. 670-684.
179. Bin Hu, M.P.M., Masaru Tomita., *Hsp70-Mediated Protein Refolding in E-Cell*, in *E-Cell System: Basic Concepts and Applications*. 2012.
180. Findly, R.C., R.J. Gillies, and R.G. Shulman, *In vivo phosphorus-31 nuclear magnetic resonance reveals lowered ATP during heat shock of Tetrahymena.* Science, 1983. **219**(4589): p. 1223-5.
181. Banecki, B. and M. Zylicz, *Real time kinetics of the DnaK/DnaJ/GrpE molecular chaperone machine action.* Journal of Biological Chemistry, 1996. **271**(11): p. 6137-43.
182. Soini, J., et al., *Transient increase of ATP as a response to temperature up-shift in Escherichia coli.* Microbial Cell Factories, 2005. **4**.
183. Creighton, T.E., *Protein folding*. 1992, New York: W.H. Freeman and Co. xix, 547 p.
184. Nicastro, K.R., et al., *The role of gaping behaviour in habitat partitioning between coexisting intertidal mussels.* BMC ecology, 2010. **10**: p. 17-17.
185. Martin, J., et al., *Chaperonin-Mediated Protein Folding at the Surface of Groel through a Molten Globule-Like Intermediate.* Nature, 1991. **352**(6330): p. 36-42.

186. Grimshaw, J.P.A., et al., *Thermosensor action of GrpE - The DnaK chaperone system at heat shock temperatures.* Journal of Biological Chemistry, 2003. **278**(21): p. 19048-19053.
187. Taipale, M., D.F. Jarosz, and S. Lindquist, *HSP90 at the hub of protein homeostasis: emerging mechanistic insights.* Nature Reviews Molecular Cell Biology, 2010. **11**(7): p. 515-528.
188. Goetz, M.P., et al., *The Hsp90 chaperone complex as a novel target for cancer therapy.* Annals of Oncology, 2003. **14**(8): p. 1169-1176.
189. Krukenberg, K.A., et al., *Conformational dynamics of the molecular chaperone Hsp90.* Quarterly Reviews of Biophysics, 2011. **44**(2): p. 229-255.
190. Krukenberg, K.A., et al., *Multiple conformations of E-coli Hsp90 in solution: Insights into the conformational dynamics of Hsp90.* Structure, 2008. **16**(5): p. 755-765.
191. Shiau, A.K., et al., *Structural analysis of E-coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements.* Cell, 2006. **127**(2): p. 329-340.
192. Graf, C., et al., *Spatially and kinetically resolved changes in the conformational dynamics of the Hsp90 chaperone machine.* Embo Journal, 2009. **28**(5): p. 602-613.
193. Sato, T., et al., *HtpG, the prokaryotic homologue of Hsp90, stabilizes a phycobilisome protein in the cyanobacterium Synechococcus elongatus PCC 7942.* Molecular Microbiology, 2010. **76**(3): p. 576-589.
194. Vaughan, C.K., et al., *Structure of an Hsp90-Cdc37-Cdk4 complex.* Molecular Cell, 2006. **23**(5): p. 697-707.
195. Young, J.C., C. Schneider, and F.U. Hartl, *In vitro evidence that hsp90 contains two independent chaperone sites.* Febs Letters, 1997. **418**(1-2): p. 139-143.
196. Ebong, I.O., et al., *Heterogeneity and dynamics in the assembly of the Heat Shock Protein 90 chaperone complexes.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(44): p. 17939-17944.
197. Falsone, S.F., et al., *The Molecular Chaperone Hsp90 Modulates Intermediate Steps of Amyloid Assembly of the Parkinson-related Protein alpha-Synuclein.* Journal of Biological Chemistry, 2009. **284**(45): p. 31190-31199.
198. Forsythe, H.L., et al., *Stable association of hsp90 and p23, but Not hsp70, with active human telomerase.* Journal of Biological Chemistry, 2001. **276**(19): p. 15571-4.

199. Street, T.O., L.A. Lavery, and D.A. Agard, *Substrate binding drives large-scale conformational changes in the Hsp90 molecular chaperone.* Molecular Cell, 2011. **42**(1): p. 96-105.
200. Rudiger, S., et al., *CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(17): p. 11085-11090.
201. Motojima-Miyazaki, Y., M. Yoshida, and F. Motojima, *Ribosomal protein L2 associates with E. coli HtpG and activates its ATPase activity.* Biochemical and Biophysical Research Communications, 2010. **400**(2): p. 241-245.
202. Street, T.O., et al., *Cross-Monomer Substrate Contacts Reposition the Hsp90 N-Terminal Domain and Prime the Chaperone Activity.* Journal of Molecular Biology, 2012. **415**(1): p. 3-15.
203. Zuehlke, A. and J.L. Johnson, *Hsp90 and Co-Chaperones Twist the Functions of Diverse Client Proteins.* Biopolymers, 2010. **93**(3): p. 211-217.
204. Mashaghi, A., P.J. Vach, and S.J. Tans, *Noise reduction by signal combination in Fourier space applied to drift correction in optical tweezers.* Review of Scientific Instruments, 2011. **82**(11).
205. Deng, W.L., et al., *Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor.* Cell, 2012. **149**(6): p. 1233-1244.
206. *Poisson's ratio at 200.* Nature Materials, 2011. **10**(11): p. 807-807.
207. Lenz, O. and F. Schmid, *Structure of symmetric and asymmetric "ripple" phases in lipid bilayers.* Physical Review Letters, 2007. **98**(5).
208. Hawkins, R.J. and T.C.B. McLeish, *Coupling of global and local vibrational modes in dynamic allostery of proteins.* Biophysical Journal, 2006. **91**(6): p. 2055-2062.
209. Ghosh, A., et al., *Membrane-bound water is energetically decoupled from nearby bulk water: An ultrafast surface-specific investigation.* Journal of the American Chemical Society, 2007. **129**(31): p. 9608-+.
210. Mashaghi, A., et al., *Interfacial Water Facilitates Energy Transfer by Inducing Extended Vibrations in Membrane Lipids.* Journal of Physical Chemistry B, 2012. **116**(22): p. 6455-6460.
211. Yook, S.H., Z.N. Oltvai, and A.L. Barabasi, *Functional and topological characterization of protein interaction networks.* Proteomics, 2004. **4**(4): p. 928-942.
212. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nature Reviews Molecular Cell Biology, 2005. **6**(3): p. 197-208.

213. Monod, J., J. Wyman, and J.P. Changeux, *On the Nature of Allosteric Transitions: A Plausible Model.* J Mol Biol, 1965. **12**: p. 88-118.
214. Jakobi, A.J., et al., *Calcium modulates force sensing by the von Willebrand factor A2 domain.* Nature Communications, 2011. **2**.
215. Darwin, C., *The origin of species by means of natural selection, or, The preservation of favored races in the struggle for life.* 1896, New York: D. Appleton.
216. Jozefczuk, S., et al., *Metabolomic and transcriptomic stress response of Escherichia coli.* Molecular Systems Biology, 2010. **6**.
217. Lambert, D. and D.E. Draper, *Effects of osmolytes on RNA secondary and tertiary structure stabilities and RNA-Mg2+ interactions.* Journal of Molecular Biology, 2007. **370**(5): p. 993-1005.
218. Record, M.T., et al., *Responses of E-coli to osmotic stress: Large changes in amounts of cytoplasmic solutes and water.* Trends in Biochemical Sciences, 1998. **23**(4): p. 143-148.
219. Jain, N.K. and I. Roy, *Effect of trehalose on protein structure.* Protein Science, 2009. **18**(1): p. 24-36.
220. Elbein, A.D., et al., *New insights on trehalose: a multifunctional molecule.* Glycobiology, 2003. **13**(4): p. 17r-27r.
221. Kaushik, J.K. and R. Bhat, *Why is trehalose an exceptional protein stabilizer? An analysis of the thermal stability of proteins in the presence of the compatible osmolyte trehalose.* Journal of Biological Chemistry, 2003. **278**(29): p. 26458-26465.
222. Melo, E.P., et al., *Trehalose favors a cutinase compact intermediate off-folding pathway.* Biochemistry, 2003. **42**(24): p. 7611-7617.
223. Habib, S., M.A. Khan, and H. Younus, *Thermal destabilization of stem bromelain by trehalose.* Protein Journal, 2007. **26**(2): p. 117-124.
224. Tanaka, M., et al., *Trehalose alleviates polyglutamine-mediated pathology in a mouse model of Huntington disease.* Nature Medicine, 2004. **10**(2): p. 148-154.
225. Singer, M.A. and S. Lindquist, *Multiple effects of trehalose on protein folding in vitro and in vivo.* Molecular Cell, 1998. **1**(5): p. 639-648.
226. Spiess, A.N., N. Mueller, and R. Ivell, *Trehalose is a potent PCR enhancer: Lowering of DNA melting temperature and thermal stabilization of Taq polymerase by the disaccharide trehalose.* Clinical Chemistry, 2004. **50**(7): p. 1256-1259.
227. Hart, J.L., Z.M. Harris, and S.M. Testa, *Analyzing and Predicting the Thermodynamic Effects of the Metabolite Trehalose on Nucleic Acids.* Biopolymers, 2010. **93**(12): p. 1085-1092.

228. Zhu, B., et al., *Natural DNA mixed with trehalose persists in B-form double-stranding even in the dry state.* J Phys Chem B, 2007. **111**(20): p. 5542-4.
229. Iyer, P.V. and L. Ananthanarayan, *Enzyme stability and stabilization - Aqueous and non-aqueous environment.* Process Biochemistry, 2008. **43**(10): p. 1019-1032.
230. Koga, Y., K. Nishikawa, and P. Westh, *Relative Hydrophobicity/Hydrophilicity of fructose, glucose, sucrose, and trehalose as probed by 1-propanol: A differential approach in solution thermodynamics.* Journal of Physical Chemistry B, 2007. **111**(50): p. 13943-13948.
231. Sowerby, S.J., et al., *Differential adsorption of nucleic acid bases: Relevance to the origin of life.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(3): p. 820-822.

Summary

Single molecule investigations of chaperone assisted protein folding

Major experimental and theoretical efforts in the past decades have brought detailed insights into the folding of small model proteins in the test-tube, a process that has often been represented by the underlying folding energy landscape. Now, an emerging scientific frontier is to understand how proteins fold in cells. Folding of large proteins (> 100 aa) is believed to involve multiple intermediate states, more rugged folding landscapes, and increased risks of misfolding. Moreover, interactions between proteins can lead to aggregation. Aggregates reduce the number of active proteins, and present a source of toxicity for cells. Despite these risks, complex proteins normally fold efficiently in cells, owing to a dedicated chaperone-machinery that guides the folding process.

Understanding how folding pathways are mediated by chaperones presents considerable challenges. Biochemical approaches have been highly successful in unraveling the life-cycle of chaperones, their structures, binding partners, and their effects on the yield of natively folded protein. However, they are not well-suited to study the protein conformational dynamics and multiple transitions between folded states that determine folding pathways and outcomes. This thesis is a first single-molecule optical tweezers study of chaperone assisted folding. Three main chaperone classes namely Trigger factor, Hsp70 and Hsp90 are investigated that function at early to late phases of de novo protein folding in vivo and on hundreds of substrate proteins. In this thesis, we looked for generic mechanisms that underlie the functioning of these chaperones.

Chapters 1 gives an overview of the current literature on chaperone-assisted single protein folding, introduces the key open questions and discusses how the single-molecule assays can be employed to address them. This is followed in chapters 2 and 3, by introduction of two technical developments in the single-molecule optical tweezers: chapter 2 discusses a method for reducing measurement noise by combining signals; chapter 3 describes a new linkage strategy based on StrepTactin/StrepTag system for constructing protein tethers. This linkage shows an improved mechanical stability which is essential to make protein tethers for long pulling experiments

Chapter 4 describes the protein contact topology. The mathematical concept is illustrated by providing simple polymer models with binary intra-molecular contacts. It is shown how the topology put constraints on unfolding pathways of these

polymers. The notion is then applied to study the unfolding path of Firefly luciferase, a model protein with 550 aminoacids. Finally we speculate how a chaperone like Trigger factor might modulate the topology of a protein.

Chapter 5 addresses protein misfolding. We first show that unfolded luciferase has a trapped state on its folding pathway which limits the folding yield. We then show that chaperones suppress entry into this state. The underlying mechanism is described in the subsequent chapters (6 and 7). In chapter 6, we discuss how Trigger factor reshapes the landscape of a protein and how that helps the protein to avoid a misfolded state. The Trigger Factor interacts with most newly synthesized protein chains and protects them from aggregation. After release from Trigger Factor, the substrate will gain its native fold, or be transferred to the Hsp70 (Heat Shock Protein 70) chaperone and co-chaperone Hsp40 (DnaK and DnaJ in E. coli). Using optical tweezers, we have shown that Trigger Factor can remain associated as the protein chain folds via a number of intermediate states, thus playing a more intricate role than previously observed.

In chapter 7, molecular mechanism of DnaK system is studied. The DnaK system undergoes ATP-driven cycles of substrate binding and release. In the ATP bound state, DnaK has an open conformation with high on and off rates. Hydrolysis to ADP closes the lid and stabilizes peptide binding. DnaK works in concert with the nucleotide exchange factor and co-chaperone DnaJ which recruits DnaK to substrates. Using optical tweezers, we systematically address the role of different nucleotide states of DnaK and its co-chaperones in folding of single MBP. We find that the DnaK system interacts with both unfolded stretches and folded regions of a substrate protein. We propose a new model for the functioning of DnaK system both during dé novo folding and during heat stress response.

In chapter 8 we study the Hsp90 chaperone and its influence on the unfolding and folding pathway of luciferase. The ATP-consuming Hsp90 chaperone is believed to act downstream of Hsp70, and interact with near-native protein states. Eukaryotes contain a wide range of Hsp90 chaperones, which are known to have regulatory functions. In this thesis we focused on the E. coli variant, HtpG, as it is the only Hsp90 chaperone in E. coli and appears less involved in regulatory functions. Not much is known about the function of Htpg, though interestingly it has recently been shown to interact with incompletely folded proteins, which suggest a chaperone role in guiding folding. We find that HtpG interacts with the unfolded chain in an ATP independent manner. We find steps during refolding and unfolding that depend on HtpG and its ATPase function. The emergence of steps suggests interaction of HtpG with the folded regions of proteins or loop induction in polypeptide chains.

In chapter 9 we discuss stress response processes at a broader level. We address the role of lipid membranes and sugars (e.g. trehalose) in stress response. The influence of trehalose on mechanical properties and melting transition of double stranded DNA is explored. We also discuss protein vWF which is a stress sensor in the extracellular environment.

Samenvatting

Enkele molecule studies van chaperone gestuurde eiwit vouwing

In de afgelopen decennia hebben grote experimentele en theoretische inspanningen gezorgd voor uitgebreide inzichten in het vouwen van kleine model eiwitten in de reageerbuis, een proces dat vaak is weergegeven door middel van het onderliggende energie landschap van het vouwen. Een nieuw wetenschappelijk gebied dat nu naar voren komt is het begrijpen hoe eiwitten in cellen zich vouwen. Het wordt verondersteld dat het vouwen van grote eiwitten (> 100 aa) betrekking heeft op meerdere tussenliggende toestanden, ruigere landschappen van het vouwen en grotere kansen op misvouwen. Daarbij kunnen wisselwerkingen tussen eiwitten leiden tot aggregatie. Aggregaten verminderen het aantal actieve eiwitten, en leveren een vorm van giftigheid voor cellen. Ondanks deze gevaren vouwen complexe eiwitten normaliter efficiënt in cellen, dankzij een toegewijde chaperonne-eiwit machinerie die het vouwproces begeleidt.

Begrijpen hoe chaperonne-eiwitten bemiddelen langs welk pad eiwitten zich vouwen levert aanzienlijke uitdagingen. Biochemische aanpakken zijn zeer succesvol geweest in het ontrafelen van de levensduurcyclus van chaperonne-eiwitten, hun structuren, bindingspartners, en hun effect op het aantal gevouwen eiwitten in hun geschikte toestand. Maar deze methodes zijn niet geschikt om de conformationele dynamica van eiwitten en de meerdere overgangen tussen gevouwen toestanden, welke de paden waarlangs eiwitten zich vouwen en hun uitkomsten bepalen, te beschrijven. Deze scriptie is een eerste single molecule optical tweezers onderzoek over chaperonne-eiwit geassisteerd vouwen. Drie chaperonne-eiwit hoofdklassen, namelijk Trigger factor, Hsp70 en Hsp90, die functioneren in vroege of late stadia van het de novo eiwit vouwen in vivo en op honderden substraat eiwitten, worden onderzocht. In deze scriptie zochten we naar algemene onderliggende mechanismen voor het functioneren van deze chaperonne-eiwitten.

Hoofdstuk 1 geeft een overzicht van de huidige literatuur over het chaperonne-eiwit geassisteerd vouwen van enkele eiwitten, presenteert de belangrijkste open vragen en bediscussieert hoe de single molecule analyses gebruikt kunnen worden om ze te beantwoorden. Dit wordt opgevolgd door hoofdstukken 2 en 3, door de introductie van twee technische ontwikkelingen in de single molecule optical tweezers: hoofdstuk 2 bespreekt een methode voor het verminderen van meetruis door het combineren van signalen; hoofdstuk 3 beschrijft een nieuwe binding strategie gebaseerd op het StrepTactin/StrepTag systeem voor het

construeren van eiwit kettingen. Deze binding geeft een verbeterde mechanische stabiliteit die essentieel is voor het maken van eiwit kettingen voor lange trek experimenten.

Hoofdstuk 4 bespreekt de eiwit contact topologie. Het wiskundige concept wordt geïllustreerd door het leveren van simpele polymeer modellen met binaire intramoleculaire contacten. Er blijkt hoe de topologie beperkingen oplegt voor de paden waarlangs deze polymeren zich kunnen ontvouwen. Dit gegeven is dan toegepast bij het bestuderen van het pad waarlangs Firefly luciferase zich ontvouwt, een model eiwit met 550 aminozuren. Tot slot speculeren we hoe een chaperonne-eiwit zoals Trigger factor de topologie van een eiwit zou kunnen regelen.

Hoofdstuk 5 richt zich op het misvouwen van eiwitten. Eerst laten we zien dat ontvouwen luciferase een gevangen toestand heeft op het pad waarlangs dit vouwt dat een limiet geeft op de opbrengst van het vouwen. Dan laten we zien dat chaperonne-eiwitten toegang tot deze toestand onderdrukken. Het onderliggende mechanisme wordt beschreven in de hierop volgende hoofdstukken (6 en 7). In hoofdstuk 6 bespreken we hoe Trigger factor het landschap van een eiwit vervormt en hoe dit het eiwit helpt om een misgevouwen toestand te vermijden. De Trigger factor heeft een wisselwerking met de meeste nieuwe gesynthetiseerde eiwit kettingen en beschermt ze van aggregatie. Nadat Trigger factor vrijkomt, verkrijgt het substraat zijn geschikte vouw, of zal deze overgedragen worden aan het Hsp70 (Heat Shock Protein 70) chaperonne-eiwit en co-chaperonne-eiwit Hsp40 (DnaK en DnaJ in E. coli). We hebben, gebruikmakend van optical tweezers, laten zien dat Trigger factor geassocieerd kan blijven terwijl de eiwit ketting vouwt, langs een aantal tussenliggende toestanden, waaruit blijkt dat het een meer ingewikkelde rol speelt dan eerder is waargenomen.

In hoofdstuk 7 wordt het moleculaire mechanisme van het DnaK systeem bestudeerd. Het DnaK systeem ondergaat ATP-gedreven cycli van substraat binding en ontbinding. In de ATP gebonden toestand heeft DnaK een open conformatie met hoge on en off rates. Hydrolyse naar ADP sluit de deksel en stabiliseert peptide binding. DnaK werkt samen met de nucleotide uitwisselingsfactor en co-chaperonne-eiwit DnaJ , welke DnaK werft aan substraten. Gebruikmakend van optical tweezers bespreken we systematisch de rol van verschillende nucleotide toestanden van DnaK en zijn co-chaperonne-eiwitten in het vouwen van een enkele MBP. We hebben gevonden dat het DnaK systeem wisselwerking ondervindt met zowel de ongevouwen stukken en gevouwen gebieden van het substraat eiwit. We stellen een nieuw model voor voor het functioneren van het DnaK systeem zowel gedurende de novo vouwen als warmte stress respons.

In hoofdstuk 8 bestuderen we het Hsp90 chaperonne-eiwit en zijn invloed op het pad waarlangs luciferase vouwt en ontvouwt. Het wordt geloofd dat het ATP consumerende Hsp90 chaperonne-eiwit actie ondervindt downstream van Hsp70, en wisselwerking heeft met bijna-geschikte eiwit toestanden. Eukaryoten bevatten een grote scala aan Hsp90 chaperonne-eiwitten, waarvan men weet dat ze regelgevende taken hebben. In deze scriptie keken we naar de E. coli variant, HtpG, omdat het het enige Hsp90 chaperonne-eiwit is in E. coli en minder invloed heeft op regelgevende taken. Er is niet veel bekend over de functie van Htpg, hoewel er recent is aangetoond dat het een wisselwerking heeft met onvolledig gevouwen eiwitten, wat een rol als chaperonne-eiwit zou kunnen suggereren in het sturen van het vouwen. We hebben gevonden dat Htpg wisselwerking heeft met de ongevouwen ketting op een ATP onafhankelijke manier. We vinden stappen tijdens het hervouwen en ontvouwen die afhankelijk zijn van de HtpG en zijn ATPase taak. Het verschijnen van stappen suggereert wisselwerking van HtpG met de gevouwen regio's van eiwitten of lus inductie in polypeptide kettingen.

In hoofdstuk 9 bediscussiëren we het stressrespons proces op een breder niveau. We richten ons op de rol van lipide membranen en suikers (bijv. trehalose) bij stressresponsen. De invloed van trehalose op mechanische eigenschappen en de smelt overgang van dubbelstrengs DNA worden verkend. Ook praten we over het eiwit vWF welke een stress sensor is in de extracellulaire omgeving.

Acknowledgements

As physicists with interests in biological systems, Sander and I shared an enthusiasm for understanding the protein folding problem. A basic tenet of protein folding has been the seminal *in vitro* work of Christian Anfinsen, which demonstrated that the sequence of a protein alone determines its native structure and the folding is a spontaneous energy-independent process. Does this apply to protein folding in vivo? Is this relevant to proteins of arbitrary size and number of domains? Why do chaperones hydrolyze ATP then? Does a given protein fold into identical states when it folds -- in a single molecule assay -- in the absence and presence of chaperones? While studying a biological system, there is often a risk of being distracted by specific details. "Is this generic?" we often ask ourselves in our meetings when looking at the pattern in the collected data from a certain experimental condition. Despite the hard work, long days and the difficulty of collecting data on the tweezers, I was always excited to do more and more experiments, thanks to the exciting meetings I had with Sander and our shared sense of beauty. I am immensely thankful to Sander for his support and continuous interest throughout this project. Sander, thanks for always being open and critical to new ideas and giving me the scientific freedom that was vital for my development. Thank you, it has been a great time.

I have been fortunate to work with leading figures in the chaperone field in the course of my PhD project. I am grateful to our collaborators in Heidelberg, Bernd Bukau, Guenter Kramer, Matthias P. Mayer and Axel Mogk. Bernd, it was a pleasure to work with you and I am thankful to your critical input on the TF and the DnaK stories. Guenter, it was really nice working with you, thanks for all your valuable contributions and countless scientific discussions. Matthias, thanks for your time, inputs to the TF and the DnaK stories and critical reading of my thesis chapters.

During last four years I enjoyed working with a number of masters students, PhD students and postdocs within our tweezers team: Philipp, Samaneh, Peter, Sergey, Fatemeh and Roeland. It was a pleasure to work with you and I am thankful for all the great moments and your scientific input. My special thanks to my friend and my collaborator Arjen Jakobi and his mentor

Eric G. Huizing. It was a great experience for me to work with you on the vWF project. I am grateful to Wiet, a very nice friend and co-worker for all the moments we spent together thinking about protein networks.

I also take this opportunity to thank my collaborators in the theory group led by Peter Bolhuis at the University of Amsterdam: Peter, Kush, Jocelyne. It was very nice to work together and to "visualize" the chaperone assisted protein folding.

My special thanks go to Andrew, Katja, Chris, Stephen, Eva, Francois, Philippe, Nils, Bjorn, Thomas and Baldo for critical reading of my thesis. I can only offer my sincerest gratitude to you. Many thanks for your time and your kindness. Many thanks to Bela, Jeroen, Gijsje, Marileen, Tom, Mirjam and Peter Rein for lots of discussions. In particular I am thankful to Bela and Jeroen for being kind, and investing time to give me input on many different problems.

Biological reactions often take place in the presence of water molecules. A fundamental problem in biology concerns the interplay between water and biomolecules. I initiated a project on interfacial water together with Huib Bakker and Mischa Bonn during my stay in AMOLF. I take this opportunity to thank Mischa and Huib for their support, insightful discussions and kindness. I can't thank you enough.

I would like to thank all the people in the Sander Tans group, Daan, Marjon, Bertus, Vanda, Alex, Sara, Sebastian, Noreen and Manju. Thanks for the good moments and your support. I also take this opportunity to thank Sana, Jose, Andrej, Stef, Jeanette, Roland, Martin, Julia, Piere, Sophie, Nuria, Feng, Bob, Kamila, Sanli, Nicholas and Svenja and my nice office-mates Chris and Milena. You have been kind to me during my stay at AMOLF and I really appreciate it. My special thanks go to Abbas for being a great friend and his lovely parents for the kindness extended to me. You do not expect an institute director to be aware of the work of a PhD student who is not even working in his/her field. I find Albert an exception! Thanks for your kindness and support. I am thankful to all AMOLF people and particularly the technical support facility. During my lifetime I have been very often full of inner happiness all due to my lovely parents, grandparents and sisters.

Curriculum Vitae

Alireza Mashaghi Tabari was born in Iran, during the 1979 revolution. His school time overlapped with the devastating Iran-Iraq war and economic sanctions. In 1994, he entered a highly competitive M.D. programme at Tehran Medical School. He graduated in 2002 with a thesis on immunogenetics of Vitiligo. While studying at the medical school he developed interest in pursuing a parallel track science career and became the first Iranian student with dual education. He studied chemistry and physics at Tehran and Sharif Universities and finished his master's studies with a thesis on statistical physics of complex networks. He then moved to Europe and worked on the development of Fluorescence Correlation Spectroscopy to study cortical polarity proteins in C. elegans single egg cells at the Max Planck Institute for Molecular Cell Biology and Genetics. He then moved to Max Planck Institute for Biophysical Chemistry where he used high resolution NMR spectroscopy to study the structure of spliceosome, an RNA-protein complex involved in mRNA processing. Further, he did a postdiploma at ETH Zurich, where he initiated various research projects in material science, nanotechnology and surface science and supervised several undergraduate and graduate students. At the start of 2009 he started his PhD at FOM institute AMOLF and Delft University of Technology.

List of publications

Related to this thesis*

1 *Mashaghi A, Tans SJ,* Single molecule study of chaperone assisted protein folding. (Review article)

Chapter 1

2 *Mashaghi A†, Vach PJ†, and Tans SJ,* Noise reduction by signal combination in Fourier space applied to drift correction in optical tweezers, **Reviews of Scientific Instruments 82**, 115103 (2011)

Chapter 2 †equal contribution

3 *Moayed F†, Mashaghi A†, Tans SJ,* A polypeptide-DNA hybrid with selective linking capability applied to single molecule nano-mechanical measurements using optical tweezers, **submitted**.

Chapter 3 †equal contribution

4 *Mashaghi A, Tans SJ,* Protein contact topology. (in preparation)

Chapter 4

5 *Mashaghi A, Mashaghi S, Tans SJ,* Chaperones Trigger factor and heat shock protein suppress entry into misfolded state. (in preparation)

Chapter 5

6 *Mashaghi A, Kramer G, Bechtluft P, Zachmann-Brand B, Driessen AJM, Bukau B, and Tans SJ,* Reshaping of a protein folding landscape by the chaperone Trigger Factor, **submitted**.

Chapter 6

7 *Mashaghi A, Bezrukavnikov S, Kramer G, Mayer MP, Bukau B, Tans SJ,* Molecular mechanism of the DnaK chaperone system. (in preparation)

Chapter 7

8 *Mashaghi A, Kramer G, Tans SJ,* Protein folding under the influence of HtpG. (in preparation)

Chapter 8

9 *Bezrukavnikov S†, Mashaghi A†, van Wijk R, Tans SJ,* The influence of trehalose on DNA overstretching forces: a single molecule study (in preparation)

Chapter 9 †equal contribution

*Selected list of publications on single-molecule optical tweezers experiments.

Other publications

1 *Ramezanpour A, Karimipour V and Mashaghi A*.* Generating correlated networks from uncorrelated ones, **Physical Review E** 67, 046107 (2003)
2 *Mashaghi A*, Ramezanpour A and Karimipour V.* Investigation of a protein complex network, **European Physical Journal B** 41, 113–121 (2004)
3 *Khan TR, Grandin HM, Mashaghi A, Textor M, Reimhult E and Ilya Reviakine.* Lipid Redistribution in Phosphatidylserine-containing Vesicles Adsorbing on Titania. **Biointerphases** 3(2) pp FA90-FA95 (2008)
4 *Mashaghi A, Swann M, Popplewell J,* et al., Optical anisotropy of supported lipid structures probed by waveguide spectroscopy and its application to study of supported lipid bilayer formation kinetics. **Analytical Chemistry** 15;80(10):3666-76 (2008)
5 *Petrásek Z, Hoege C, Mashaghi A, Ohrt T, Hyman A & Schwille P,* Characterization of protein dynamics in asymmetric cell division by scanning fluorescence correlation spectroscopy. **Biophysical Journal** 95(11), 5476-5486 (2008)
6 *Amstad E, Zürcher S, Mashaghi A, Wong JY, Textor M, Reimhult E,* Surface Functionalization of Single Superparamagnetic Iron Oxide Nanoparticles for Targeted Magnetic Resonance Imaging, **Small** 5(11) 1334–1342, (2009)
7 *Zamani M, Tabatabaiefar MA, Mosayyebi S, Mashaghi A, Mansouri P,* Possible association of the CD4 gene polymorphism with vitiligo in an Iranian population. **Clinical and Experimental Dermatology** 35, 521-425 (2010)
8 *Baumann MK, Amstad E, Mashaghi A, Textor M, and Reimhult E,* Characterization of supported lipid bilayers incorporating the phosphoinositides phosphatidylinositol 4,5-biphosphate and phosphoinositol-3,4,5-triphosphate by complementary techniques. **Biointerphases** 5, 114-119 (2010)
9 *Nalam PC, Clasohm JN, Mashaghi A and Spencer ND,* Macrotribological Studies of Poly(L-lysine)-graft-Poly(ethylene glycol) in Aqueous Glycerol Mixtures **Tribology Letters** 37, 541-552 (2010)
10 *Jakobi, AJ, Mashaghi, A, Tans SJ, Huizinga EG,* Calcium modulates force sensing by the von Willebrand factor A2 domain. **Nature communications** 2, 385 (2011)
11 *Jadidi T, Mashaghi A*, Seyyed-Allaei H, Maass P, Tabar MRR,* Poisson's ratio and local strain in the phospholipid membrane, Submitted (2012)
12 *Mashaghi A*, Partovi-Azar P, Jadidi T, Maass P, Nafari N, Rahimi Tabar MR, Bonn M and Bakker HJ,* Hydration strongly affects the molecular and

electronic structure of membrane phospholipids. **Journal of Chemical Physics** 136(11), 1-5 (2012)

13 *Kaufmann S, Ilg K, Mashaghi A, Textor M, Priem B, Aebi M and Reimhult E,* Supported lipopolysaccharide bilayers. **Langmuir** 28 (33), 12199–12208 (2012)

14 *Mashaghi A*, Partovi-Azar P, Jadidi T, Maass P, Nafari N, Esfarjani K, Rahimi Tabar MR, Bakker HJ and Bonn M,* Interfacial water facilitates energy transfer by inducing extended vibrations in membrane lipids, **Journal of Physical Chemistry B** 116(22), 6455–6460 (2012)

15 *Singhal K, Vreede J, Mashaghi A, Tans SJ, Bolhuis PG,* Hydrophobic Collapse of Trigger Factor Monomer in Solution, submitted (2012)

**corresponding author*