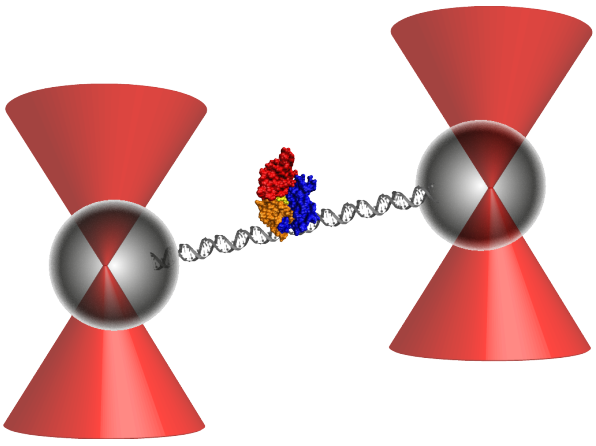
DANIELA INGRID BAUER

SINGLE-MOLECULE FORCE SPECTROSCOPY REVEALS THE DOMAIN HIERARCHY OF DNAK NBD AND IDENTIFIES A FOLDING SEED AND MINIMAL ATP-BINDING DOMAIN OF DNAK

DISSERTATION AT THE PHYSICS DEPARTMENT E22
TECHNICAL UNIVERSITY OF MUNICH



Fakultät für Physik Technische Universität München Lehrstuhl für Biophysik

SINGLE-MOLECULE FORCE SPECTROSCOPY REVEALS THE DOMAIN HIERARCHY OF DNAK NBD AND IDENTIFIES A FOLDING SEED AND MINIMAL ATP-BINDING DOMAIN OF DNAK

Daniela Ingrid Bauer

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

Vorsitzender: Prof. Dr. Martin Zacharias

Prüfende/-r der Dissertation:

1. Prof. Dr. Matthias Rief
2. Prof. Dr. Hendrik Dietz

Die Dissertation wurde am 07.08.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 05.10.2018 angenommen.

ABSTRACT

The folding energy landscape of large proteins often exhibits partially stable intermediates. These can have profound effects on the timescale of folding to the native state. In this thesis, the folding and unfolding of the nucleotide-binding domain of DnaK (EcNBD) was investigated at a single-molecular level using optical tweezers. This method is a very powerful tool due to its high temporal and spatial resolution. Hence, ligand binding events as well as the energetics of structural changes can easily be monitored. EcNBD (45kDa) consists of two lobes: Lobe I and lobe II, which can further be subdivided into a/b. Lobe II harbors most of the coordinating residues of the active site, which binds Mg-ATP with nano-molar affinity.

The unfolding as well as the refolding pathway of this large protein domain could be dissected. In the case of the unfolding pathway, ligand binding can shift the whole domain hierarchy of EcNBD by stabilizing lobe II. The c-terminal helix acts like a molecular glue and is the major determinant of the high mechanical stability.

The refolding of EcNBD follows a sequential pathway populating two fast forming refolding intermediates. Using loop-insertion variants and different pulling geometries, mapping of the refolded portions to the lobe II region was possible. Moreover, we could show that lobe II alone binds Mg-ATP with micromolar affinity (33 µM) and provides a folding platform for lobe I. To prove that lobe II is a folding nucleus, a chimeric protein was designed by grafting lobe II from *E.coli* into the refolding incompetent mitochondrial NBD of yeast (mt-NBD). The designed chimera showed native-like ATPase activity and enhanced refolding ability.

Furthermore, a minimal ATP-binding domain consisting of lobe II was created. Using a limited proteolysis assay and standard protein kinetics methods, we showed that this domain can specifically bind adenosine nucleotides with a micromolar affinity. To this end, the structure of that domain using small-angle X-ray scattering (SAXS) and X-ray crystallography was determined. These experiments provided structural information and detailed molecular insights into the ligand binding geometry and its effects on the overall structure.

In summary, optical tweezers data deciphered a detailed unfolding and refolding pathway for EcNBD. We showed that lobe II is a stable nucleotide-binding folding nucleus for the NBD fold. Furthermore, these findings provide concepts which may also be relevant for the folding behavior of other members of this protein superfamily and their evolution.

ZUSAMMENFASSUNG

Die Faltungsenergielandschaft von großen Proteinen weist oft partiell gefaltete Strukturelemente auf. Diese können Einfluss auf die Gesamtfaltungszeit zum nativen Zustand nehmen. In dieser Arbeit wurde der Faltungspfad der Nukleotid-bindenden Domäne von DnaK (EcNBD) mithilfe von Einzelmolekülkraftspektroskopie entschlüsselt. Diese Methode ist wegen ihrer hohen Zeit- und Raumauflösung bestens geeignet um Ligandbindungsereignisse und energetische Änderungen von Substrukturen aufzuzeichnen. EcNBD lässt sich strukturell in zwei, etwa gleich große, Lobes unterteilen (I und II), welche wiederum in zwei weiter Substrukturen untergliedert werden (a und b). Lobe II stellt dabei die Großzahl der koordinierenden Residuen, die nötig für die Ligandenbindung (Mg-ATP) sind. Die hohe Affinität von EcNBD für Mg-ATP liegt im nanomolaren Bereich.

Im Rahmen dieser Arbeit ist es gelungen, sowohl den Entfaltungspfad als auch den Rückfaltungspfad zu charakterisieren. Im Falle der Entfaltung bestimmt der Ligand wesentlich die Domänenhierarchie. Ist der Ligand gebunden, so wird Lobe II gegenüber Lobe I stabilisiert und entfaltet später. Die C-terminale Helix ist für die hohe mechanische Stabilität der NBD verantwortlich und wirkt als molekularer Kleber.

Die Rückfaltung involviert die Population von zwei Intermediaten unter Kraft. Unter Zuhilfenahme von Schleifeninsertionsvarianten und Varianten mit verschiedenen Ziehgeometrien, konnten die Intermediate strukturell der Lobe II Region zugeordnet werden. Lobe II kann drüberhinaus bereits Mg-ATP mit einer Affinität von 33 µM binden. Die Hypothese, dass Lobe II ein Faltungskeim für Lobe I ist, konnte mit der faltungsinkompetenten homologen Variante aus Mitochondrien (mtNBD), belegt werden. Durch das Design eines chimären Proteins bestehend aus mtNBD und Lobe II aus EcNBD, konnte eine faltungskompetente Variante mit nativer ATPase Rate kreiert werden.

Lobe II ist in Isolation eine minimale ATP-bindende und stabile Domäne. Kinetik- und Proteolyseexperimente zeigen, dass diese Domäne mit mikromolarer Affinität spezifisch Adenosinnukleotide bindet. Die Strukturaufklärung mithilfe von SAXS und Röntgenkristallographie ergab detailierte molekulare Einblicke in die 3D-Struktur und das aktive Zentrum.

Die vorgestellen Einzelmolekül- und biophysikalischen Resultate liefern wichtige Informationen für das Faltungsverhalten anderer Mitglieder der Proteinsuperfamilie geben. Basierend auf unseren Ergebnissen, sind verschiedene Evolutionsszenarien möglich.

PUBLICATIONS

Some ideas and figures have been adapted from the following publications:

- **Daniela Bauer**, Dale R Merz, Benjamin Pelz, Kelly E Theisen, Gail Yacyshyn, Dejana Mokranjac, Ruxandra I Dima, Matthias Rief, and Gabriel Žoldák.
  "Nucleotides regulate the mechanical hierarchy between subdomains of the nucleotide binding domain of the Hsp70 chaperone DnaK."
  Proceedings of the National Academy of Sciences of the United States of America, 112 (33), (Aug. 2015), pp. 10389–10394

- **Daniela Bauer**, Sarah Meinhold, Roman P Jakob, Johannes Stigler, Ulrich Merkel, Timm Maier, Matthias Rief, Gabriel Žoldák.
  "A folding nucleus and minimal ATP binding domain of Hsp70 identified by single-molecule force spectroscopy".
  Proceedings of the National Academy of Sciences of the United States of America, 115 (18), (May 2018), pp. 4666-4671

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AFM Atomic Force Microscopy

ATP adenosine triphosphate

ADP adenosine diphosphate

AMP adenosine monophosphate

BS Schlierf-Berkemeier-Rief model

CV column volume

dsDNA double-stranded deoxyribonucleic acid

ssDNA single-stranded deoxyribonucleic acid

DNA deoxyribonucleic acid

EcNBD *E.coli* NBD

EcNBD-nc *E.coli* NBD n-/c-terminal pulling direction

eWLC extensible worm-like chain

SAXS small-angle x-ray scattering

FRET Förster resonance energy transfer

HMM Hidden Markov model

MABA 8-[(4-Amino)butyl]-amino-adenosine-5′-triphosphate

mtNBD mitochondrial NBD

PCR polymerase chain reaction

PBS phosphate-buffered saline

SDS-PAGE sodium dodecyl sulfate –polyacrylamide gel electrophoresis

SEC size-exclusion chromatography

TCEP Tris-(2-carboxy-ethyl)-phosphin

50TKM 50 mM TRIS pH 7.5, 50 mM KCl, 5 mM $MgCl_2$

50TKM-AS 50TKM 1M ammonium sulfate buffer

TRIS 2-Amino-2-(hydroxymethyl)propane-1,3-diol

WLC worm-like chain

Part I

INTRODUCTION

1

PROTEIN FOLDING AND THE ROLE OF MOLECULAR CHAPERONES

Life depends on the function of proteins, whose function in turn depends on their structure. Many proteins need to adopt a fine-tuned 3-dimensional structure. This 3D structure is built up by α-helices, β-sheets and random coils. However, some proteins are not necessarily completely folded. There are many examples, like the pKID-KIX system [29], where one partner folds upon binding to the other [56]. To maintain large structural rearrangements, the protein structure needs to be very dynamic. These dynamics are crucial for the function of most proteins. Despite all the investigations of the protein structure/-function relationship, which have happened over the past 50 years, the fundamental question how a long chain of amino acids finds its 3D structure is not yet fully understood.

The current view of how proteins are made in a cell and how they fold is outlined in the following. Proteins are made by synthesizing a chain of amino acids by the RNA-protein machinery called the ribosome (Figure 1). In the ribosome, this chain is translated from an mRNA template into an amino acid chain. In this translation process, single tRNA bound amino acids are shuffled to the peptidyl-transferase center (PTC) of the ribosome. Here, the new amino acid gets connected via a peptide bond to the growing chain. From the PTC, the nascent chain proceeds through a narrow tunnel to the exit point where first secondary structures can form (Figure 1). The secondary structural elements that from early on are dictated by the amino acid composition of the chain. Some amino acids favor, for example, α-helical conformations, others disfavor this structural element. The formation of these small substructures in the tunnel is much faster than translation. Hence, the first folding events can already happen in a quasi-equilibrium in the restricted environment of the tunnel. Due to the small width of the tunnel, the formation of larger structural elements is restricted. It has to be mentioned that these structural elements at this stage are not necessarily in their native state. Larger structural units can be formed at the exit port and build up the native protein structure in a highly cooperative manner [60, 86].

In a very basic picture, the formation of larger folding units is driven by hydrophobic and hydrophilic interactions, hydrogen bonding and van der Waals forces (Figure 1). The formed secondary structural elements may align cooperatively to cover hydrophobic patches within the structure, and more hydrophilic ones face the surrounding aqueous solution (*hydrophobic effect*) [114]. Further interactions between these

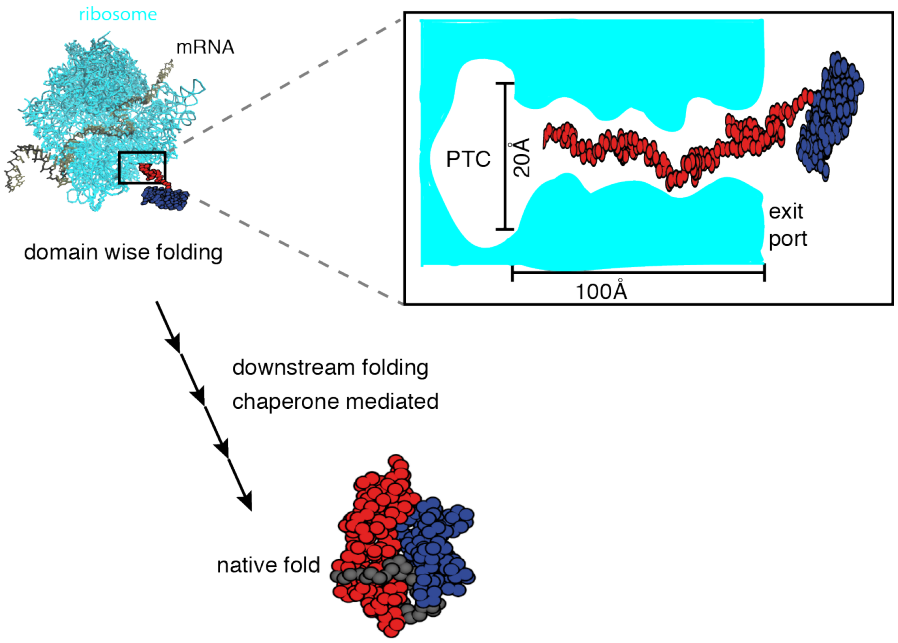


Figure 1: Protein translation and folding into the designated tertiary structure. The protein is synthesized by a set of protein machinery called the ribosome (cyan). The ribosome uses mRNA (black helix) as a template to generate the encoded amino acid chain (blue and red chain). The synthesized amino acid chain can fold into small secondary structures such as α-helices, β-sheets or loop regions within the ribosome exit tunnel (zoom). The peptidyl transferase center (PTC) in the ribosome is about 20 Å wide and runs into a small channel of 100 Å length. After the exit port, larger domains can form. Further folding to higher order structure may involve the action of chaperones [8].

units might lead to multi-domain proteins with a distinct final 3-dimensional orientation. Therefore, the folding process constrains the number of degrees of freedom of the chain and minimizes the entropic energy term [40]. This loss in entropic energy is coupled with a gain of enthalpic energy[1]. This enthalpic term includes bonding and non-bonding interactions within the chain. Therefore, the folded state is characterized by a more negative Gibbs free energy compared to the unfolded state. The described process is only a rough picture of how protein folding works and further research will be necessary to understand the folding process in more details.

[1] *The Gibbs free energy is defined as* $\Delta G = \Delta H - T\Delta S$.

The goal of understanding protein folding is to make predictions how a protein structure forms and what are crucial folding intermediates which can be responsible for malfunctions. To be able to predict

protein folding pathways, descriptive models are needed. The first theories were developed 55 years ago, when Anfinsen *et al.* demonstrated that proteins could fold by themselves [3]. Additionally, Levinthal suggested that an amino acid chain can adopt a high number of possible folds. Random searching, through the many plausible conformations, would not allow folding in a reasonable timescale [63]. Controversially, proteins are indeed able to fold on a time scale of milliseconds. This controversy led Levinthal to the conclusion that proteins follow a programmed structure formation pathway [63]. This pathway can be described by the folding funnel theory, which is the currently used model.

In the folding funnel theory, an unfolded protein chain folds to the native state (global energy minimum) by passing through local minima along the folding trajectory. These local minima are called intermediates [121]. Folding intermediates are submolecular parts of the structure which can form semi-independently and lead the rest of the protein chain to the native fold. The folding pathway of a single-domain or even a multi-domain protein can then be described as a sequential folding process of individual units, which progressively guide the way to the native structure [66, 70]. Importantly, these intermediates can also accommodate critical biological functions, such as the folding intermediate of cytochrome c, which drives the alkaline transition equilibrium [71]. Folding intermediates are populated, especially, along the folding pathway of larger protein domains (>100 amino acids). As most of the proteins include more than 100 amino acids[2], the majority of them will fold to the native state by populating folding intermediates.

[2] *Median number of amino acids per protein in E.coli ≈ 310 and in human ≈ 560 amino acids.*

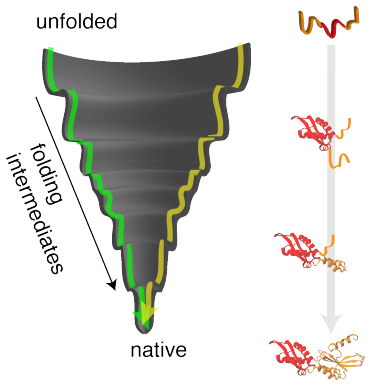


Figure 2: Protein folding funnel. Folding of a protein from a completely unfolded state to the functional native state populates so-called folding intermediates. These transient intermediates can be on- or off-pathway, and are differently populated in time according to their stability. The right side depicts, an example of a productive folding pathway with two on-pathway intermediates.

It is often indispensable that during the folding process, hydrophobic patches get buried; otherwise they tend to aggregate in the highly

crowded environment [31]. The cell has developed an elaborate system to prevent aggregation and misfolding. Already during protein synthesis, kinetic traps are precluded by the ribosomal architectural features [8]. After translation, the protein chain associates with a specialized machinery, called molecular chaperones. These proteins ensure proper folding and prevent aggregation during and after translation as well as under stress conditions[3]. Under such conditions, more proteins are prone to un- or misfold, and therefore chaperones have to make sure that these proteins stay soluble. Chaperones can act as a *foldase*, or *holdase*. Foldases support the folding of client proteins in an adenosine triphosphate (ATP)-dependent manner (for example the Hsp70 or the GroEL/ES chaperone system). Holdases bind certain folding intermediates and prevent aggregation during the folding process, but they do not actively support folding (for example DnaJ or Hsp33). If this system fails, protein aggregation can lead to severe consequences in the cell. Very prominent diseases which are caused by protein aggregation include neurodegenerative diseases like Alzheimer's or Parkinson disease [45, 102].

[3] *Stress situations are elevated temperatures or oxidative stress. Often the term heat-shock proteins (Hsp's) is used.*

In summary, folding is the most crucial step for proper protein function. Some proteins can fold in a single step, while others, especially larger domains, populate folding intermediates along their folding pathway. Some protein chains depend on the function of chaperones which help to proceed with the folding path to the native fold. To understand the folding process in more detail is an essential part of ongoing research. The presence of on- and off-pathway intermediates will play an important role to understand protein miss-folding and associated diseases.

2

HEAT SHOCK PROTEIN 70 IS A MOLECULAR CHAPERONE

Heat shock proteins have many essential tasks in the cell. Figure 3 summarizes most essential functions of Hsp's, such as protein translocation or protein-assisted folding. To fulfill these tasks, Hsp's need to bind their clients. This binding site is a specific amino acid sequence, which is usually hydrophobic. By binding the protein client to Hsp, it can, e.g., fold more efficiently, be translocated through pores or prevented from aggregation [61]. The following chapter will focus on the Hsp70 system, in *E.coli* (DnaK) and *S.cerevisiae* (Ssc1).

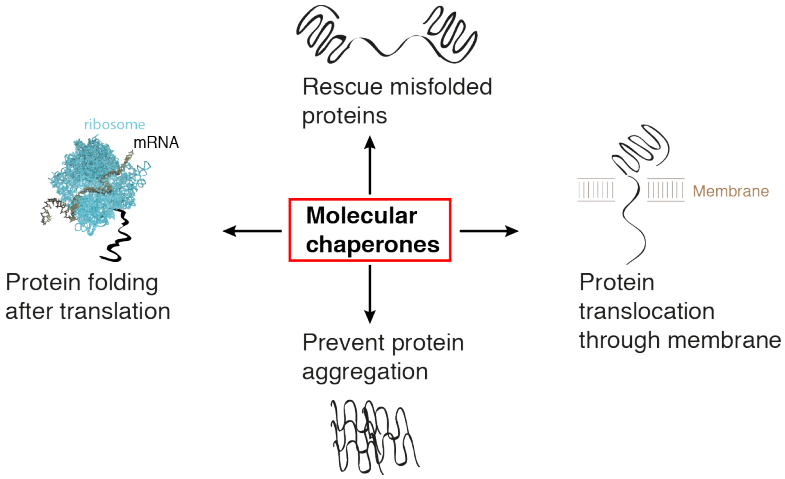


Figure 3: Heat-shock proteins fulfill many important tasks in the cell. First, binding of misfolded proteins and help to refold them into their native structure. Second, translocation of proteins through a membrane. Third, prevention of protein aggregation in the cell. Last, helping newly synthesized proteins to adapt their functional fold.

2.1 STRUCTURE AND FUNCTION OF THE E.COLI HSP70

Structure and function of proteins are highly correlated. This correlation will in the next sections be described on the example of *E.coli* Hsp70 (DnaK). DnaK can specifically bind ligand, which induces certain conformational changes in the NBD. These changes enable designated structural elements to serve as signal transducers to the substrate-binding domain. This allosteric communication couples the nucleotide state of the NBD to structural rearrangements of the SBD. The communication between NBD and SBD plays an important role in DnaK's

functions as a foldase or a holdase and its capacity as a housekeeping system [24].

2.1.1 *Overall structure of DnaK*

From a structural point of view, DnaK consists of two domains connected via a flexible linker (Figure 4A). The C-terminal 25-kDa substrate-binding domain (SBD) incorporates two parts, a β-sandwich domain harboring the substrate-binding site and an α-helical lid domain. The 45-kDa N-terminal ATPase domain (NBD) is necessary for ligand binding and hydrolysis.

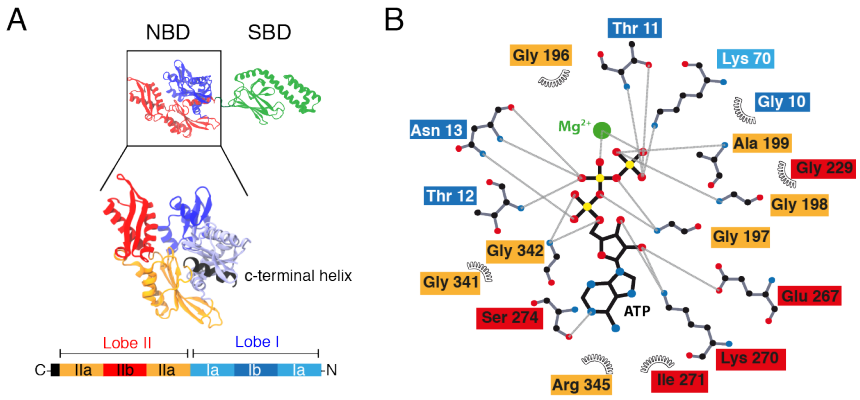


Figure 4: (A) Crystal structure of Hsp70 in the open state (PDB: 2KHO). Hsp70 can be subdivided into the C-terminal substrate-binding domain (SBD) and the N-terminal nucleotide binding domain (NBD). (B) The coordinating residues in the NBD are color-coded as introduced in (A). The LigPlot shows residues involved in ligand binding [126]. Most are located in lobe II. Lobe I residues are predominantly coordinating the magnesium ion and the third phosphate group.

NBD consists of two lobes: lobe I (blue) and lobe II (orange/red) in Figure 4A. The lobes are further subdivided into *a* and *b*, where the b-subdomains are inserted into the a-domains[4], as the amino acid sequence organization in Figure 4A, bottom, shows. This structural organization is also reflected in ligand binding. NBD specifically binds adenosine nucleotides (ATP and adenosine diphosphate (ADP)) in a magnesium-dependent manner. Lobe I of NBD mainly coordinates the magnesium ion. Lobe IIa and lobe IIb position the adenosine and lobe I mainly coordinates the phosphate groups, as visualized by the LigPlot in (Figure 4B) [5]. The nucleotide state in NBD defines if SBD is in the open or closed conformation: The lid domain is bound to the NBD, or the lid domain covers the substrate-binding site, respectively. This dependency is important for the chaperone cycle of DnaK, which is described in the next paragraph.

[4] *Lobe Ia: residues 1-37 and 120-183; lobe Ib: residues 38-119; lobe IIa: residues 184-227 and 311-368; lobe IIb: residues 228-310 [35]*

[5] *A LigPlot illustrates all residues coordinating a ligand molecule or interacting domain [65, 126].*

2.1.2 *Chaperone cycle of DnaK*

In the 1960s, Georgopoulos and coworkers discovered the Hsp70 system in *E.coli* (DnaK system), which includes the chaperone DnaK and two cochaperones DnaJ and GrpE [46, 75]. Since then, many fundamental discoveries have been made in understanding the complex mechanism of this powerful protein machinery. The action of the ATP-dependent chaperone system of DnaK can be described in a chaperone cycle (Figure 5). In this cycle, the molecular machine undergoes

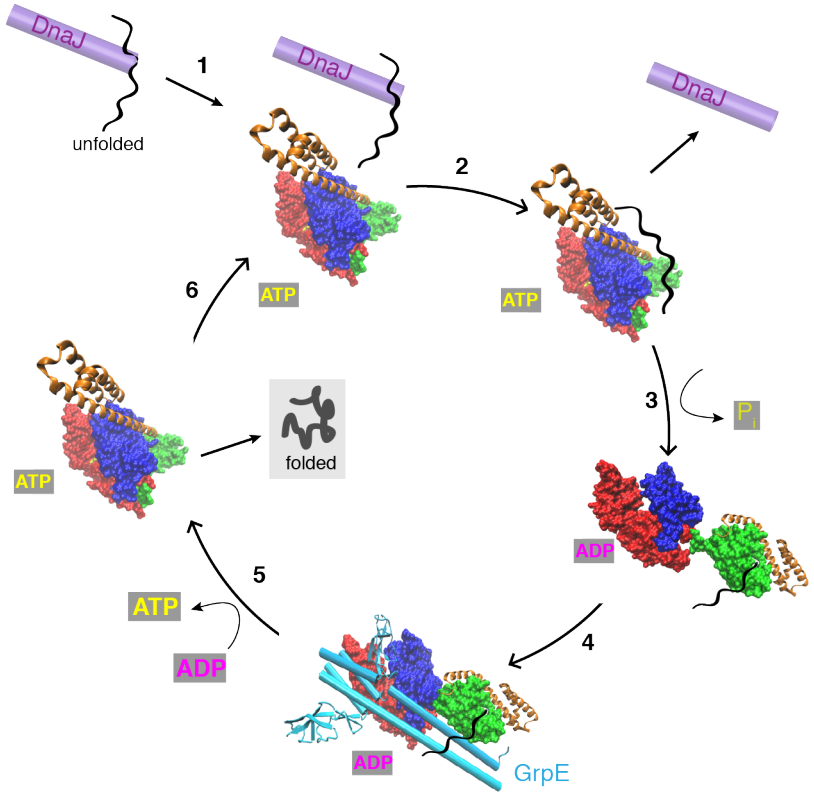


Figure 5: DnaK chaperone cycle. DnaK can bind and hydrolyze ATP to ADP. This reaction is associated with a large structural rearrangement. The cochaperones DnaJ and GrpE are regulatory tools in this cycle and influence the ATPase rate and client refolding rate of DnaK. (PDB: 2KHO (open conformation), 4B9Q (closed conformation) and 1DKG (GrpE bound to NBD).

several transitions upon ligand binding. As mentioned before, the two domains, NBD and SBD, are allosterically coupled by nucleotide-dependent motions of the SBD. In the ADP-bound state, DnaK populates an open state, where NBD and SBD are spatially more distant from each other. Upon ATP binding a closed state is achieved by binding the α-helical lid domain of the SBD to the interface of NBD [117]. The co-chaperones, DnaJ and GrpE, highly regulate the

frequency of the transitions from the open to the closed state. DnaJ and GrpE modulate DnaK's activity at specific points within the chaperone cycle to enhance its action. The chaperone cycle can be subdivided into six steps. In the first step, an unfolded client protein is bound to DnaJ and shuttled to the ATP-bound DnaK (open conformation). The substrate-binding cleft of the SBD is accessible, and the peptide exchange rate is high with a low affinity for substrates. In step two, the client is bound to the SBD and DnaJ enhances the hydrolysis of ATP to ADP. In step three, ATP gets hydrolyzed to ADP and inorganic phosphate. Hence, the lid domain detaches again from the NBD, and the substrate-binding cleft closes. At this point, the client binding site is in the high-affinity state, with a low substrate exchange rate. In step four, GrpE binds to the ADP-bound DnaK and accelerates the exchange of ADP with ATP. In step five, DnaK adapts the open conformation and the folded client protein is released. In the last step, DnaK is in the open conformation with no substrate bound to SBD and the cycle can start over. This cycle would also proceed without the action of co-chaperones, but much slower. Co-chaperones accelerate certain reaction rates in this Hsp70 cycle and make it more efficient. GrpE acts as a nucleotide-exchange factor and stimulates the release of bound ligand (step four and five). DnaJ enhances the ATP turnover rate ($0.045\,\mathrm{min}^{-1}\mathrm{DnaK}^{-1}$ to $0.57\,\mathrm{min}^{-1}\mathrm{DnaK}^{-1}$), and couples the hydrolysis more tightly with the substrate association (step one) [77].

2.1.3 *Mechanism of ATP binding, hydrolysis and nucleotide release in NBD*

In the last section, the chaperone cycle was discussed in detail. It was stated that the nucleotide state of the NBD is important for the proceeding of the cycle. Therefore, we will deal next with the detailed mechanism of adenosine binding, hydrolysis and the action of GrpE on nucleotide release.

2.1.3.1 *Nucleotide binding and hydrolysis of NBD*

The highly conserved ATP binding cavity harbors a coordinated magnesium ion and two potassium ions [88, 130] to stabilize the coordination of ATP (1:1 binding stoichiometry ATP and DnaK). ATP binding occurs in two steps. First, a weak complex forms rapidly, then a second slow step occurs, which is important for chaperone activity. In this second step, DnaK switches to the high-affinity state for substrate (open conformation of the SBD) [117]. Ligand binding includes the coordination of the adenine ring, the ribose, and the phosphate groups. The adenine ring is sandwiched between the side chains of Arg-272 and Arg-342 while adopting the anti-conformation. The ribose is bound in the 2'-endo pucker conformation in the active site [36]. NBD is not only capable of ATP-binding but can also hydrolyze ATP to ADP and

inorganic phosphate. ATP hydrolysis (γ-phosphate cleavage) is the rate-limiting step in the whole ATP-binding, ATP-hydrolysis, and ATP-release reaction. This rate-limiting step also influences the chaperone cycle of DnaK [77]. The ATPase catalytic activity of the NBD domain is $0.0015\,s^{-1}$ [117]. The dissociation constant of ATP is 0.09 µM and the off-rates of ATP and ADP are $0.019\,s^{-1}$ or $0.014\,s^{-1}$, respectively [112, 117]. These rates are one order of magnitude faster than ATP hydrolysis, and hence, the chaperone cycle depends on the hydrolysis rate of DnaK [117].

2.1.3.2 *Nucleotide release and the effect of nucleotide exchange factors*

The dissociation rate of nucleotides is the rate-limiting step for substrate release and therefore a tool for regulation. In the case of DnaK, the structure of the binding cavity is organized in a mouse-trap-like fashion [74] and includes the hydrophobic patch (L257-V59), the salt bridges (K55-E267 and R56-E264) and the long loop region in lobe IIb (A276-R302). These elements result in a tight interaction of lobe I with lobe II. Therefore, the association of ATP is rapid and the dissociation of ATP and ADP is slow [16, 74]. To enhance the dissociation of ADP, the nucleotide-exchange factor GrpE plays a critical role. This co-chaperone partially breaks the tight lobe-lobe interactions and accelerates nucleotide exchange [16]. GrpE is composed of a C-terminal globular domain and an N-terminal extended helical rod domain, which is a temperature sensitive element. The C-terminal part interacts with the NBD by inserting one β-sheet domain into the nucleotide-binding cavity and forcing the rotation of subdomain IIb by an angle of 14°. This rotation is enabled by a high degree of rotational freedom of lobe IIb [120] and results in a displacement of Ser-274, Lys-270 and Glu-267 by 2–3 Å [43]. These residues are involved in the coordination of the adenosine and play a role in the mouse-trap-like mechanism of ligand binding (salt bridge and hydrophobic patch). The disruption of these interactions results in nucleotide release.

In contrast to *E.coli* DnaK, human Hsp70 (Hsc70, pdb:4H5T) does not have the extended loop region in lobe IIb and misses one salt bridge, which results in one order of magnitude faster nucleotide dissociation rates (Figure 6) [16]. Another member of the Hsp70 family, HscA (*E.coli* Hsc66) lacks all of the ATP-trapping features and therefore has high dissociation rates and is independent of nucleotide-exchange factors [14, 105]. These are excellent examples how structural features regulate ligand exchange and how co-chaperones alter structural arrangements to fulfill specific tasks.

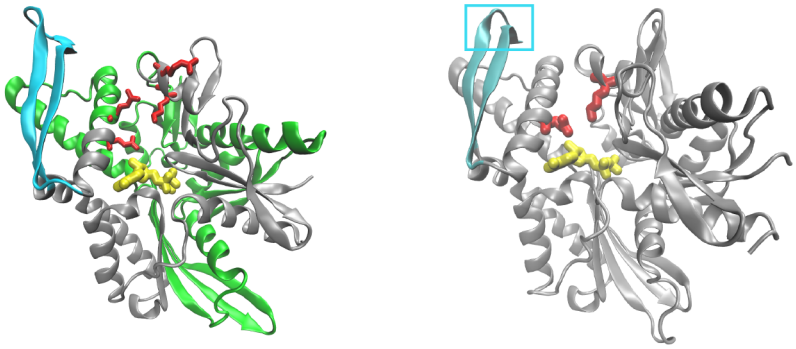


Figure 6: The ATP binding cavity is organized in a mouse-trap-like fashion. Left: Structure of DnaK NBD (pdb:4B9Q). Colored are the elements which establish a close lobe-lobe interaction network. Hence, the nucleotide (ATP in yellow) is selectively trapped in the active site with a very narrow exit tunnel (green: hydrophilic patch, cyan: loop region, red: salt bridges). Right: The structure of human Hsc70 (pdb:4H5T) shows a shorter loop region compared to DnaK (cyan) and only one salt bridge (red).

2.1.4 *Duties of DnaK within the cell*

DnaK is an ATP-dependent chaperone that acts as a foldase or a holdase and operates as a housekeeping system [24]. It is one of the three crucial chaperones in *E.coli*: DnaK, GroEL/ES and trigger factor (TF). Under all conditions, especially under stress, the presence of DnaK is essential [125]. Under non-stress conditions, the DnaK concentration in the cell is $\sim$50 µM, which is in the same order of magnitude as the ribosome concentration [47]. The action of the ribosome and DnaK are coupled. DnaK receives nascent chains from the ribosome and TF after translation. Therefore, DnaK plays an essential role in *in vivo* folding of many nascent protein chains [116]. Proteomic analysis revealed that more than 700 proteins interact with DnaK during their folding process [21]. Due to DnaK's important role in folding it is not surprising that the deletion of DnaK and TF are lethal for bacterial cells. The deletion of only TF leads to the takeover of DnaK [116]. When the interaction with DnaK does not lead to a fully folded protein, DnaK shuttles preexisting peptide chains to the GroEL/ES machinery. The interaction with the ribosome and TF as well as with the downstream chaperone GroEL/EL makes DnaK the central hub for chaperone-assisted folding [21].

DnaK does not only act as a foldase, as described above, it has also holdase activity. Holdases keep the protein chain in a more extended or unfolded state. Hence, these chains can be more easily transported through membranes. DnaK is involved in the protein export process

into the outer membrane or the periplasm. DnaK and DnaJ can held the client protein in a pre-state to keep it transportable. One example is the alkaline phosphatase, which is held in a transportable state by DnaK and DnaJ [131, 132]. After transport, the protein adopts its mature form.

Heat-shock response is a vital tool for a cell to deal with changes in the environment. DnaK and its co-chaperones interact with specific proteins under different stress conditions (e.g., heat shock or oxidative stress). Hence, making it a master regulatory element to maintain balanced growth. To do so the transcription of the heat-shock proteins must be highly regulated. DnaK regulates its transcription level in dependence of the concentration of free/unbound DnaK. This feedback mechanism includes the transcription regulator $(\sigma^{32})^6$, which associates with the RNA-polymerase and induces the translation of the DnaK machinery [28]. When the temperature increases more DnaK molecules are associated with a substrate. This decrease in free DnaK molecules and co-chaperones lead to the production of more DnaK and induces the heat shock response [20]. At 46 °C DnaK together with GroEL/ES, represent 15-20% of the total protein concentration in the cell [39, 42]. Therefore, free DnaK, GrpE, and DnaJ act as a cellular thermometer that has a negative regulatory effect on σ^{32} production and a positive effect on its degradation. DnaK and its co-chaperones are hence master regulatory elements to maintain balanced growth by quality-controlled protein homeostasis under stress conditions.

[6] *A protein needed for the transcription initiation where it mediates RNA Polymerase binding to the promotor region*

In contrast to heat shock, oxidative stress shuts the ATP-dependent DnaK system off, and Hsp33 comes into play [133]. Hsp33 activity is regulated by the redox-potential. It is located in the cytoplasm and harbors highly reactive cysteines, which coordinated a zinc ion. Under oxidizing conditions, disulfide bridges are formed, which leads to the release of the coordinated zinc and leads to the activation of Hsp33 chaperone function. Under the influence of reactive oxygen species, the ATP level in the cell decreases and hence, the ATP independent Hsp33 takes over [51]. When normal conditions are restored, ATP levels increase and DnaK can again act as a chaperone.

To sum up, DnaK is an essential heat-shock protein within the bacterial cell. It has a highly fine-tuned structure which undergoes a ligand dependent motion allowing the progression of the so-called chaperone cycle. In this cycle, the refolding of the substrate is coupled to ATP hydrolysis. The binding of Mg-ATP and its hydrolysis is executed in the NBD whereas substrate binding and release take place in the SBD. The ligand bound state is mainly sensed by lobe I of NBD and transmitted via allosteric coupling to the SBD. Lobe II is primarily involved in stable and specific ligand binding. The duties of DnaK involve the maintenance of protein homeostasis in the cell under stress and non-stress conditions. Here, the DnaK is involved in the refolding of denatured proteins and the folding of newly synthesized protein chains to pre-

vent protein aggregation (foldase activity). Last, DnaK can act as a holdase to allow protein transport through membranes.

2.2 MITOCHONDRIAL HSP70 OF S. CEREVISIAE (SSC1)

Mitochondria are endosymbiotic organelles which are crucial for every cell. They play an indispensable role in energy production, cell signaling, stress response, and apoptosis as well as the biosynthesis of metabolites. Most proteins need to be imported from the cytoplasm to become active in the mitochondrial matrix. In this import process, the class of Hsp70 chaperones plays an important role. In yeast three, essential classes of Hsp70 exist: Ssc1, Ssq1, and Ecm10. The most abundant form is Ssc1. This protein is involved in the protein import through the TIM-TOM complex and is needed for proper folding of imported proteins [53, 68, 82].

Ssc1 is essential for proper mitochondrial function. If its function is impaired or abolished, the viability of the cell shuts down [26]. Sequence analysis of the Ssc1 DNA showed that it is closely related to the DnaK in *E.coli*. Compared to DnaK, Ssc1 has an N-terminal mitochondrial import sequence [27]. DnaK carrying this leader sequence is, like the Ssc1, imported into the mitochondrial matrix [13]. This shows that only the leader sequence is necessary for the import into the mitochondrial matrix.

In vivo and *in vitro*, Ssc1 needs the co-chaperone Hep1 to sustain its structure and function. Hep1 neither influences the ATPase activity of Ssc1 nor interacts with Ssc1 at high ATP levels. Therefore, Hep1 must act as a chaperone for Ssc1, ensuring folding of Ssc1 in the mitochondrial matrix [98, 104]. Hep1 interacts with the nucleotide-binding domain and the interdomain linker of Ssc1. This interaction prevents aggregation of Ssc1 and induces folding. Adenosine binding to the folded NBD of Ssc1 triggers Hep1 release [11, 13].

Ssc1, like other Hsp70s, ligand-dependent cycles between ATP and ADP states [48, 84]. Single-molecule Förster resonance energy transfer (FRET) studies showed that Ssc1 adapts a heterogeneous ADP-bound state and a uniform ATP-bound state. This phenomenon has not been observed for the DnaK system [73]. Similar to DnaK, Ssc1 also needs the action of co-chaperones. In the case of the mitochondrial Hsp70, the nucleotide-exchange factor is called Mge1 [81] and the associated J-domain co-chaperones are Mdj1 and Tim14. Mdj1 aids for the protein folding process [30], while Tim14 assists in the translocation process [119].

To sum up, Hsp70 in mitochondria (Ssc1) is vital for the proper folding of mitochondrial proteins and is involved in the protein translocation via the TIM-TOM-complex into the mitochondria matrix. Ssc1 needs the action of its co-chaperone Hep1 to be folded and fully func-

tional in the mitochondrial matrix. Like other Hsp70 proteins, it is an ATP dependent chaperone and associates with J-domain co-chaperones and nucleotide-exchange factors.

2.3 THE SUGAR-KINASE FAMILY: STRUCTURAL FEATURES AND EVOLUTION

Proteins are classified in a variety of protein super-families, according to either sequence conservation, functional tasks or structural similarities. In all cases, a common evolutionary origin is the starting point for a particular protein super-family, a so-called common ancestor. From this common ancestral domain, further sub-families have evolved and specialized to distinct tasks in the cell.

2.3.1 *Structural feature: highly conserved butterfly-like fold*

All Hsp70 NBDs, actin, and hexokinase are members of the sugar-kinase family (Figure 7, top). The structural characteristics of this family are their butterfly-like 3D structure and a common ATP-binding site. Here, the lobe-lobe domain architecture, the lobe-lobe connection, and the ATP-binding loops are highly conserved. The amino acid sequence follows the architecture: lobe Ia — inserted lobe Ib — lobe IIa — inserted lobe IIb (brown, Figure 7). In both lobes, subdomain b is pasted into subdomain a, arguing for a domain-insertion event during evolution. To connect lobe I with lobe II, the C-terminal helix, arising from lobe II, is structurally inserted into lobe I (brown in Figure 7). Furthermore, a loop region between lobe I and lobe II connects both lobes with each other (light green in Figure 7). The ATP-binding site comprises three highly conserved ATP-coordinating loops, which build the active site (green in Figure 7) [6, 35, 36]. In summary, the sugar-kinase family is characterized by its lobe-lobe architecture, which is stabilized by two structural features (C-terminal helix and loop region). Moreover, the ATP-binding site is highly conserved in all members.

2.3.2 *Evolution of the sugar-kinase family*

As mentioned above, the sugar-kinase family share three major structural features and have the same lobe-lobe architecture. Hence, Bork *et al.* proposed a common ancestral homodimeric ATP-binding domain to be the evolutionary origin of this protein family [6]. This domain further developed by gene duplication and divergent evolution, into the lobe I and II structures we know today. From here, two main branches emerged: the sugar kinases and the actin/heat-shock protein branch. In the sugar kinase branch, lobe I and II developed in a very different way compared to the actin/heat-shock protein branch. Hence, the

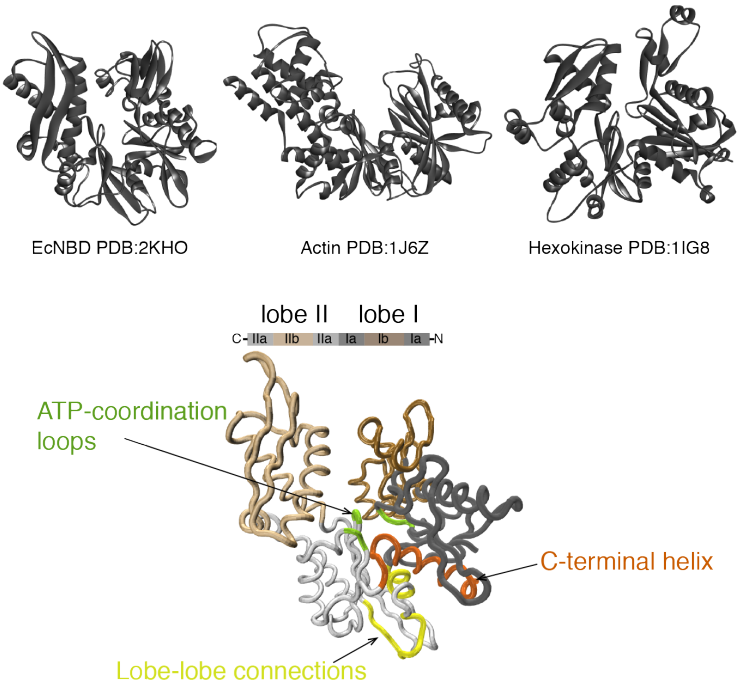


Figure 7: Common structural features of the sugar-kinase family members. The ATP-coordination loops are green, the C-terminal helix is orange, and the lobe-lobe connection loop is yellow. These three elements are common features conserved in the sugar-kinase protein family as well as the lobe I and lobe II domain structure.

similarity between Hsp70 NBD and actin is more evident than Hsp70 NBD and hexokinase. Most of these significant differences occurred in regions that are more specific for the function of the protein. The ATP-binding site stayed conserved.

To sum up, Hsp70 NBD is a member of the sugar-kinase family. All members share the common butterfly-like fold and bind Mg-ATP in the highly conserved binding pocket mainly located in lobe II. From an evolutionary point of view, the sugar-kinase family is thought to have developed by a common ancestral domain, via gene duplication and divergent evolution. This common ancestral ATP-binding domain was able to dimerize. One lobe adapted to maintain a specific task. The other is mainly involved in ATP binding. All members of this super-family undergo a conserved ligand-dependent domain motion and adopt an open or closed form by domain rotation [49].

3

OPTICAL TWEEZERS: A TOOL TO STUDY PROTEIN FOLDING AT THE SINGLE-MOLECULE LEVEL

Investigating the protein folding pathway is one of the central questions in structural biochemistry and biophysics. The understanding of protein folding would give the power to design proteins with a particular function. These designed proteins could, for example, be used in therapeutics to deliver drugs. Until now, the protein folding problem has not been completely solved, especially for large protein complexes harboring many functional domains. Artificially designed structures are often rather rigid and less dynamic than naturally designed ones[7, 23, 96]. To study protein folding and their native structures, researchers use common techniques including spectroscopic methods, NMR, X-ray crystallography or SAXS. These techniques focus on an ensemble of molecules and have the disadvantage that fast phases or rare events are often hidden. Single-molecule approaches such as, Förster Resonance Energy Transfer, Atomic Force Microscopy (AFM) or optical tweezers, easily detect rare events. The power of optical tweezers lies in their high temporal (micro-seconds) and spatial resolution (nanometer and piconewton). This method provides insights into unfolding and refolding pathways of proteins, domain motions and ligand-binding events [41, 97, 101, 107].

James C Maxwell theoretically showed in 1864, that light can exert a force on a particle, but he could not validate his theory experimentally [69]. This validation was possible after the invention of laser technology in the 1960s. In 1970, Ashkin demonstrated experimentally that radiation pressure can stably trap particles [4]. Based on this ground-breaking work on radiation pressure, it was possible to develop more advanced experiments. The first biological application of optical tweezers was the manipulation of viruses and motile bacteria [5]. Later, macro-molecules like RNA or DNA were investigated in the light of their polymer behavior [127, 128]. More recently, the folding behavior of proteins [50, 55, 110, 138], protein-ligand interactions [107] or protein-protein interactions [41] were investigated using single-molecule optical tweezers. Even the investigation of whole protein machineries [67, 85, 93, 118] or the action of enzymes is now possible [91].

The physical basis of optical tweezers is that a dielectric particle is trapped by a highly focused laser beam, which travels through a microscope objective. This trapped particle experiences two forces: a scattering force and a gradient force (Figure 8). The scattering force acts along the direction of the laser beam. To realize 3-dimensional

trapping, the laser beam has to be highly focused using a high numerical aperture microscope objective. The gradient force pushes the particle to the point of highest laser intensity. Finally, the bead is held in place because the vector of the scattering and gradient force are equal. To sum up, a trapped particle in a highly focused laser beam experiences two forces: a scattering force and a gradient force. Both force components cancel out and hence hold the particle trapped in all three dimensions.

To describe the gradient force different models can be used. Is the wavelength of the incident light in the same range or smaller as the particle size, the Mie Scattering is prominent. For scattering at, for example, molecules, the Rayleigh Scattering model applies. The diffraction of the laser beam at the bead surface, which has a higher refractive index than the surroundings, transfers momentum to the bead. This transfer of momentum acts towards the center of the Gaussian intensity distribution of the laser beam.

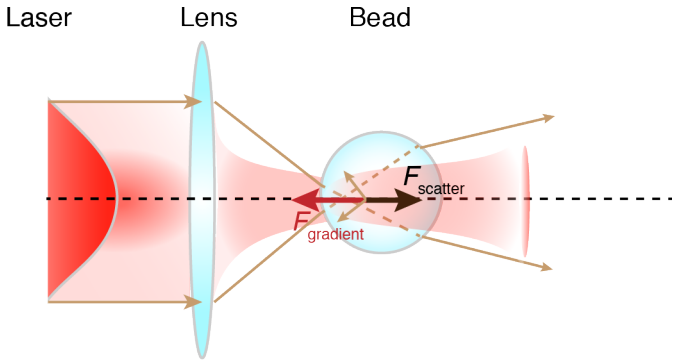


Figure 8: Optical trap in a ray optical view.

When the bead is displaced from the trap center, a restoring force acts on it. The trap behaves in this case, in approximation, like a Hookean spring ($F = k \times x$). Therefore, at small displacements, the restoring force is proportional to the bead deflection. When the displacement and the stiffness of the trap are known, the determination of the acting force is possible. The force range from a typical optical tweezers setup is 1 to 100 pN. The stiffness of an optical tweezer is in the range of about $\approx$ 0.04–0.7 pN nm^{-1} allowing to investigate small forces and length changes [9, 19, 85, 138].

A dual beam optical tweezers system can be used to study single proteins. To do so, the protein of interest has to be fused via DNA to beads. This beads can be trapped in one of the laser beams Figure 9. By moving one beam, the bead follows and force is exerted onto the protein molecule. In this dumbbell geometry different experiments

can be performed to probe for protein stability, ligand binding effects, folding pathways, protein-protein interactions and many more. In this thesis the un- and refolding pathway of the DnaK NBD protein is investigated as well as the ligand binding and its effect on the protein stability. For further details see Section 13.5.1.

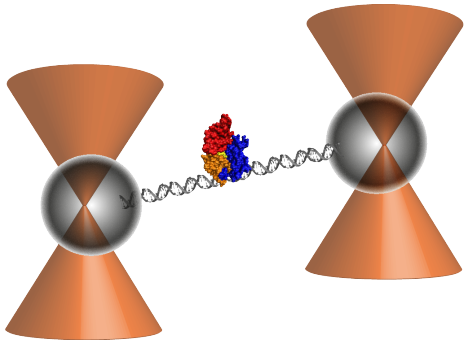


Figure 9: Dumbbell geometry in optical tweezers measurements. Two laser beams (orange) trap two functionalized beads. The protein in the middle (blue, red and orange) is tethered to the beads via DNA-handles (gray). By moving one bead force is exerted onto the protein molecule.

In summary, optical tweezers is a valuable tool to study protein folding and protein interactions. The physical principle of optical trapping is that micron-sized beads can be trapped and positioned in all three dimensions in a highly focused laser beam. This stable positioning is possible because the scattering force and the gradient force cancel each other out. Assembly of two laser traps enables the application in biophysics to study single-molecule protein unfolding and refolding as well as ligand binding events. Here, it is necessary to be able to displace one trap in order to exert a force on the protein dumbbell[7]. With the knowledge of the trap stiffness and bead displacement, the force can be calculated.

[7] *The protein dumbbell is in more detail described in Section 13.5.1.*

Part II

RESULTS

4

UNFOLDING PATHWAY OF ECNBD

This chapter presents the results obtained for the unfolding pathway of *E.coli* NBD (EcNBD)[8] and the influence of ligand binding on the domain hierarchy. Results presented in this chapter are already published partly in my master's thesis and detailed in Bauer et al. [9] and therefore will be presented very briefly. To study unfolding pathways of proteins at the single-molecule level, the protein is fused to silica beads, which can be manipulated by highly focused laser beams (for details see Chapter 3). Briefly, to assemble a so-called dumbbell, the EcNBD protein is fused to two DNA-handles, which are connected with the beads surfaces. In the following experiments, force is applied to the EcNBD protein N-/C-terminally (EcNBD-nc).

[8] *EcNBD construct for trap measurements: MACKGS-NBD(1-394)-C15A-SKCL-H6.*

4.1 TYPICAL UNFOLDING PATTERN OF ECNBD-NC

EcNBD-nc can be unfolded and refolded several times in constant velocity experiments ($20\,\mathrm{nm\,s^{-1}}$)[9]. Natively folded *E.coli* NBD n-/c-terminal pulling direction (EcNBD-nc) protein results in a high force single step unfolding. Figure 10A shows the typical unfolding pattern of the natively folded protein. First, the dumbbell is stretched out and, at a distance of roughly 350 nm, the force increases. This first stretching curve follows the typical DNA polymer behavior discussed in Section 13.5.3.1. Further increasing the distance increases the force on the protein and an abrupt rip can be observed at ≈30 pN (Figure 10A and Figure 10B). This rip corresponds to the unfolding of the natively folded EcNBD-nc protein. Unfolding appears to happen in a single step to the fully unfolded peptide chain. The event results in a contour-length increase of ≈134 nm which is consistent with the complete unfolding of a protein of 383 amino acids in size[10]. Upon closer inspection, the unfolding event shows short-lived intermediates which are populated during the unfolding process. We could determine two major unfolding intermediates ($I_1^{apo}$and I_2^{apo}Figure 10C). These intermediates can be characterized by their contour lengths (L_C) and lifetimes (τ) and are listed in Table 1.

[9] *In constant velocity experiments, the resulting curves are called stretching and relaxing curves. The pulling velocity and the start-and-end distance for each pull are set manually.*

[10] *The length of one amino acid is* 0.35 nm.

The two well-populated intermediates $I_1^{apo}$and $I_2^{apo}$are well resolved in the experiments[11]. Sometimes also supplementary intermediates can be detected in the data sets, but those are very short-lived and close to the temporal resolution limit (blue arrows in Figure 10D). In the following, the focus will be on the well-populated intermediates $I_1^{apo}$and I_2^{apo}.

[11] 30 kHz *sampling rate results in a lifetime resolution of* >200 µs.

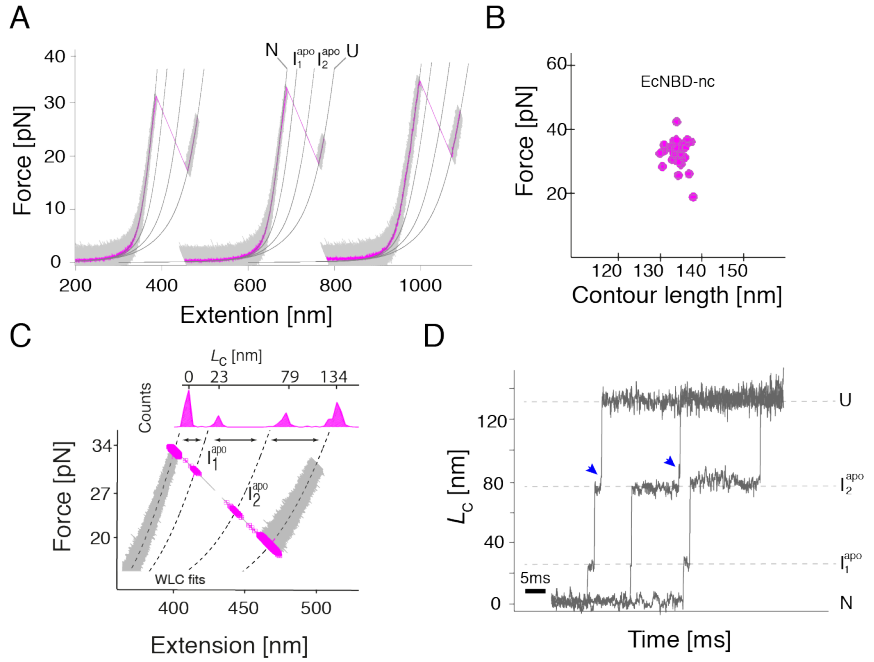


Figure 10: Unfolding of the natively folded EcNBD protein. (A) Typical force-extension unfolding traces of natively folded NBD (recorded with a pulling velocity of 20 nm s^{-1}. Only stretching cycles are shown). The gray traces are the raw data recorded with an acquisition frequency of 30 kHz and the pink traces are smoothed traces (box-car smoothing: 20). (B) Force vs contour length plot of the unfolding event. EcNBD-nc unfolds with a mean force of 32.7 ± 3.5 pN and a contour-length gain of 133.9 ± 2.6 nm. (C) Zoom into the unfolding intermediates. Raw data are shown in gray. An overlay of many unfolding events is shown in pink. Top: unfolding contour-length distribution. (D) In the shown contour length versus time plot the populated intermediates are very short lived (range of milliseconds). The two well resolved intermediates I_1^{apo} and I_2^{apo} have contour lengths of 23.5 ± 1.0 nm and 79.0 ± 1.2 nm, respectively.

State	L_C**[nm]**	$N_{\textbf{amino acids}}$	$\tau[\mathbf{s}^{-1}]$
Native (N)	0	0	N.A.
I_1^{apo}	23.5 ± 1.0	60	0.8 ± 0.2
I_2^{apo}	79.0 ± 1.2	160	11 ± 3
Unfolded (U)	134.0 ± 1.0	376	N.A.

Table 1: List of contour length and life times of the unfolding intermediates of EcNBD-nc.

The unfolding always follows the same pattern, indicating an obligatory sequence of unfolding events populating I_1^{apo} and then I_2^{apo}. Mapping the unfolding events to structural parts of the protein is not

straightforward. The contour-length gains of the unfolding events ΔL_C (I_1^{apo}, I_2^{apo}) and ΔL_C(I_2^{apo}, U) are of very similar size (≈55 nm). From a structural point of view, this is not surprising due to the lobe-lobe architecture of EcNBD: lobe I and lobe II are of similar size. Those two domains might act as individual unfolding units. To test this hypothesis, we designed loop insertion variants which make it possible to assign the structural element related to each unfolding event.

4.2 STRUCTURAL MAPPING OF UNFOLDING INTERMEDIATES

Data obtained with optical tweezers is susceptible to contour-length changes. With a 20-residue-long unstructured Gly-Ser-rich loop (L_C= 7 nm) it is possible to elongate chosen regions of the protein structure to map the unfolding pathway. This loop was inserted in pre-existing loops of EcNBD-nc. We designed four insertion variants of EcNBD-nc: Ins45, Ins183, Ins290, Ins364 (Figure 11A). For Ins45, we did not see native refolding within the experimentally accessible timescale. All other variants showed native-like folding behavior. Unfolding of the region that included the loop insert led to a 7 nm longer contour length. The unfolding lengths and lifetimes are listed in Table 2. By comparing contour lengths of loop insertion variants with EcNBD-nc, it is possible to dissect the structural unfolding units. The contour length versus time plot in Figure 11B revealed first of all that all loop-insertion variants are 7 nm longer than EcNBD-nc (L_C≈141 nm), as expected. To dissect in which intermediate the longer contour length can be detected, we looked closer into each single unfolding event and compared them with the corresponding events in EcNBD-nc.

Name	L_C (I_1^{apo}) [nm]	τ (I_1^{apo}) [s^{-1}]	L_C(I_2^{apo}) [nm]	τ (I_2^{apo}) [s^{-1}]
EcNBD	23.5 ± 1.0	0.8 ± 0.2	79.0 ± 1.2	11 ± 3
Ins364	31.4 ± 0.6	0.04 ± 0.004	83.8 ± 0.8	3.2 ± 0.9
Ins290	24.0 ± 1.0	0.5 ± 0.03	80.4 ± 0.7	6.1 ± 0.7
Ins183	31.8 ± 0.7	0.3 ± 0.03	83.1 ± 1.0	7.6 ± 0.7

Table 2: List of contour lengths and lifetimes of insertion variants (apo).

Ins364 was designed to probe for the unfolding of the c-terminal part of NBD and showed a 7 nm longer unfolding event in I_1^{apo}. Form this it can be concluded, that the c-terminal part unfolds first. Also, Ins183 showed a 7 nm more extended I_1^{apo}intermediate, arguing for a concerted unfolding of the c-terminal part of NBD and the separation of lobe I and lobe II. Ins290 shows a prolonged I_2^{apo}event meaning that lobe II unfolds in that step. In conclusion, first, the c-terminal

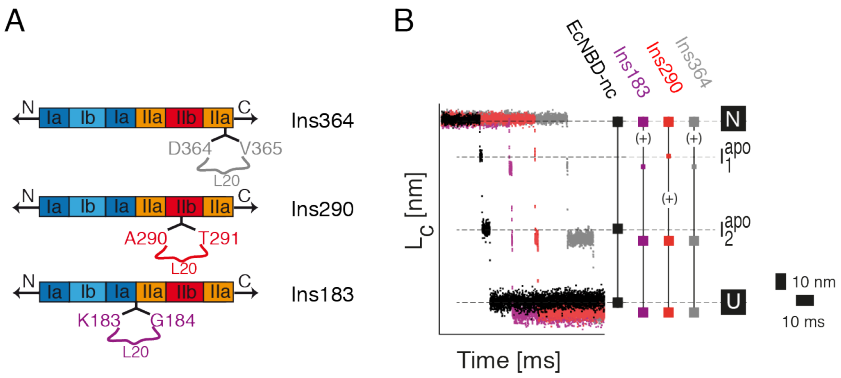


Figure 11: Structural mapping of unfolding events under apo conditions. (A): illustration of the design of the loop insertion variants of NBD. At each position (Ins364, Ins290 and Ins183) a 20 amino acid unstructured Gly-Ser-rich-loop is inserted. (B): Unfolding traces of the insertion variants in a contour length vs time plot. The EcNBD-nc unfolding (black) is used as a reference to show which intermediate gets prolonged due the the loop insertion. All insertion variants show a longer unfolded contour length as expected. Ins183 shows a longer I_1^{apo}as well as Ins364 and Ins290 shows a longer I_2^{apo}unfolding event.

helix unravels, followed by the separation of the lobes. Second, lobe II unfolds and lobe I unfolds in the last step. The whole unfolding pathway derived with these experiments is summarized in Figure 14.

4.3 LIGAND BINDING SHUFFLES THE DOMAIN HIERARCHY

As introduced earlier, the NBD of DnaK is able to bind ATP with nanomolar affinity and hydrolyzes it to ADP [117]. In a next set of experiments we were curious to see if ligand binding affects the stability of the protein and hence can change the hierarchy of the unfolding pathway. We did the same constant velocity unfolding experiments as before, now in the presence of 1 mM ATP or ADP and 5 mM $MgCl_2$ (holo-state). We found that the unfolding forces are not affected by the presence of ligand: $F_{unf,\,apo} = 33.1 \pm 4.2$ pN (n=73), $F_{unf,\,ADP} = 32.1 \pm 4.0$ pN (n=53) and $F_{unf,\,ATP} = 32.7 \pm 3.5$ pN (n=77) (Figure 12B). However, a closer look into the unfolding event reveals three major unfolding intermediates: I_1^{holo}, I_2^{holo}, and I_3^{holo}. The data obtained for the EcNBD-nc and the insertion variants are shown in Figure 13.

A comparison of the unfolding contour length in apo and holo conditions shows that the ligand affects the stability of specific substructures within the NBD. As a consequence, also the unfolding pathway changed in the presence of ligand (Figure 12A). However, the unfolding forces are not affected by the bound ligand (Figure 12B)

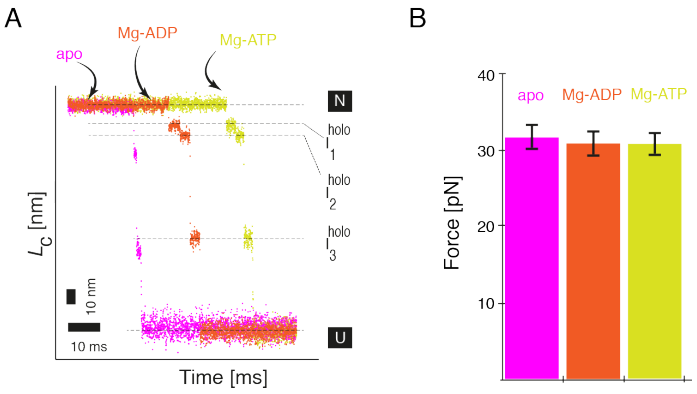


Figure 12: Unfolding events under holo (ATP and ADP) conditions in comparison with the apo unfolding. (A) In pink the unfolding sequence of apo EcNBD-nc is shown. In comparison Mg-ADP^{-} (orange) and Mg-ATP^{2+} (yellow) are shown. Both holo forms show the same unfolding pattern. This pattern is different from the apo unfolding pattern because additional unfolding intermediate get populated. (B) Unfolding forces of apo (pink), Mg-ADP^{-} (orange) and Mg-ATP^{2+} (yellow) are plotted.

To decipher the unfolding pathway in the presence of ligand, we again used the insertion variants to map the unfolding substructures along the unfolding pathway (Figure 13). The first unfolding event is again the unfolding of the c-terminal part (Ins364) followed by the separation of the two lobes (Ins183). In contrast to the apo form, in the holo unfolding pathway, the unfolding of the c-terminal part and the lobe separation occurs separately. In the second step, lobe I unfolds (Ins45) and in the third step lobe II unfolds (Ins290).

Next, we wanted to know if the hydrolysis state plays a role. We performed experiments not only in the presence of ATP but also with ADP (Figure 12A). Both unfolding patterns match up, without significant difference, indicating that only the occupation of the ligand binding site leads to the shift in the domain hierarchy. The hydrolysis state of ATP plays a minor role.

4.4 SUMMARY OF THE UNFOLDING PATHWAY OF ECNBD

The unfolding of apo EcNBD occurs in a sequential pathway populating two intermediates ($I_1^{apo}$and I_2^{apo}). In the case of the ligand-bound unfolding of EcNBD, three intermediates are populated (I_1^{holo}, I_2^{holo}, and I_3^{holo}). To decipher the exact unfolding pathway, loop insertion variants of EcNBD were designed and helped to dissect the structural parts involved in the unfolding process. Both pathways (apo and holo) are summarized in Figure 14. Ligand binding significantly alters the domain hierarchy of the unfolding pathway by stabilizing lobe II

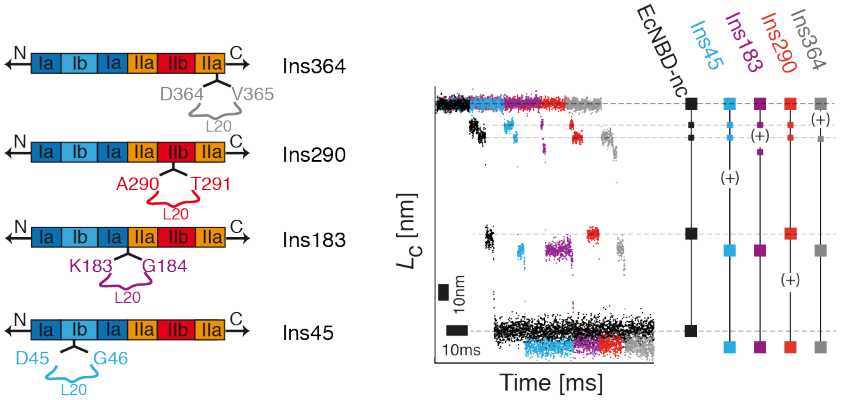


Figure 13: Structural mapping of unfolding events under holo conditions. The unfolding contour length of all insertion variants and EcNBD-nc are plotted over time. Again, the EcNBD-nc unfolding curve is a reference for determining which unfolding event of the insert variants in longer due to the additional loop length. Ins364 shows a longer I_1^{holo}, Ins183 has a longer I_2^{holo}, Ins45 a longer $I_3^{holo}$and Ins290 shows a longer unfolding to the unfolded state. With this information we can show that first, the c-terminal helix is unfolding then the lobes are separated followed by the unfolding of lobe I, and finally lobe II unfolds.

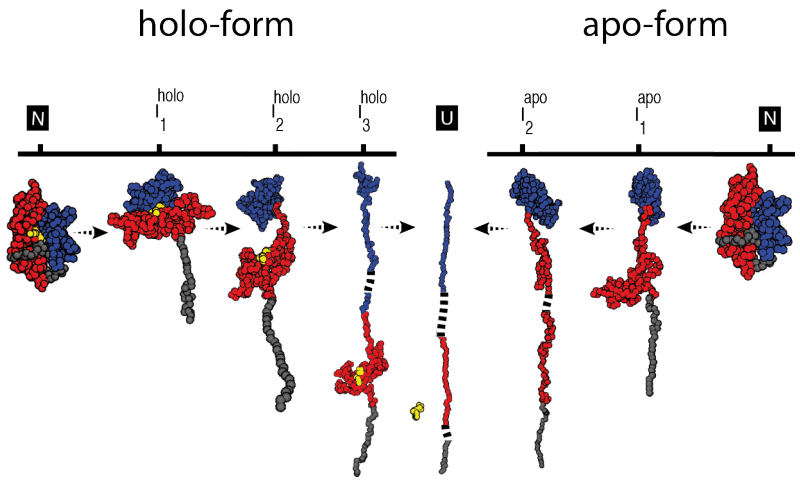


Figure 14: Unfolding pathway of NBD in apo and holo conditions. The unfolding pathway in apo follows the sequential unfolding of the c-terminal helix and separation of the lobes. Second, lobe II unfolds, and last, lobe I unfolds this results in a fully unfolded peptide chain. The presence of nucleotide shifts the domain hierarchy and hence alters the unfolding pathway. Again, the c-terminal helix unfolds, followed by the lobe separation and unfolding of lobe I. Lobe II is in this unfolding pathway stabilized due to the interactions with nucleotides and unfolds in the last step.

which leads to a switch in the unfolding pattern. Both pathways have in common that the c-terminal helix is the critical structural element for the high stability of EcNBD and acts as a molecular glue between lobe I and II. Furthermore, the separation of lobes is in both pathways present and in the holo case happening in a disjointed step and in the apo case concerted with the unfolding of the c-terminal helix.

5

REFOLDING PATHWAY OF ECNBD

The unfolding of a natively folded protein was discussed in detail in Chapter 4. In this chapter, the focus is on the refolding process and the involved refolding intermediates. Optical tweezers, enable the recording of multiple un- and refolding events of one protein molecule. Furthermore, rarely populated intermediates and misfolded states, which may or may not lead to natively folded proteins, can be detected. The obtained results show that the refolding process of EcNBD involves two fast forming on-pathway intermediates.

5.1 ECNBD REFOLDS VIA TWO FAST FORMING INTERMEDIATES

Optical tweezers experiments are not only able to decipher the unfolding pathway of a protein, but they can also provide insights into the refolding path of a protein. EcNBD-nc was unfolded and, by gradual force release, parts of the structure were allowed to refold under load. Figure 15A depicts a typical unfolding and refolding cycle of EcNBD-nc. The stretching of the protein is colored in dark brown, and the relax cycle is colored in light brown. After unfolding the protein at high force, parts of the structure can refold under force, as the relaxation cycle shows (two humps in Figure 15A).

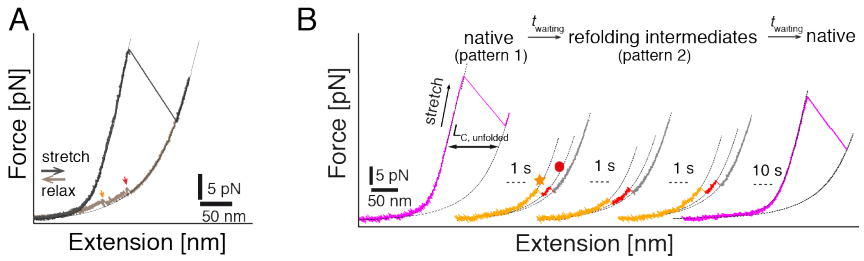


Figure 15: EcNBD unfolding and refolding sequence. (A) The first constant velocity cycle of an EcNBD-nc experiment. The stretching cycle shows the unfolding of the natively folded protein (≈30 pN). The relaxing cycle shows that parts of the unfolded chain start to refold in the lower force regime (red and orange arrow). (B) Illustrated are the two major patterns seen in the stretching cycles of EcNBD-nc. Pattern 1: Unfolding of the natively folded protein (pink). Pattern 2: Unfolding of partially refolded structural parts of NBD (orange and red parts). The waiting time between each stretching cycle was 1 s and the pulling velocity 200 nm s^{-1}.

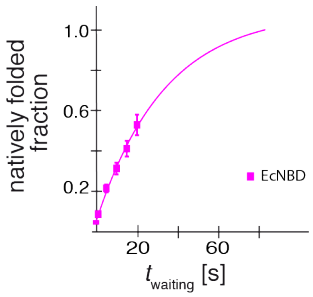


Figure 16: This plot shows the natively folded fraction of NBD dependent on the waiting time at zero force. The data can be fitted with a single-exponential fit to obtain the folding rate of natively folded EcNBD-nc (pink), which yields $0.03 \pm 0.02\,s^{-1}$.

The analysis of many stretching curves reveals two main patterns (Figure 15B). Pattern 1 (pink) is the unfolding of the natively folded EcNBD protein, as discussed in Chapter 4. Pattern 2 is the unfolding of partially refolded parts of EcNBD (orange and red, Figure 15B). Two fast forming refolding events are determined: RFI1 (red) and RFI2 (orange). Folding to the native state takes a long time and only occurs at zero load. The refolding time to the native state can be determined by performing unfolding-refolding experiments with different waiting times between cycles. By plotting the natively refolded fraction against waiting time, the refolding time is calculated by fitting the data with a single-exponential equation (Figure 16). The mean refolding rate of natively folded EcNBD-nc is $0.03 \pm 0.02\,s^{-1}$.

Next, the focus will be on the fast folding portions RFI1 and RFI2 and their characterization according to unfolding forces, contour lengths and structural assignment. The overlay of pattern 1 and pattern 2 clearly shows that only half of the structure is refolded and the other half (73 nm) remains unfolded (Figure 17A). The contour lengths of the well-defined refolded parts are $L_C = 73.0 \pm 7.2$ nm for RFI2, $L_C = 110.1 \pm 4.7$ nm for RFI1 and $L_C = 134.3 \pm 1.9$ nm for the unfolded protein (Figure 17A, gray). RFI1 and RFI2 typically unfold at 6.7 ± 2.0 pN and 4.7 ± 2.0 pN, respectively. Plotting the unfolding forces against the contour length of several pulls and different molecules reveals the reproducibility of the refolding process (Figure 17B).

Furthermore, the order of these refolding events is always the same. Having a look at several stretching curves of one molecule (Figure 18), one can see that sometimes only RFI1 (red) refolds. The case where RFI2 refolds on its own was never observed. This observation led us to conclude that the order of the refolding events follows a sequential order: U — RFI1 — RFI2. Occasionally, refolding to a state called RFI3 is visible (blue, Figure 18) and will be discussed in Section 5.2.2.

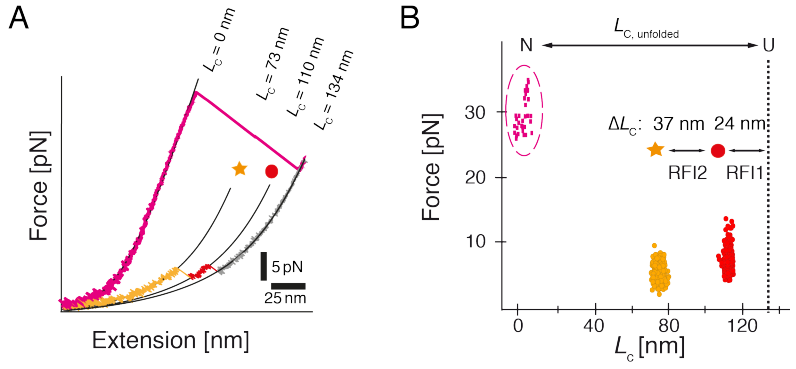


Figure 17: Characterization of the EcNBD-nc refolding intermediates. (A) In pink, the native unfolding of EcNBD-nc is shown (pattern 1). The overlay of pattern 1 (pink) with pattern 2 (orange, red) shows that both patterns result in a fully unfolded protein chain (gray). The first WLC fit for the DNA handles was kept constant. The contour length of the refolding intermediates can be determined: $L_C = 73.0 \pm 7.2$ nm and $L_C = 110.1 \pm 4.7$ nm for RFI1 and RFI2, respectively. (B) Force versus contour length plot of 10 molecules. In pink the native (N) high force unfolding of EcNBD-nc is plotted. The refolding intermediate RFI2 is shown in orange and RFI1 in red. The contour-length gain (ΔL_C) of RFI2 $\approx$ 37 nm and of RFI1 is $\approx$ 24 nm. The mean unfolding force of RFI2 4.7 ± 2.0 pN and of RFI1 is 6.7 ± 2.0 pN. The unfolded peptide chain (U) has a contour length of 134 nm, consistent with a fully unfolded structured part of NBD (residues 4-380).

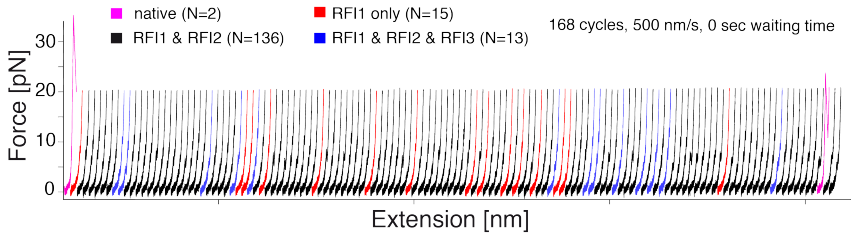


Figure 18: Consecutive stretching curves for EcNBD-nc in constant velocity mode (500 nm s^{-1}, $t_{waiting}$= 0 s). Repeatedly stretching and relaxing the EcNBD-nc protein results in the displayed traces. First, the natively folded protein unfolds (pink, pattern 1). In the next cycles, partially folded parts unfold and can be categorized in the unfolding of RFI1 and RFI2 (black, pattern 2). Occasionally, two other patterns can be observed: unfolding of only RFI1 (red) and a third refolding intermediate, RFI3 (blue).

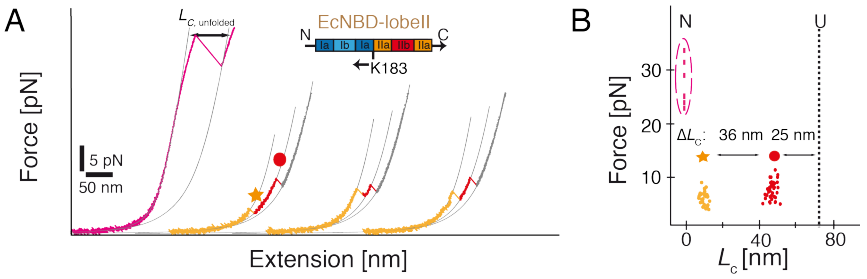


Figure 19: Constant velocity data of EcNBD-lobeII pulling variant. (A) EcNBD-lobeII pulling variant with attachment sides located at the c-terminus and at position K183C. Unfolding of the natively folded protein at high forces (pink). Refolding events are colored in orange and red. (B) Plot of force against contour-length. The contour length gain of RFI1 is 25 nm and RFI2 is 36 nm. The completely unfolded protein chain results in a contour length of ≈ 70 nm.

5.2 STRUCTURAL AND KINETIC CHARACTERIZATION OF THE REFOLDING PATHWAY

5.2.1 *The refolded portion corresponds to lobe II*

The question now arising is: which structural parts of EcNBD do refold? The refolded length of RFI2 is $L_C = 73.0 \pm 7.2$ nm, which corresponds to 174 amino acids[12]. The structure of NBD comprises two lobes which have each the size of roughly 174 residues. To determine which lobe refolds, an EcNBD variant was designed where only lobe II is under load. The attachment sides are at the C-terminus and position K183C (EcNBD-lobeII). A typical unfolding cycle of EcNBD-lobeII shows Figure 19A. The first pull shows the native unfolding of the protein with a contour length of ≈ 70 nm, as expected[13]. In the native fold the c-terminal helix is still inserted into lobe I and has to be pulled out before any further unfolding can happen[14]. This helix results in the high unfolding force of the natively folded protein. The consecutive stretching cycles show the unfolding of refolded portions. This pattern is similar to pattern 2, we determined for the EcNBD-nc variant. The unfolding forces versus contour-length plot shows two distinct intermediates (RFI1 and RFI2) (Figure 19B). EcNBD-lobe II unfolding events have the same contour-length increases and unfolding forces for RFI1 and RFI2 as for EcNBD-nc (Figure 17B). This similarity leads to conclude that the refolding intermediates include the entire lobe II region.

So far we showed that lobe II is the structural part which refolds and is under load. Longer waiting times and zero load, allow the correct positioning of lobe I and the c-terminal helix. Now, the question remains: which structural parts are incorporated in RFI1 and RFI2? To answer this question we again used the loop insertion variants of

[12] $(134 - 73\,\text{nm})/0.35\,\text{nm}\,aa^{-1} = 174\,aa$

[13] $200\,aa \cdot 0.35\,\text{nm}\,aa^{-1} = 70\,\text{nm}$

[14] *In the unfolding pathway analysis (Chapter 4) the c-terminal helix acts as the glue between lobes and has to unfold first to allow further unfolding steps.*

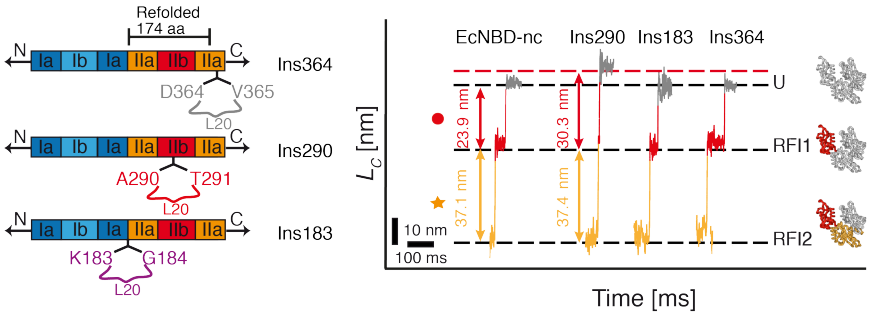


Figure 20: Loop insertion variants of EcNBD help to decipher the complete refolding pathway. Ins290, Ins183, and Ins364 helped to investigate the regions and boundaries of the refolding events. Only Ins290 shows a longer RFI1 event compared to EcNBD-nc. Hence, lobe IIb refolds first. The refolded regions depicted on the right show refolded parts in color.

EcNBD-nc (Ins183, Ins290, and Ins364)[15]. Unfolding or refolding of the structural region around a loop insertion results in a 7 nm longer event compared to EcNBD-nc. The contour-length transformations for the refolding sequence of the EcNBD variants are shown in Figure 20.

Comparing the contour-length changes of the variants with EcNBD-nc reveals that only Ins290 shows a longer unfolding event of 7 nm in the RFI1 state. Hence, RFI1 corresponds to the refolding of lobe IIb. Ins183 and Ins364 do not show a more extended event either in RFI1 nor in RFI2 but show the expected longer total contour length. These two variants determine the boundaries of the refolded portions. Therefore, the c-terminal helix as well as lobe I refold slowly.

To sum up, RFI1 corresponds to folded lobe IIb, and RFI2 includes the already folded lobe IIb plus the structured part of lobe IIa without the c-terminal helix. A schematic representation of the whole refolding pathway is depicted in Figure 21. This illustration also shows that from a structural point of view, it is reasonable to refold lobe IIb first, because the amino acid sequence of lobe IIb is inserted into lobe IIa. When lobe IIb folds, the ends of lobe IIa get closer which enables folding of lobe IIa. Further refolding to the native state includes refolding of lobe I and proper positioning of the c-terminal helix. This last step needs a long time and zero force to proceed and is the rate-limiting step in the refolding path.

[15] *An unstructured 20 amino acid long Gly-Ser-rich loop was inserted in pre-existing loop structures of EcNBD. See also Chapter 4.*

5.2.2 *Rarely occurring events in the refolding pathway*

Figure 18 shows consecutive stretching cycles which expose besides the already discussed patterns (pattern 1 and pattern 2), a third pattern. Pattern 3 (blue) includes the refolded portions of RFI1 and RFI2 plus a refolded part of lobe I. Figure 22A depicts, the stretching cycles

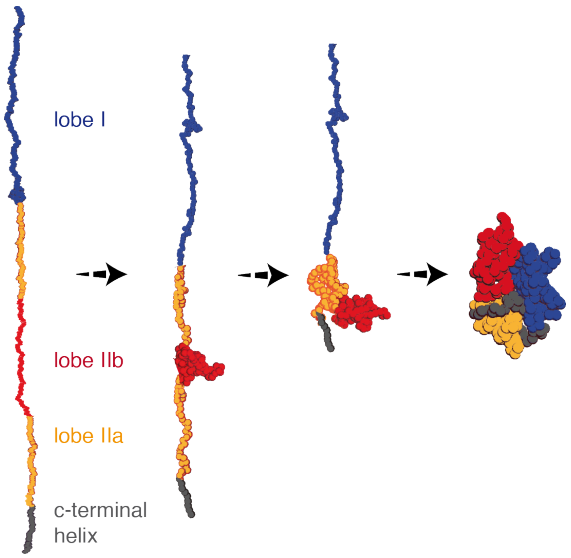


Figure 21: The refolding pathway of EcNBD. The refolding of NBD occurs along a sequential pathway. First, lobe IIb refolds followed by refolding of lobe IIa in the next step. Last, the native fold is achieved by refolding of lobe I and the insertion of the c-terminal helix.

resulting in RFI3 and the corresponding contour-length transformations of these traces (Figure 22B). The obtained contour length for RFI3 is $L_C \approx 22\,\text{nm}$. The contour-length gain of the refolded portion of RFI3 is 48 nm corresponding to 137 amino acid residues. The entire lobe I incorporates 188 residues. Hence, 51 amino acids are still unfolded after formation of RFI3. Which parts remain unfolded is speculative, but one could argue that the c-terminal helix refolds in the last step because it needs to be inserted into lobe I.

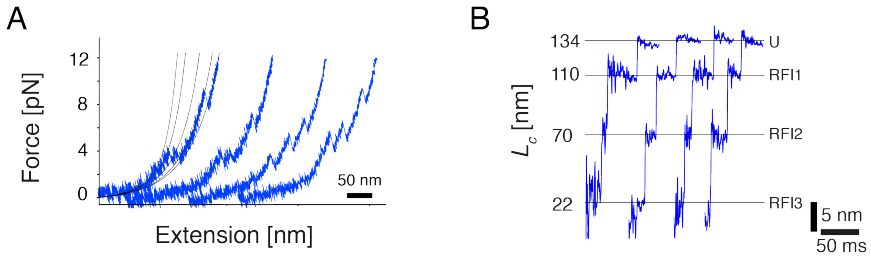


Figure 22: RFI3 is a very rarely occurring refolding event. (A) Constant velocity cylces ($200\,\text{nm}\,\text{s}^{-1}$) show the additional refolding event RFI3. (B) The contour length of RFI3 is 22 nm or 137 refolded residues of lobe I. This results in a remaining unfolded portion of 51 amino acids.

5.2.3 *Kinetic investigation of the refolding process*

To investigate the kinetics of the refolding process we used passive mode experiments. By holding the trap positions constant, the force-dependent probabilities of populating a particular state are recorded. This passive mode experiments can be done for many trap positions. In this way, the dwell-time distribution of each state at a specific force can be fitted by a single-exponential (if it is a one-step process) and the un- or refolding rate can be calculated. These force-dependent rates can be plotted against the force which results in a so-called chevron plot. This plot allows to make conclusions about the kinetics of the un- and refolding process. First, the folding and unfolding branch can be fitted, and the force-independent rates can be obtained. Second, the intersect of the unfolding and folding branch mark the force where both states are in equilibrium (equal probability to populate state 1 and state 2).

State	$\Delta G_{apo} k_B T$
RFI1	-15.2
RFI2	-6.6
RFI2f	-3.6

Table 3: List of un- and refolding rates of states U, RFI1 and RFI2 as well as the folding free energies.

The obtained kinetics of EcNBD-nc refolding process in apo conditions are shown in Figure 23A. First, the RFI1 (red) and the unfolded state (gray) are populated (Figure 23A top). At a force of around 4.4 pN population of the unfolded state, RFI1 and rarely RFI2 (orange) is visible. At lower forces ($\approx$ 4 pN) RFI1 is well populated and refolding to the RFI2 state occurs more frequently (red to orange). A closer look at the lifetime distribution reveals that the RFI2 state can be divided into two kinetically different species with the same contour length (further called RFI2 (orange) and RFI2f (green)). RFI2f is very unstable and unfolds very rapidly compared to RFI2. The force-dependent un- and refolding rates for all states are shown in Figure 23B. These rates can be fitted with the Schlierf-Berkemeier-Rief model (BS)-model to obtain zero force un- or refolding rates (see Equation 9). All obtained rates are listed in Figure 24. RFI1 forms rapidly with a rate of 12 400 s^{-1}. The formation of RFI2 is slower with 124 s^{-1}. The unfolding rate of RFI2f is two orders of magnitude faster compared to the unfolding rate of RFI2, which makes it a more fragile state than RFI2. These results show that RF1 and RFI2 are kinetically separated. RFI1 is fast forming and RFI2 forms on a slower time scale. This difference in kinetics further supports the proposed sequential refolding pathway, where RFI1 then RFI2 and in the end, the native state is populated.

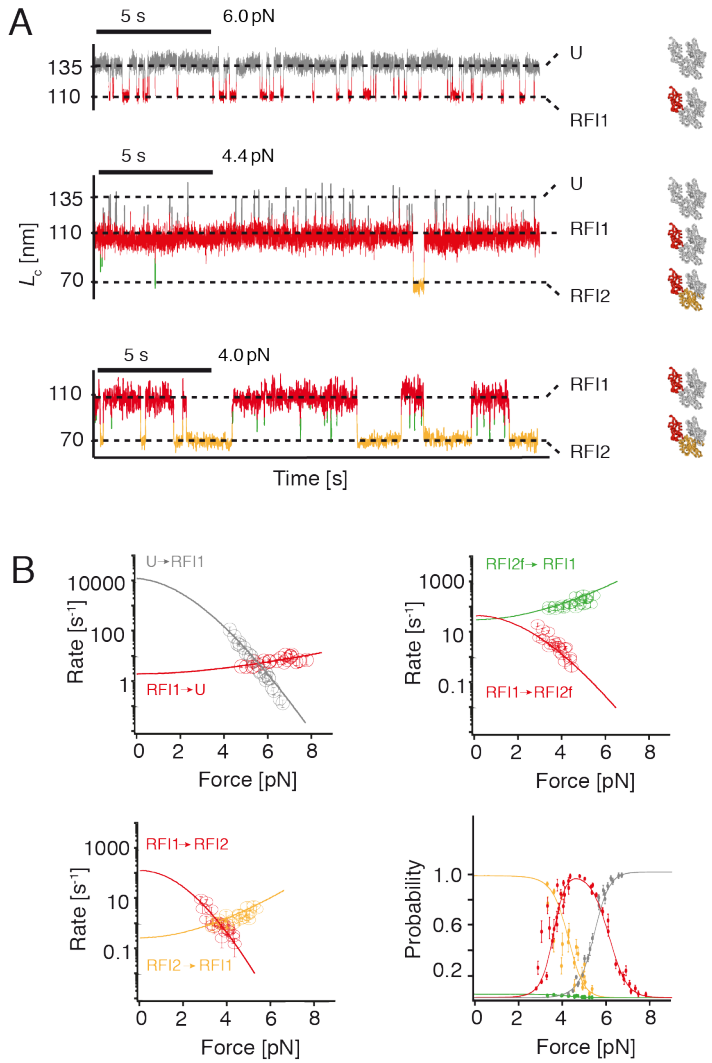


Figure 23: Passive mode experiments on the refolding process of EcNBD-nc in apo. The color code is unfolded state (gray), RFI1 (red), RFI2 (orange) and RFI2f (green). (A) Passive mode experiments at different force levels: 6 pN to 4 pN (The mean force populated by RFI1 defines the force level). In the upper panel, the unfolded state and the RFI1 state can be populated at 6 pN. At lower force (4.4 pN) predominantly RFI1 gets populated and occasionally the unfolded state or RFI2. At 4 pN, RFI1 and RFI2 get populated. For RFI2 we see two kinetically different species: RFI2 (orange) and RFI2f (green). (B) All rates can be plotted against force and can be fitted with the BS-model (solid lines). The probability of populating a specific state over force is shown in the lower right.

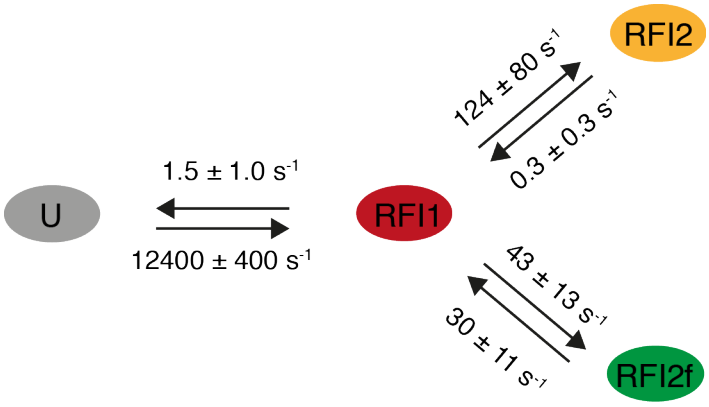


Figure 24: Schematic of transition rates of apo states populated in passive mode experiments.

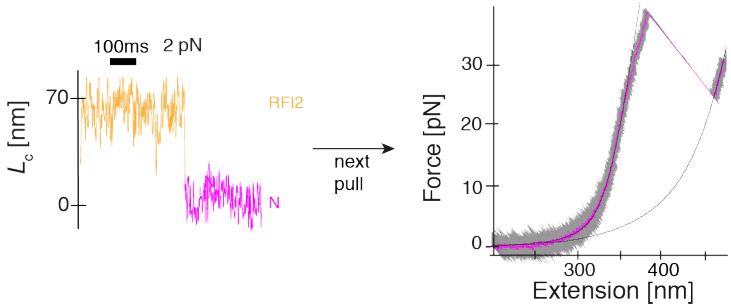


Figure 25: The refolding to the native state from RFI2. Right: passive mode experiment at 2 pN. RFI2 is stably populated (orange) and after some time it can fold to the completely native state (pink). This is confirmed by the constant velocity trace on the left (20 nm s^{-1}).

If the forces in passive mode are low and the waiting time long enough, we sometimes observe refolding to the native state from the RFI2 state ($\approx$ 2 pN) (Figure 25, left). To confirm that this state has to unfold in constant velocity measurements via unfolding pattern 1, which is indeed the case (Figure 25, right). Therefore, productive folding to the native state can only be accomplished when lobe II is stably folded. Furthermore, this data proves that RFI2 is an on-pathway intermediate.

Passive mode experiments provide information about the energetics of the refolding process. One can calculate not only rates but also equilibrium free energies for each state (ΔG)[16] from the probability distribution (Figure 23B, bottom left). Those determined energies are listed in Table 3. As expected, RFI1 has a much higher thermal stability ($\Delta G_{apo} = -15.2\, k_B T$) compared to RFI2 ($\Delta G_{apo} = -6.6\, k_B T$) and RFI2f ($\Delta G_{apo} = -3.6\, k_B T$).

[16] *Gibbs-Helmholtz equation:* $\Delta G = \Delta H - T \cdot \Delta S$ *with* ΔG*: change in Gibbs free energy,* ΔH*: change in enthalpy and* ΔS*: change entropy.*

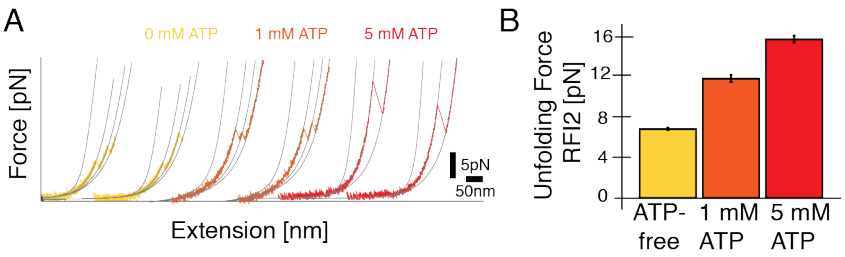


Figure 26: RFI2 is able to bind nucleotide.(A) Constant velocity stretching cycles (500 nm s^{-1}) of EcNBD-nc refolding intermediates in the absence (0 mM ATP, orange) and presence of nucleotide (1 mM ATP, orange; 5 mM ATP, red). (B) Bar diagram of the RFI2 unfolding force depending on the nucleotide concentration.

5.3 THE REFOLDING INTERMEDIATES ARE NATIVELY FOLDED

After we knew what structural parts of EcNBD refold, we were curious about if the refolded portions are natively folded and therefore functional subunits of the EcNBD. The LigPlot (Figure 4B) reveals that the ligand Mg-ATP is mostly coordinated by lobe II. Furthermore, lobe II refolds almost completely with a fast rate and under load. If the RFI2 structure is actually native-like folded, the ligand should be able to bind to the well-positioned residues in the active center.

5.3.1 *Ligand binding to the refolded state*

To test if the refolded portions are natively folded, we performed the first set of constant velocity experiments in the presence of Mg-ATP (0, 1 and 5 mM ATP). As Figure 26 shows is the unfolding force of RFI2 significantly shifted up to higher forces the more ligand is added. This increase is due to the additional stabilizing interactions with the ligand. This result proves that RFI2 is natively folded and the active site is positioned correctly, which allows ligand binding.

Ligand binding does not only influence the RFI2 state it also affects the overall refolding time of EcNBD. The refolding time to the native state can be plotted in the presence of ligand (1 mM Mg-ATP), and fitted with a single-exponential equation. The time constant obtained from this fit shows that the refolding rate is accelerated by a factor of two compared to apo conditions (0.03 s^{-1} (apo) and 0.07 s^{-1} (holo), Figure 27). Hence, the presence of ligand stabilizes lobe II and enhances lobe I refolding, which together leads to an increase of the folding rate of the whole EcNBD.

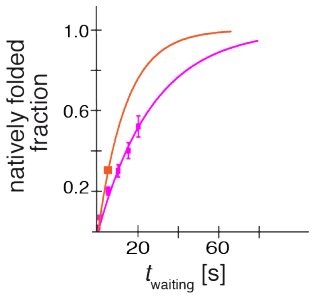


Figure 27: Refolding probability in apo (pink) and holo (orange) to the native state. The obtained refolding rates are $0.03\,s^{-1}$ for apo and $0.07\,s^{-1}$ for holo.

5.3.2 *Kinetics of ligand binding to the refolding intermediates*

In the next set of experiments, the ligand Mg-ATP was added in different concentrations to the passive mode experiments (Figure 28). For the nucleotide-bound states longer dwell times due to ligand binding are expected, which we indeed observed for the RFI2 state. The more ligand is added (top to bottom, Figure 28), the more RFI2 is stabilized, and hence the longer are the dwell times of this state. In the case of RFI2, the refolding branch (RFI1→RFI2) is not affected by ligand. However, the unfolding branch (RFI2→RFI1) is significantly shifted to slower kinetics (Figure 28B). This effect is also visible in the equilibrium free energy for RFI2 (Table 4). In the presence of ligand it is significant increased ($\Delta G_{holo} = -9.3\,k_BT$), as a result of a more stable structure than without ligand.

In contrast, the data clearly shows that RFI1, as well as RFI2f, are not ATP binding competent. For RFI1 and RFI2f, the unfolding, as well as the refolding branch of the chevron, match the ones obtained for the apo form (Figure 28A). Hence, the formed structures of RFI1 and RFI2f are not capable of nucleotide binding, because essential residues or not correctly positioned or missing in the folded portion.

State	$\Delta G_{\mathbf{apo}}$ $[k_BT]$	$\Delta G_{\mathbf{holo}}$ $[k_BT]$
RFI1	−15.2	−14.5
RFI2	−6.6	−9.3
RFI2f	−3.6	−3.8

Table 4: Free energies of the refolding process in the presence and absence of ligand.

The fact that ligand is stabilizing RFI2 can be used to determine the dissociation constant of the ligand. Therefore, the ligand concentration is successively increased and the lifetimes at a specific force are plotted. This scatter plots are shown in Figure 29A for all states. It is

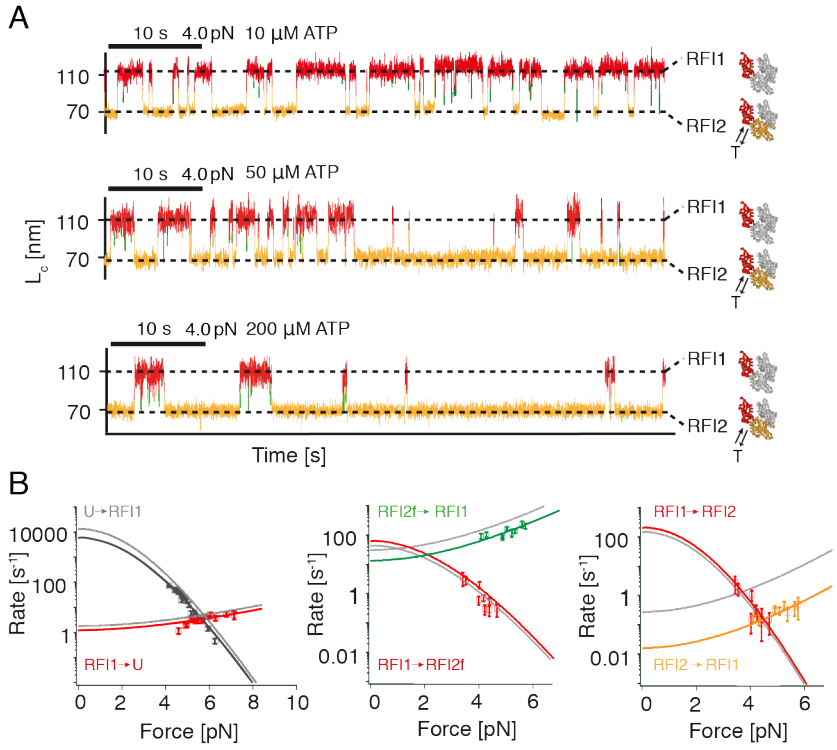


Figure 28: Passive mode experiments of the refolding process of EcNBD-nc. (A) at 4 pN with different Mg-ATP concentrations: 10, 50 and 200 µM (top to bottom). RFI1 is shown in red, RFI2 in orange and RFI2f in green. (B) Kinetics of the refolding process of EcNBD-nc in the presence of 200 µM Mg-ATP (T). The unfolded protein is colored in black, RFI1 in red, RFI2 in orange and RFI2f in green. Overlayed in gray is the unfolding and refolding branches of the apo EcNBD-nc.

clearly visible that more ligand leads to longer lifetimes of the RFI2 state (orange). At very high ATP concentrations (higher than 200 µM), only very few events show lifetimes below 1 s.

The question was now is the exchange with ligand fast or slow. To this end we performed kinetic simulations with the same parameters as used experimentally[1]. These simulations show that only RFI2 shifts to longer lifetimes (Figure 29B), as we found in our experiments. All other states behave independent of ligand and appear like in the apo form. Furthermore, experiment and simulation show that the exchange of ligand with the RFI2 state is fast. A shift of the whole population to longer lifetimes can be observed when the ligand is present. The scenario of a slow exchange would result in two lifetime populations: apo and holo. This is not the case for our system. Hence, ligand exchange in our kinetic network is fast.

1 Simulations were performed in collaboration with Dr. Johannes Stigler.

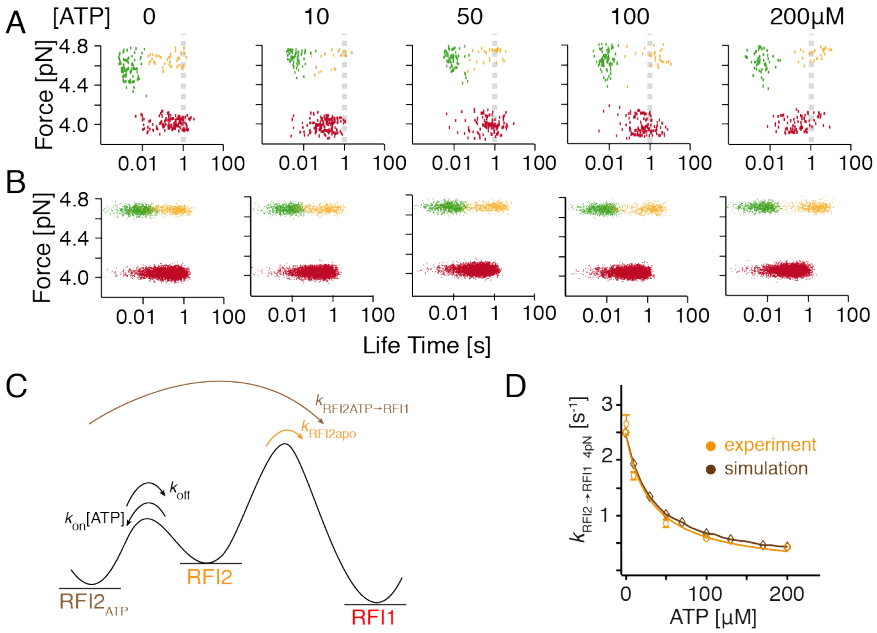


Figure 29: Kinetics of ATP binding to RFI2: experiment and simulation. (A) experimentally obtained lifetimes of RFI1 (red), RFI2f (green) and RFI2 (orange) in the absence and presence of different Mg-ATP concentrations. (B) simulated lifetime distribution of RFI1, RFI2f and RFI2 under the same conditions as in the experimental setup. (C) illustration of the ATP-binding model. (D) unfolding rate of RFI2 plotted against the ATP concentration. Lines are fits of the data (experiment and simulation) with the ATP-binding model to obtain the dissociation rates ($K_D = 33 \pm 10\,\mu M$ and $K_{D,sim} = 30\,\mu M$).

The simulations provide insights into the range of association and dissociation rate-constants. To extract the dissociation constant of this titration data, a simple ATP binding model was used (Figure 29C). In this model, the ligand can only bind to RFI2, and the exchange is faster than the unfolding rate of RFI2. Kinetic simulations were performed, to set the range for the dissociation and association rate of ligand. The lower range of the ligand's on-rate is calculated to be $10^6\,M^{-1}\,s^{-1}$, which is close to the diffusion limit. For details see the Section 13.5.3.5 in the method chapter. Plotting $k_{RFI2 \rightarrow RFI1}$ at 4 pN[17] over the ATP concentration, a binding curve can be obtained (Figure 29D). The experimental and simulated data fitted with the ATP-binding model, derive the dissociation constant of ATP of $K_D = 33 \pm 10\,\mu M$ and $K_{D,sim} = 30\,\mu M$, respectively. These values are 50-times lower than for EcNBD ($K_D = 0.7\,\mu M$) [83, 117], which is due to missing lobe I.

[17] *The given force is the force of the RFI1 state*

Next, we were curious if the presence of Mg-ADP or Mg-adenosine monophosphate (AMP) have a different ligand-binding kinetics compared to Mg-ATP. These ligands have one or two phosphate groups

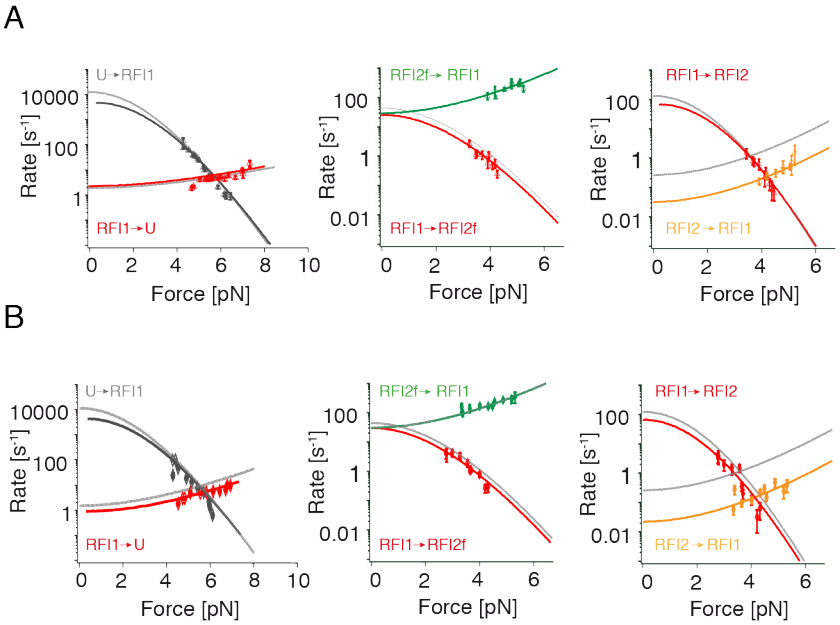


Figure 30: Passive mode experiments of the refolding process of EcNBD-nc. (A) presence of 200 µM Mg-ADP. (B) presence of 200 µM Mg-AMP. In gray the chevron plots of apo EcNBD-nc are overlayed.

removed compared to ATP. Passive mode experiments with Mg-ADP and Mg-AMP should reveal if the sensing of bound ligand is due to energetic discrimination in lobe II. Therefore, we performed these experiments in the presence of 200 µM Mg-ADP or Mg-AMP. The chevron plots in Figure 30A and Figure 30B show that the unfolding and refolding rates of each state match precisely the ones obtained for Mg-ATP. The unfolding branch of RFI2 is shifted to slower kinetics, like in the presence of Mg-ATP. All other transitions have the same kinetics as the apo form. These facts lead to the conclusion that lobe II alone is not able to distinguish which adenosine nucleotide is bound.

5.4 SUMMARY OF THE REFOLDING PATHWAY OF ECNBD

In this chapter, the refolding pathway of EcNBD could be deciphered. We found that the active refolding portion of the protein corresponds to lobe II. Lobe I, in contrast, needs longer time to refold under no load and is the rate-limiting step of the refolding process. The refolded portion can be subdivided into RFI1 and RFI2. Structurally RFI1 corresponds to lobe IIb, and RFI2 harbors the pre-folded lobe IIb and lobe IIa domains (Figure 21). In the last step, lobe I and the proper positioning of the c-terminal helix occur. The addition of ligand to the refolding process does not change the earlier mentioned sequential refolding pattern (U — RFI1— RFI2). The ligand shifts up the unfolding force and lifetime of RFI2. This stabilizing effect indicates ligand binding to RFI2. Hence, RFI2 is natively folded because only if the active site is established correctly, ligand-binding is possible. To describe the ligand-binding process in more detail, we combined experimental data and kinetic simulations. We could characterize this ligand-binding process by determining the dissociation constant experimentally and with simulations ($K_D \approx 33\,\mu M$ and $K_{D,sim} = 30\,\mu M$). The simulation allowed us to further characterize the ligand binding by setting the range of the ligand on-rate to $10^6\,M^{-1}\,s^{-1}$. This results show that the ligand exchange is fast and is limited by the off-rate. The accurately folded lobe II enables further refolding towards the native state. We think that RFI2 is an obligatory on-pathway intermediate and might even act as a folding seed for lobe I. To prove the folding-seed theory, chimeric proteins were designed, which I will discuss in the next chapter Chapter 6.

6

YEAST MITOCHONDRIAL HSP70 NBD: MAKING A FOLDING INCOMPETENT PROTEIN A FOLDING-COMPETENT ONE

The yeast mitochondrial Hsp70, Ssc1, is closely related to DnaK in *E.coli*. The sequence identity of both NBDs is more than 60% (emboss needle tool [95]) and the three-dimensional structures are highly similar. However, as the following sections will show, the refolding abilities of these two NBD analogs are strikingly different.

6.1 UNFOLDING AND REFOLDING ABILITY OF SSC1-NBD

The mitochondrial NBD (mtNBD) protein was, as the EcNBD-nc protein, fused n- and c-terminally to dsDNA handles for optical tweezers experiments. Figure 31A depicts the unfolding pattern of mtNBD in apo form. The mtNBD-nc unfolding pattern shows a high force ($\approx$ 30 pN) unfolding event and two transiently populated, but well-defined unfolding intermediates (Figure 31B). As Figure 31B shows, the unfolding intermediates are similar to the EcNBD-nc (pink) ones. Hence, apo mtNBD-nc and apo EcNBD-nc follow the same unfolding pathway (Chapter 4). Furthermore, unfolding experiments of mtNBD-nc in the presence of ligand (1 mM Mg-ATP) show, like EcNBD-nc, three major unfolding intermediates with equal contour length and lifetime to those measured in EcNBD-nc intermediates (Figure 31C). This argues that the mtNBD domain hierarchy gets, like for EcNBD, shifted upon nucleotide binding (Figure 31D). To sum up, the unfolding pathway of mtNBD-nc shows the same domain hierarchy and ligand dependencies in the unfolding pathway as the homolog EcNBD-nc.

A big disparity between EcNBD and mtNBD lies in their ability to refold. EcNBD refolds readily with a rate of 0.03 s^{-1} to the native state by populating two on-pathway intermediates (Chapter 5). In contrast, mtNBD is not able to refold in the experimentally accessible time ($\approx$ 15 min, Figure 32A). Some partially collapsed states (green) scatter over a broad range of contour lengths, but no well-defined intermediates as for the EcNBD-nc could be detected (Figure 32B). From literature it is known that Ssc1 needs the chaperone Hep1 to refold completely *in vivo* and *in vitro* [12]. To test, if nucleotide or nucleotide and Hep1 are needed for proper refolding, we added those to the assay (Figure 32C and Figure 32D, respectively). As the scatter plots show, neither Mg-ATP nor Mg-ATP and Hep1 improve the refolding ability of mtNBD in our experiments. All tested conditions show that

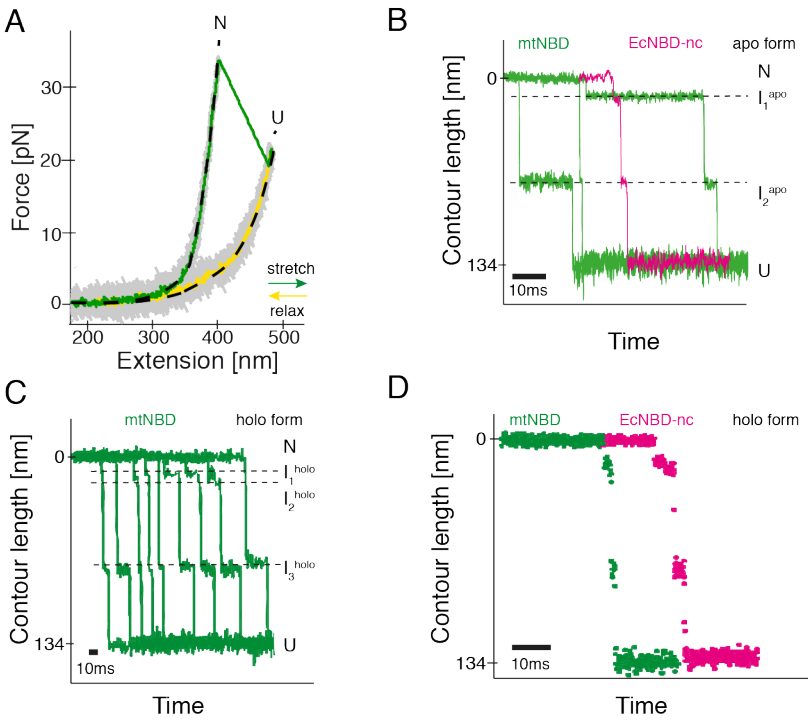


Figure 31: Unfolding pattern of mtNBD in apo and holo form. (A) Example constant velocity unfolding trace of mtNBD with a pulling speed of 20 nm s^{-1}. The stretching cycle is shown in dark green, and the relax cycle in yellow. (B) Contour-length transformation of apo mtNBD (green) unfolding events. Superimposed is an example trace of EcNBD (pink). (C) Contour-length transformations of unfolding events from mtNBD in the presence of 1 mM MgATP. (D) Comparison of the holo unfolding pattern of EcNBD-nc with mtNBD. The unfolding events follow in both proteins the same unfolding pathway.

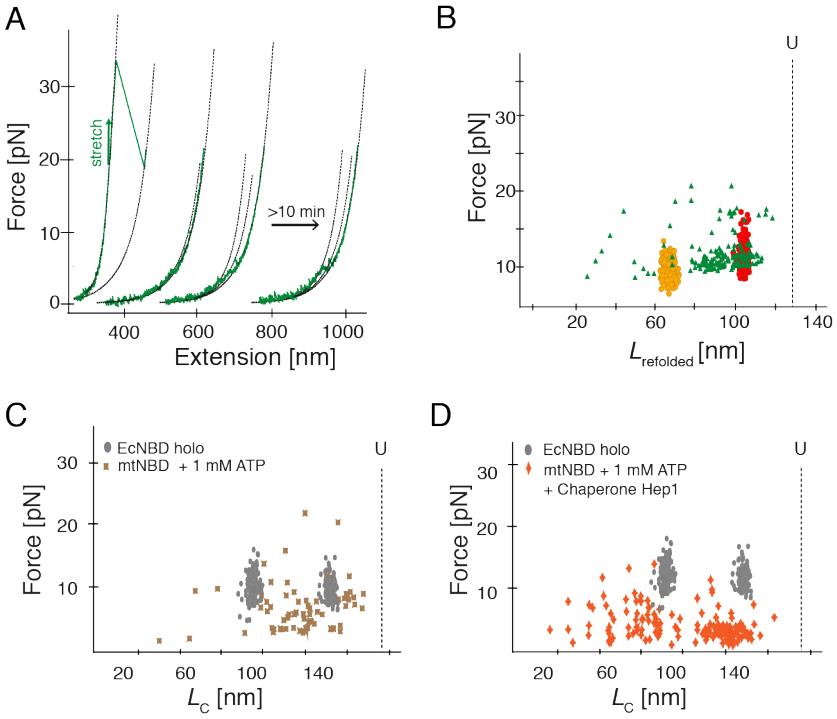


Figure 32: Unfolding pattern of mtNBD in apo and holo form. (A) Subsequent unfolding cycles of mtNBD-nc in constant velocity experiments with a pulling speed of 200 nm s^{-1} (only stretch cycles (green)). (B) Scatter-plot of the unfolding force versus the contour length of each event. In green, the refolding attempts of mtNBD-nc are shown and, for comparison, the well-defined RFI1 (red) and RFI2 (orange) of apo EcNBD-nc are included. The unfolded contour length is marked as a dashed line. (C) Scatter-plot of the contour lengths of mtNBD-nc (brown) and ATP-bound EcNBD (gray) in the presence of 1 mM Mg-ATP. (D) Scatter-plot of the contour lengths in the presence of 1 mM Mg-ATP and chaperone Hep1 (dark orange) and of ATP-bound EcNBD-nc (gray).

mtNBD-nc does not refold to the native state. Again, some refolding attempts of weakly formed collapsed states are visible, but no well-defined refolding intermediates, like for EcNBD-nc (gray). The presence of Hep1 led to folding attempts which result in longer unfolding contour length, compared to the refolded portions formed in apo or in the presence of ATP. All refolding attempts measured for mtNBD are very unstable and unfold at forces below 10 pN. In summary, the refolding of mtNBD-nc is in comparison with EcNBD-nc drastically different. While EcNBD-nc populates two well-defined refolding intermediates, mtNBD-nc refolds only to weakly folded collapsed states. Even in the presence of ligand or Hep1, no native refolding in an experimentally accessible time frame is possible.

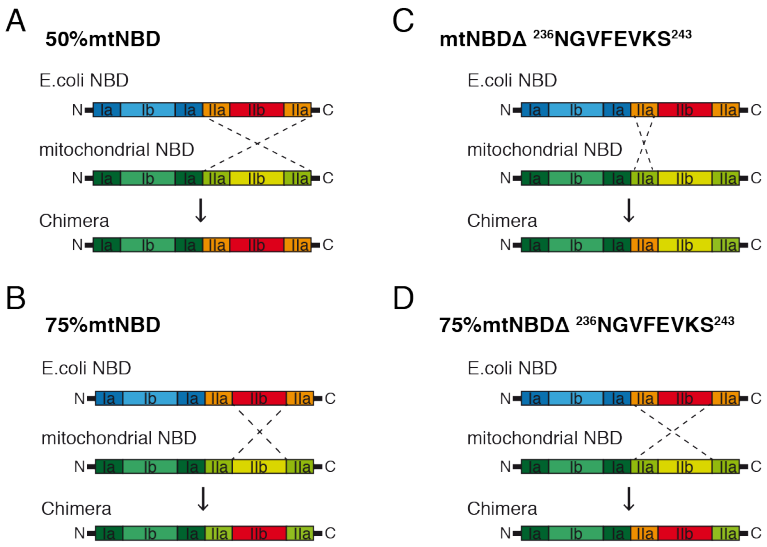


Figure 33: Design of chimeric proteins. (A) 50%mtNBD has lobe II of EcNBD grafted into the mtNBD sequence. (B) 75%mtNBD has lobe IIb of EcNBD grafted into the mtNBD sequence. (C) mtNBDΔ^{236}NGVFEVKS^{243} has a stretch of amino acids exchanged with the corresponding EcNBD sequence. (D) 75%mtNBDΔ^{236}NGVFEVKS^{243} has lobe IIb of EcNBD grafted into the mtNBD sequence and a stretch of amino acids replaced by the EcNBD sequence. For further detail see the amino acid sequences listed in the supplemental part (Section A.3)

6.2 LOBE II IS A CRUCIAL REFOLDING INTERMEDIATE FOR NBD

Why is the refolding of mtNBD so different than EcNBD? For the refolding path of EcNBD-nc, we found that lobe II plays a critical role. Only when this sub-structure is folded correctly, further folding to the native state is possible. In mtNBD, lobe II does not refold well, as the partially collapsed refolding attempts showed (Figure 32). To test the hypothesis of lobe II being a folding seed for the NBD fold, we designed chimeric proteins. The first chimeric protein, 50%mtNBD, is composed of lobe I, the c-terminal helix sequence from mtNBD and the lobe II sequence from EcNBD (Figure 33A).

6.2.1 *The chimeric protein 50%mtNBD is natively folded*

First, the chimeric protein fold was analyzed using CD spectroscopy and compared with the spectra of EcNBD. Figure 34A shows that the chimeric protein is well structured with an overall α-helical structure. The CD spectrum matches the one from EcNBD, and hence the chimera is properly folded. This result shows that the chimeric protein

can be expressed in *E.coli* without any co-expression of chaperones, as is necessary for mtNBD.

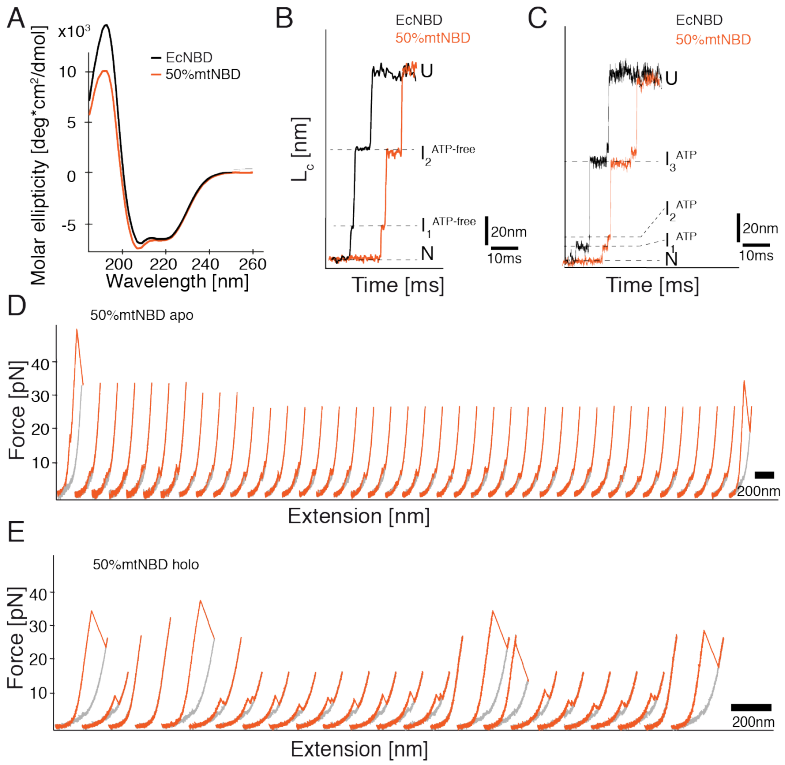


Figure 34: Results of 50%mtNBD CD spectra, unfolding and refolding in optical tweezers. (A) CD spectra of EcNBD-nc and 50%mtNBD recorded at 20 °C. (B) Unfolding pattern of EcNBD-nc and 50%mtNBD in comparison show that in the ATP-free case both proteins populate the same unfolding intermediates. (C) Unfolding pathway of EcNBD-nc and 50%mtNBD in the presence of 1 mM Mg-ATP populate the same three unfolding intermediates. (D) Subsequent stretching cycles (200 nm s^{-1}) of 50%mtNBD in the ATP-free state shows native refolding after 4 min waiting time. (E) Stretching cycles (200 nm s^{-1}) of 50%mtNBD in the presence of 1 mM Mg-ATP show a stabilization of RFI2 (higher unfolding forces) and a faster refolding to the native state ($t_{native,holo} = 2.5$ min).

To test the protein stability and refolding ability, we performed optical tweezers experiments. The native unfolding trace under apo (Figure 34B) and at holo conditions (Figure 34C) show the same behavior as we found for the EcNBD and mtNBD unfolding pathway. This proves that the chimeric protein is appropriately folded and the domain swap had no adverse effect on the protein stability.

Next, we were curious if the refolding ability of the 50%mtNBD has improved. In a first set, we unfolded and refolded the chimeric protein under apo conditions. As figure Figure 34D proves, is the chimeric protein indeed able to refold to the native state after $\approx$ 4 min. As

expected, the refolding intermediates include the well-defined refolding portions corresponding to EcNBD lobe II. In the presence of Mg-ATP, binding of the ligand to the RFI2 state is detected (Figure 34E) and is characterized by higher unfolding forces of RFI2, compared to the apo form. Hence, the RFI2 is natively folded and binds ligand. Also in the ligand-bound state native refolding of the chimera is possible. The refolding rate is even a bit faster than under apo conditions ($t_{\text{native,apo}} = 4\,\text{min}$ and $t_{\text{native,holo}} = 2.5\,\text{min}$). In summary, the chimeric protein (50%mtNBD) shows that grafting lobe II into the folding incompetent mtNBD results in a folding-competent protein. Consequently, refolding of lobe II is a critical step in the refolding pathway towards the native state.

6.2.2 *The whole lobe II is necessary to act as a folding seed for NBD*

To confirm that the whole lobe II region is necessary to act as a folding seed for NBD, we created other chimeric proteins incorporating additional parts of the mtNBD sequence: 75%mtNBD, 75%mtNBD Δ^{236}NGVFEVKS^{243} and mtNBDΔ^{236}NGVFEVKS^{243} (Figure 33).

By grafting only lobe IIb from EcNBD into mtNBD (75%mtNBD), we wanted to test if lobe IIb alone is sufficient to act as a folding seed. The protein could be expressed, purified and the overall structure was determined using CD spectra analysis (Figure 61). Like the EcNBD spectra, the chimeric protein has a pronounced α-helical shape, but some parts are unfolded, due to a slight shift of the minima at 208 nm.In Figure 35A subsequent stretching cycles show that refolding to the native state does not occur in our experimental time. The protein only folds partially. The scatter-plot in Figure 35B indicates that RFI1 is quite well-defined but further refolding attempts have a very broad contour-length and force distribution. Hence, the refolding of lobe IIa in this mtNBD chimera is not possible, even when lobe IIb is refolded and the IIa sequence is in closer proximity.

One study proposed that the replacement of a stretch of 8 amino acids with the *E.coli*sequence can improve protein solubility [13]. To test if this protein variant is more refolding competent we designed 75%mtNBDΔ^{236}NGVFEVKS^{243} and mtNBDΔ^{236}NGVFEVKS^{243}.

The chimeric variant mtNBDΔ^{236}NGVFEVKS^{243} (Figure 33D), could not be purified from *E.coli* lysate due to severe formation of aggregates.

75%mtNBDΔ^{236}NGVFEVKS^{243} was successfully purified and showed as the 75%mtNBD a pronounced overall α-helical CD spectra (Figure 61) which shows also a slight shift at the minima at 208 nm wavelength. In single-molecule folding experiments, this chimera is not natively folded. The scatter plot in Figure 35C shows that RFI2 is more well-defined compared to the 75%mtNBD protein but no further re-

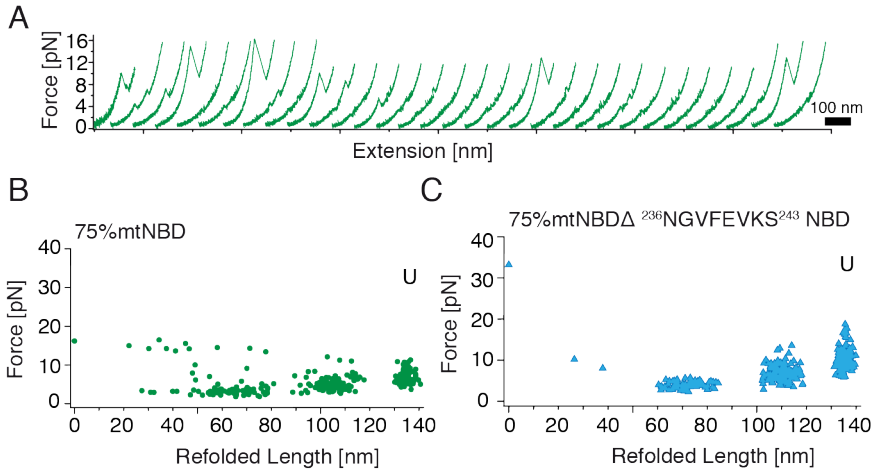


Figure 35: Optical tweezers data on the refolding ability of chimeric proteins. (A) Unfolding sequence of 75%mtNBD in a constant velocity experiment. (B) Scatter-plot of the contour length of the partially refolding events of 75%mtNBD. (C) Scatter-plot of the contour lengths of partially refolded events of the 75%mtNBDΔ^{236}NGVFEVKS^{243} variant.

folding to the native-like contour length is observable. To sum up, only the chimera 50%mtNBD can refold to the native state. All other chimeric proteins incorporating less of lobe II are refolding incompetent. This provides evidence that the whole and stably forming lobe II region is necessary to act as a folding seed for lobe I.

6.3 SUMMARY OF THE MTNBD RESULTS

In summary, the chimera 50%mtNBD is a folding-competent mtNBD chimeric protein variant. All other chimeric proteins, incorporating less of EcNBD lobe II, are folding incompetent and might need the action of Hep1, like the wildtype mtNBD protein [11]. The obtained CD spectra showed that the chimeras, with less EcNBD content, have more unstructured regions compared to 50%mtNBD or EcNBD. This is in agreement with our single-molecule data which shows that partially folded structures exist, but no native folding occurs. The proteins are always partially unfolded. In conclusion, we could show that the refolding ability of the NBD protein is critically dependent on a properly folded lobe II. If the refolding of lobe II is impaired, as in the mtNBD or the 75%mtNBD, 75%mtNBDΔ^{236}NGVFEVKS^{243} and mtNBDΔ^{236}NGVFEVKS^{243} chimeras, folding to the native state is not possible. This provides evidence that the whole lobe II acts as a folding seed for lobe I, by providing the right platform for proper formation of lobe I and insertion of the c-terminal helix into the structure.

7

SUMMARY OF THE UNFOLDING AND REFOLDING PATHWAYS OF NBD

7.1 DOMAIN HIERARCHY IN NBD UNFOLDING AND THE INFLUENCE OF NUCLEOTIDE BINDING

Optical tweezers experiments provide a very detailed folding/unfolding pattern description of proteins at a single molecular level, with a resolution in the nanometer regime (spatial) and microsecond timescale (temporal). The EcNBD protein, which can bind adenosine nucleotide with nano-molar affinity, was investigated. We found, that the presence of nucleotide shifts the domain hierarchy and alters the domain interactions. In the following the main results of the un- and refolding path are summarized.

7.1.1 *The domain hierarchy of EcNBD-nc unfolding*

The unfolding of EcNBD-nc starts in both apo- and holo-form with the unfolding of the c-terminal helix. In apo-form, this step is coupled with the separation of the two lobes. Next, lobe II and, finally, lobe I unfolds. This unfolding sequence is different under holo conditions. In holo-form, after the c-terminal helix is pulled out, the two lobes separate in a disjoint step. Next, lobe I unfolds followed by lobe II unfolding and ligand unbinding. In the holo case, the stabilities of lobe I and lobe II are shifted compared to the apo form. The binding of ligand stabilizes lobe II because more residues are involved in its coordination compared to lobe I (Figure 4B). Hence, lobe II is more force resistant in the holo case compared to lobe I and the domain hierarchy gets shifted. Furthermore, these experiments (unfolding of EcNBD in apo and holo form) show that the lobe-lobe architecture is also conserved in the stability of individual domains—both domains unfold independently of each other and in one event. This might also play an essential role for their function: lobe II as the nucleotide binding domain and lobe I as the sensing domain and its responsibility for the allosteric coupling.

Ligand binding has two significant effects on the stability of the protein structure: (1) Lobe-lobe interaction and (2) domain hierarchy. While in the apo case, the lobe-separation event is coupled to the unfolding of the c-terminal part, is in the holo case, the lobe-separation separated from the c-terminal helix unfolding. Due to the additional ligand interactions stabilizing the interface of the two lobes and hence

the c-terminal helix unfolds before the lobes separate. This stabilization influences also the domain hierarchy of lobe I and II. Lobe II has more interaction sites with the ligand than lobe I, resulting in lobe I unfolding before lobe II unfolds.

In collaboration with Prof. Ruxandra Dima[1], we compared the unfolding pathway found with the optical tweezers experiments with coarse-grained unfolding simulations. The simulation confirmed that the unfolding path in the absence of ligand follows an obligatory pathway including three major steps [9]. First, the unfolding of the c-terminal helix concerted with the separation of the two lobes. Second, lobe II unfolds, followed by the unfolding of lobe I. The lobe separation could previously only be detected in simulations. The loop-insertion variants allowed to corroborate this also with experimental data. The simulations also showed a switch in domain hierarchy upon nucleotide binding [9] as we could show with optical tweezers experimental data. In summary, simulations and experimental data could give detailed insights into the unfolding pathway of EcNBD.

7.1.2 *The c-terminal helix is essential for the high stability of NBD*

Comparing the unfolding pathway in apo and in holo form emphasizes the importance of the c-terminal helix regarding protein stability. The whole NBD structure has a high resistance against force, due to the c-terminal helix, which connects lobe I and lobe II. If this structural element is unfolded, the unfolding of the domains happens very fast (ms-time scale). The importance of the c-terminal helix has also been shown by an EcNBD-nc variant missing this c-terminal helix (EcNBD-$\Delta\alpha$)[18]. This construct showed in optical tweezers experiments only very weakly folded protein, which unfolds around 10 pN. Hence, the proper positioning and formation of the c-terminal helix is key to the high stability of the NBD fold by acting as a molecular glue between the two lobes.

[18] *Shows Figure 59 in this thesis and [80]*

7.2 REFOLDING OF NBD FOLLOWS A SEQUENTIAL PATHWAY POPULATING REFOLDING INTERMEDIATES

7.2.1 *EcNBD populates two on-pathway refolding intermediates*

Single-molecule optical tweezers also provide detailed insights into the refolding process of EcNBD. The whole EcNBD refolds on a much slower time-scale (order of seconds). For refolding two major refolding intermediates could be determined, which are fast and stably forming. First, lobe IIb folds rapidly and supports the refolding of lobe IIa. Those two steps occur within milliseconds. This sequential refolding

1 Chemistry Department, University of Cincinnati, OH 45221

pattern is necessary due to the sequence topology. The sequence of lobe IIb is inserted in-between lobe IIa (Figure 4A). Hence, folding of lobe IIb brings the two parts of lobe IIa in closer proximity and enables folding. In contrast, lobe I can not fold on its own. Slow refolding of lobe I is observable when lobe II has already formed. Therefore, lobe II serves as a folding nucleus for lobe I and the proper positioning of the c-terminal helix. This last step is the rate-limiting step and translates to an overall folding rate of $\approx 0.03\,s^{-1}$.

7.2.2 *Non-natively folded RFI2f is a result of the improper positioning of lobes IIa and IIb*

In the passive mode experiment besides RFI2, a fast-unfolding portion (RFI2f) gets populated (Figure 23). This state has the same contour length as RFI2, but the unfolding rate is 100 times faster. To determine whether this state is an on- or off-pathway intermediate is challenging. Because RFI2f has the same contour length as RFI2, we think, it is conceivable that the same structural parts in RFI2 and RFI2f are refolded. We speculate that the arrangement of lobe IIa with respect to IIb is not correct. This would explain why the RFI2f state is incompetent for ligand binding and the unfolding rate is higher due to missing stabilizing interactions with lobe IIb. Chang *et al.* [22] found in a mutagenesis approach that residues in the hinge region (residue 225 to 230) are involved in the rotation upon ATP hydrolysis. Mutation of Glu-225 or Asp-231 resulted in a much more flexible protein and abolished the concerted action of ATP hydrolysis and refolding ability of DnaK. If this hinge region is not properly positioned due to mutagenesis or application of force, the structure is very flexible. This would result in a high-speed unfolding event as observed for RFI2f.

7.3 FOLDING KINETICS OF ECNBD AND THE CONTRIBUTION OF LIGAND BINDING

7.3.1 *Folding to the native state*

The folding kinetics of EcNBD can be studied using passive mode and constant velocity experiments. Refolding to the native state is limited by the refolding of lobe I and the c-terminal helix. Under apo conditions, EcNBD needs $\approx 0.03\,s^{-1}$ to refold. In contrast, under holo conditions, refolding speeds up to $\approx 0.07\,s^{-1}$. This faster refolding rate is due to the more compact and stable fold in the ligand-bound form. Insertion of loops can result in a significant drop in refolding ability. However, the insertion variants were properly folded as the CD-spectra in Figure 60 prove and first pull analysis in constant velocity data showed native unfolding behavior. We found that especially for

variant Ins45 native refolding under apo conditions could not be detected. Only in the presence of ligand, refolding could be observed on a time-scale around 5 min. We think that the introduced flexibility by the extension of the designated loop region affects the refolding time-scale in some cases more dramatically (Ins45, Ins364) than in others (Ins183, Ins290). Ins45 and Ins364 have the loop introduced in lobe I or before the c-terminal helix. Increasing the flexibility in these structural parts is more dramatic compared to insertions in the lobe II regions. Hence, the introduction of more flexible parts in lobe I has a profound influence on its stability. Also the elongation of the c-terminal helix influences the stability of NBD. As discussed before, is the c-terminal helix acting as a molecular glue between the lobes. Hence, the elongation of this important element results in a less rigid and proper positioning of this helix which leads to a less tight connection of the lobes.

7.3.2 *Natively folded RFI2 enables ligand binding*

Our experiments showed that RFI2 could bind Mg-ATP with a micromolar affinity. Hence, the populated refolding intermediates are natively folded. The active site is functional and binds explicitly ATP. The LigPlot of coordinating residues reveals that most of the amino acids involved in the coordination of the base and ribose of ATP are located in lobe II (red and yellow in Figure 4B). This provides further evidence that lobe II is the ligand binding unit and lobe I is necessary for the detection of the phosphate backbone (ATP or ADP bound). This result also shows that the refolding intermediates are already natively folded and do not rearrange in a last step of the folding process to form an active protein structure.

7.3.3 *Ligand-binding kinetics to RFI2 in optical tweezers assay*

We found that RFI2 (lobe II) can bind ATP with a $K_D = 33\,\mu M$, but a precise kinetic characterization of on- and off-rates is elusive. Kinetic simulations of the ATP-binding process suggested a lower boundary for the off-rate of $3\,s^{-1}$.The off-rate of ATP for EcNBD is reported to be $0.019\,s^{-1}$ [111], which is two orders of magnitude slower compared to lobe II alone and can be attributed to the lack of lobe I. The on-rate is limited by diffusion ($1 \times 10^6\,M^{-1}\,s^{-1}$) as our kinetic simulation data showed.

Next, we were interested how the two lobes contribute energetically. The K_D is reported to be $0.06\,\mu M$ [83, 117] for DnaK and $0.6\,\mu M$ for DnaK-NBD-T199A [113]. Therefore, the ATP binding free energies under standard conditions can be approximated for full-length EcNBD ($-14.2\,k_BT$) and lobe II ($-10.8\,k_BT$). Hence, assuming that the energies

are additive, the binding free energy of lobe I is $-3.8\,k_BT$. The contribution of lobe I to the binding of ligand is only a third, compared to lobe II. Among the 18 residues, involved in nucleotide binding, 13 are located in lobe II. This is 2/3 of the total residues involved in ligand binding and is also consistent with our approximations of the binding free energies. Lobe II can specifically bind Mg-ATP with a lower micro-molar affinity, but the off-rate of ligand is increased. All these effects are due to the lack of lobe I. The presence of lobe I increases the affinity by a factor of two to allow a more regulated ligand interaction.

7.4 LOBE II AS A FOLDING SEED FOR NBD

Refolding experiments conducted with the EcNBD-nc unveiled that lobe II is a fast refolding portion of NBD. Therefore, I speculated that this domain acts as a folding seed for lobe I. I provide evidence for this theory by developing a chimeric protein of the folding-incompetent mtNBD variant by grafting lobe II of EcNBD into the mtNBD sequence. This chimeric protein (50%mtNBD) resulted in a folding-competent NBD protein. Using shorter fragments of the EcNBD lobe II region resulted again in folding deficient chimeric variants. Hence, the whole lobe II region is necessary to act as the predicted folding seed for lobe I.

8

A MINIMAL ATP-BINDING DOMAIN (MINI-NBD) DERIVED FROM OPTICAL FORCE SPECTROSCOPY DATA

In the last chapters, lobe II was introduced as a crucial folding seed for NBD found by single-molecule force experiments. Accurate and fast formation of lobe II is the crucial step along the folding pathway of the NBD fold. In single-molecule data, we could provide evidence that this domain is already able to bind Mg-ATP. We were curious if this lobe II sub-domain alone can be expressed in *E.coli*. We provided evidence that lobe II alone is already a properly folded domain which can bind specifically Mg-ATP in the active site. Last, the crystal structure of this domain will provide molecular details of the structural arrangement of lobe II as well as of the ligand binding site.

8.1 BIOCHEMICAL CHARACTERIZATION OF THE MINI-NBD

Mini-NBD was expressed as a fusion construct with the solubility tag SUMO-His_6. The specific protease Senp2 could remove this tag. The purification was realized in four steps: (1) Ni-NTA column, (2) Size-exclusion chromatography, (3) SUMO-His_6 tag cleavage by Senp2 and (4) the SUMO-tag was removed by hydrophobic column. Finally, pure Mini-NBD was obtained. We designed two Mini-NBD proteins: one with the c-terminal helix present (Mini-NBD-183-383) and one without (Mini-NBD-183-359).

8.1.1 *Mini-NBD is an overall α-helical well-folded domain*

After purification of the two Mini-NBD proteins (Mini-NBD-183-359 and Mini-NBD-183-383), the folding behavior of the domains was tested. First, Circular Dichroism (CD) spectra were recorded at 20 °C, to determine the overall secondary structural elements. As Figure 36 shows, both proteins are well-structured and have, like the EcNBD-nc protein, an overall α-helical fingerprint (ellipticity minima at 208 and 220 nm).

To further investigate if the proteins are native-like folded and can bind nucleotide, we performed limited proteolysis assays in the presence and absence of ligand. For EcNBD we and others could show that the protein is protease resistant in absence and presence of nucleotide (Figure 37A) [64, 103]. This stability can be attributed to the fact that the NBD structure is quite compact and the c-terminal helix stabilizes the fold. In the case of the Mini-NBDs in the apo form, the

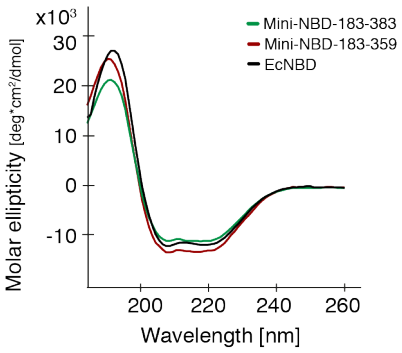


Figure 36: Mini-NBD-183-383 (green) and Mini-NBD-183-359 (red) CD spectra. In black the EcNBD spectra is shown.

protein is entirely digested after 5 min and fragmentation bands arise. In the presence of 1 mM Mg-ATP, the protein fold is stabilized, and the protease cannot digest the protein. This provides evidence that the two Mini-NBDs can bind ATP in the active site. This binding further shows that the active site must be well established to provide a specific ATP-binding site.

Next, the stability and the folding behavior of the Mini-NBDs were investigated using single-molecule force experiments. At this point, I want to briefly recall that we know from the refolding of EcNBD-nc, that RFI1 incorporates lobe IIa and RFI2 refolded lobe IIb an prefolded lobe IIa. These two intermediates should also be detectable in the un- and refolding force-extension traces of the Mini-NBDs. Figure 38 shows the results for both proteins in the absence and presence of nucleotide. The contour-length increase for the RFI1 intermediate is 28.0 ± 2.1 nm, RFI2 32.0 ± 2.0 nm and the fully unfolded contour length is 61.0 ± 1.5 nm, as expected. These results are in good agreement with the corresponding contour-length increases (RFI1 and RFI2) that we determined for EcNBD-nc. In Mini-NBD-183-359 (red, Figure 38A), the first unfolding event, which corresponds to RFI2 in EcNBD-nc, is destabilized. This destabilization results in lower unfolding forces. Sometimes this domain is even not able to refold at all (Figure 38B, light red). The addition of Mg-ATP can rescue the weak fold of lobe IIa by stabilization. This stabilizing effect of ligand, results in higher unfolding forces of lobe IIa ($F_{U,apo} = 3.1 \pm 0.5$ pN to $F_{U,holo} = 8.0 \pm 2.5$ pN).

In the case of Mini-NBD-183-383 (green, Figure 38C), clearly two unfolding events can be seen, corresponding to the RFI1 and RFI2 in EcNBD-nc. These are characterized by a contour-length increase of 27.0 ± 2.4 nm and 35.0 ± 2.3 nm, respectively (Figure 38D). The 20 amino acids ($L_C = 7$ nm) of the c-terminal region are mostly unfolded and are stretched with the DNA. The presence of ligand stabilizes RFI2 and results in higher unfolding forces ($F_{U,apo} = 4.6 \pm 0.9$ pN and

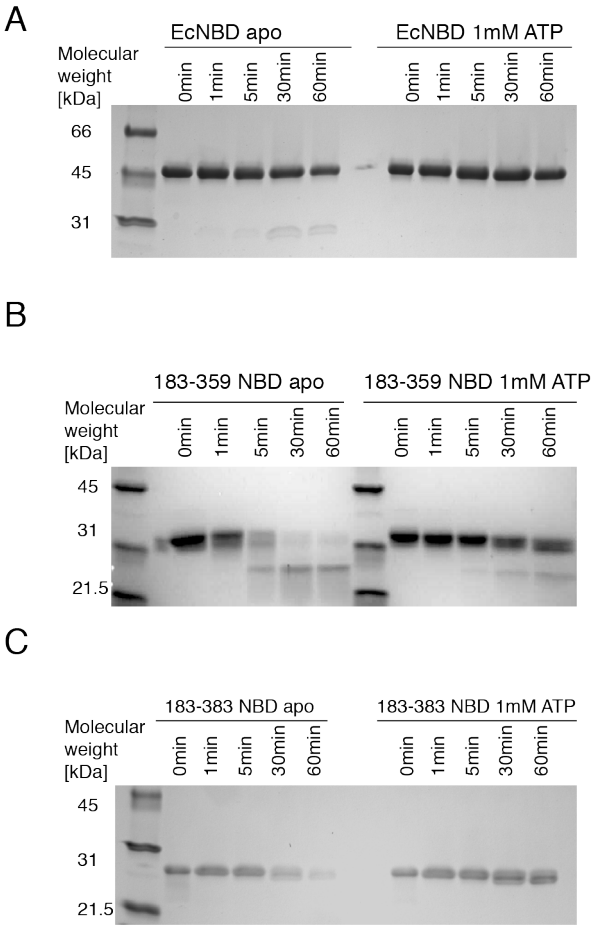


Figure 37: Proteolysis assay on the protein stability of the two Mini-NBDs. In this assay, the protein was incubated with the protease trypsin for a specific amount of time. The fragments were analyzed by SDS-PAGE. (A) Proteolysis assay of EcNBD in the absence (left) and presence of Mg-ATP (right). Each lane corresponds to a specific incubation time with the protease (0–60 min). Same experiments with (B) Mini-NBD-183-359 and (C) Mini-NBD-183-383.

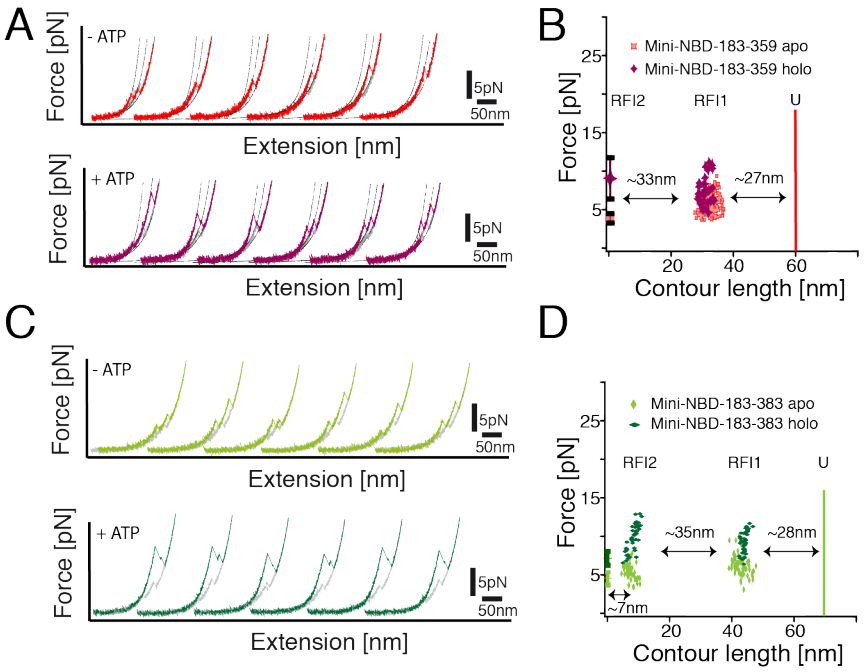


Figure 38: Mechanical stability of the two Mini-NBDs. (A) Unfolding curves of Mini-NBD-183-359 in the absence of nucleotide (upper panel) and presence of 1 mM Mg-ATP (lower panel). Under apo conditions (red) the RFI2 is very weakly folded. Due to nucleotide binding (purple), RFI2 is stabilized and more frequent. (B) Scatter-plot of the contour length of unfolding events (apo, red and holo, purple). The RFI1 intermediate is at 27.0 ± 2.4 nm, RFI2 at 33.0 ± 2.0 nm and the fully unfolded contour length is at 61.0 ± 1.5 nm. (C) Unfolding curves of Mini-NBD-183-383 in the absence (upper panel) and presence of 1 mM Mg-ATP (lower panel). (D) Scatter-plot of all unfolded contour lengths the first 7 nm correspond to the unfolded c-terminal helix. RFI1 has a contour length of 28.0 ± 2.1 nm and RFI2 35.0 ± 2.3 nm. RFI2 is populated well but the C-terminal helix is mostly unfolded (7 nm).

$F_{U,holo} = 6.9 \pm 1.0$ pN). In the presence of ligand, the 20 c-terminal amino-acid residues are, like in apo, mostly unfolded. Sometimes the c-terminal helix is only weakly bound to the RFI2 structure which results in a longer unfolding contour-length (7 nm of the c-terminal helix plus 36 nm of lobe IIa). Ligand seems to trigger the formation of the c-terminal region, but for its proper positioning lobe I is needed.

We conclude from these experiments that Mini-NBD-183-359 is folded, but the truncation at the C-terminus impairs the stability of lobe IIa. If the additional c-terminal amino acids (Mini-NBD-183-383) are included, lobe II folds correctly but the c-terminal helix is loosely bound or unstructured.

8.1.2 *Mini-NBD harbors a natively folded and adenosine specific nucleotide binding pocket*

Mini-NBD is properly folded as shown in Section 8.1. So far the question if the specificity for adenosine nucleotide is still encoded, is not answered. In the next chapters, this question will be addressed by limited proteolysis assays and ligand binding kinetic analysis. These experiments will provide further details of the ligand binding capability of the Mini-NBDs.

8.1.2.1 *ATP binding ability test: limited proteolysis assay*

In this section, I would like to address the question: can the Mini-NBDs specifically bind adenosine ligands? We performed limited proteolysis assays in the presence of different ligands, including Mg-ATP, Mg-GTP, Mg-UTP, and Mg-CTP. The digestion assay was analyzed by SDS-PAGE (Figure 39). We expected, if specifically adenosine ligands are bound, the Mini-NBDs will be digested in the presence of other non-binding ligands.

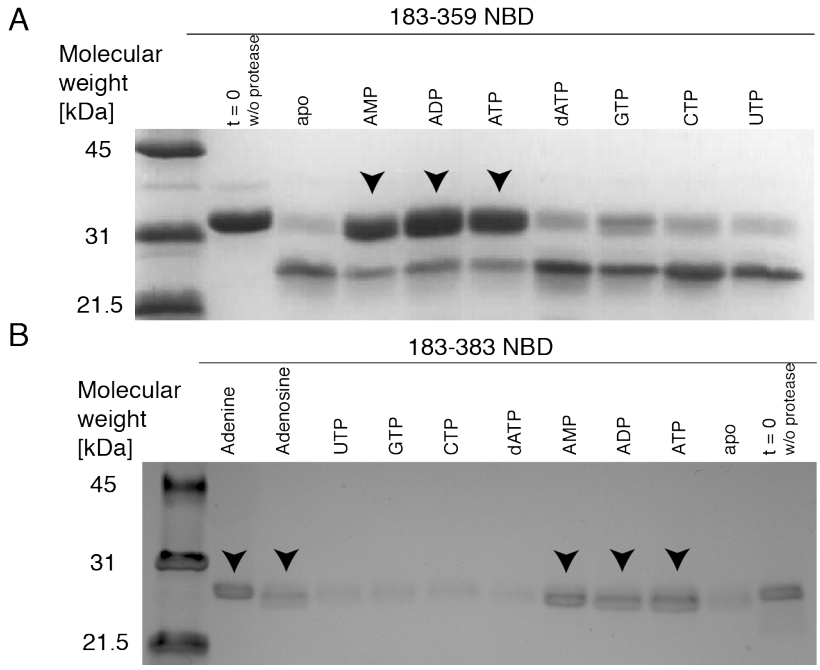


Figure 39: Nucleotide specificity of the two Mini-NBDs using limited proteolysis assay. (A) In the first lane the protein was not incubated with protease. In the following lanes the protein was incubated with different ligands before incubation with trypsin for 30 min. The digestion products were analyzed with the presented SDS-PAGE. Binding of AMP, ADP or ATP protects Mini-NBD-183-359 from digestion (black arrows). (B) The same experiments were done with Mini-NBD-183-383.

In this case, the protease trypsin will produce fragments of the protein. In the case of apo, Mg-dATP, Mg-GTP, Mg-CTP and Mg-UTP the protein is, indeed, digested to lower-molecular-weight fragments. Hence, the Mini-NBDs do not bind these ligands resulting in a less stable structural arrangement which is accessible to protease. In the case of Mg-AMP, Mg-ADP, and Mg-ATP, the protein is stabilized, due to ligand binding to the active site. These results clearly show, that Mini-NBD-183-359 and Mini-NBD-183-383 specifically bind adenosine nucleotides. Furthermore, even the specificity for ribose is harbored in the active site of the Mini-NBDs, because Mg-dATP is not able to associate.

8.1.2.2 *Characterization of the Mini-NBD ATP-binding kinetics and binding affinity*

The results above showed that a properly folded and nucleotide specific minimal-ATP binding site could be designed. Next, we set out to investigate the ligand binding process in more detail and get information about the importance of lobe I for the ligand-binding process. First, we did experiments with different amounts of Atto488-ATP to determine the dissociation constant (K_D) of Mini-NBDs and EcNBD (Figure 40A-Figure 40C).

EcNBD-nc has an K_D of $0.99 \pm 0.34\,\mu M$ for Atto488-ATP. Mini-NBD-183-359 has a K_D of $15.4 \pm 6.3\,\mu M$ and Mini-NBD-183-383 of $10.6 \pm 3.9\,\mu M$. The Mini-NBDs have one order of magnitude lower affinity for ATP compared to the EcNBD control protein (literature value for DnaK is $0.09\,\mu M$ [117] and for DnaK-NBD-T199A $0.6\,\mu M$ [113]). The dissociation constants, we could derive for lobe II in optical tweezers experiments, are in good agreement with the dissociation constant determined for the Mini-NBDs (Section 5.3).

To reveal the off-rate of ligand we used that fact that when Atto488-ATP is bound, it has a higher fluorescence signal. After ligand is bound, Atto488-ATP is chased with a high concentration of unlabeled ATP. The fluorescence signal drops in a single exponential process, due to unbinding of Atto488-ATP and replacement with a non-labeled ATP molecule. To determine the off-rate, the decay in fluorescence intensity over time can be fitted with a single-exponential. In Figure 40D-Figure 40F the results for EcNBD, Mini-NBD-183-359 and Mini-NBD-183-383 are shown. The k_{off} rates of the Mini-domains are $\approx 0.13 \pm 0.04\,s^{-1}$. The experiments show that the unbinding rate is so fast that it is close to the detection limit of our system. Hence, we underestimate the real unbinding rate. Atto488-ATP dissociates with a $k_{off} = 0.011\,s^{-1}$ from EcNBD, which is in good agreement with the literature value for MABA-ATP (unbinding rate of $0.019\,s^{-1}$) [112][19]. The Mini-NBDs show a very fast dissociation of ATP, which provides

[19] *EcNBD was used as a control to show that the Atto488-ATP ligand has the same properties as MABA-ATP (dissociation constant and dissociation rate).*

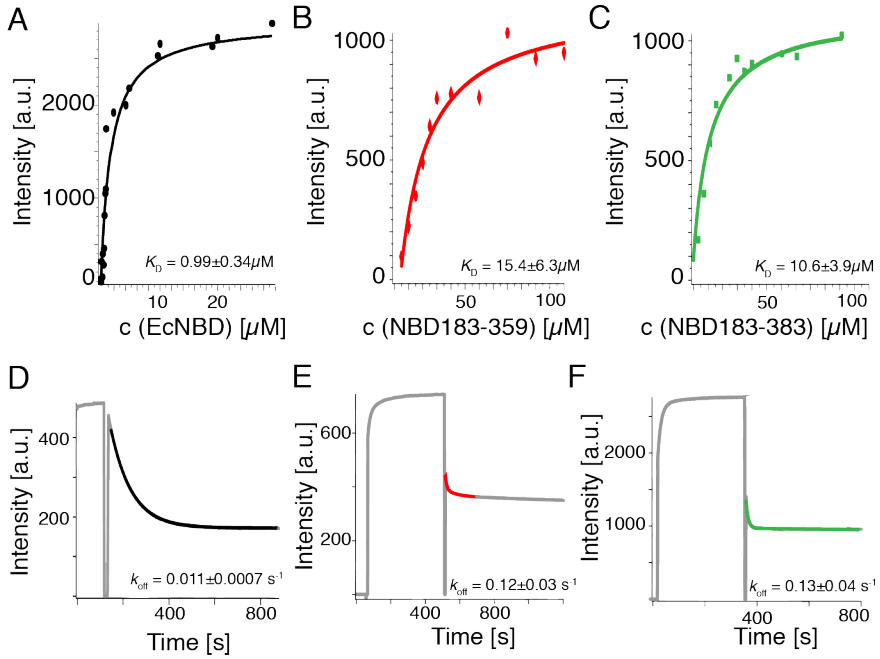


Figure 40: Dissociation-rates and off-rates of Atto488-labled ATP determined for the Mini-NBDs and EcNBD. (A) K_D determination of EcNBD reveals a value of $0.99 \pm 0.34\,\mu M$, of Mini-NBD-183-359 $15.4 \pm 6.3\,\mu M$ (B) and of Mini-NBD-183-383 $10.6 \pm 3.9\,\mu M$ (C). Determination of ATP off-rates using Atto488 labled ATP reveals rates of $0.0110 \pm 0.0007\,s^{-1}$ for EcNBD (D), $0.12 \pm 0.03\,s^{-1}$ for Mini-NBD-183-383 (E) and $0.13 \pm 0.04\,s^{-1}$ for Mini-NBD-183-359 (F).

evidence that missing lobe I leads to a faster exchange of ligand from the active site.

Finally, we were curious as to whether this protein can hydrolyze ATP to ADP. Therefore, enzyme-coupled ATPase assays (Section 13.3.3) were performed. As expected, the Mini-domains are not able to hydrolyze ATP due to the absence of lobe I, which coordinates the phosphates of ATP.

To sum up, the Mini-NBDs can specifically bind adenosine ligands in the active site but with a much lower affinity compared to the EcNBD protein. Moreover, due to missing lobe I the ability to hydrolyze ATP to ADP is not existing.

8.2 DETAILED STRUCTURAL CHARACTERIZATION USING SMALL ANGLE X-RAY SCATTERING AND X-RAY SCATTERING

In the last part, I will provide high-resolution structural data obtained for the two Mini-NBDs in absence and presence of different ligands. I wanted to address the questions: (I) is the structure altered due to the truncation? (II) Does nucleotide binding influence the structure and (III) is the nucleotide bound in the same manner as in EcNBD?

8.2.1 *Ligand induces a more compact fold of EcNBD and Mini domains*

In the limited proteolysis assays, we found that the proteins are protease resistant upon ligand binding. We were now interested why these are protected and we thought it might be due to a more compact fold of the Mini-NBDs. To test this hypothesis, we did SAXS experiments, which are capable of addressing this question. In collaboration with Dr. Tobias Madl,[1] we measured the EcNBD and the Mini-NBD proteins in apo, AMP, ADP and ATP bound form. The results obtained are shown in Figure 41 and Table 12. The density distributions ($P(r)$) show that the proteins have an overall cylindrical shape. The integral of the $P(r)$ distribution with r^2 corresponds to the radius of gyration of the protein (black line in Figure 41A,C,E). The radius of gyration and the calculated molecular mass are listed in Table 12. For EcNBD and Mini-NBD-183-383 we could obtain values reflecting a monomer. However, Mini-NBD-183-359 shows twice the molecular weight which argues for dimers in solution. When now comparing the apo form with the ligand-bound form the following gets obvious: Upon nucleotide binding, the maximum of the $P(r)$ distribution is slightly shifted to a lower radius of gyration and narrower distribution (gray box Figure 41). This provides evidence that the ligand-bound form is more compact compared to the apo form. Only a very minor difference between ATP, ADP or AMP is visible.

To get hands on the molecular structure of the Mini-NBDs, it is possible to do rigid body modeling, based on the SAXS data and crystal-structural information found in: PDB:4B9Q and PDB:2KHO[20]. This modeling relies on finding one matching structural solution for the measured SAXS data. Therefore, it is only a rough estimate of the structure in solution. These protein models show that the ligand-bound proteins are more compact, compared to the apo ones (Figure 41, right side), as proposed from the raw data. These results endorse our hypothesis that the ligand bound state is more compact and hence less proteolytically accessible.

[20] *PDB: Protein Data Base* `https://www.rcsb.org`

1 Special thanks to Gesa Richter and Dr. Tobias Madl, chemistry department TUM, who provided the SAXS data and analysis.

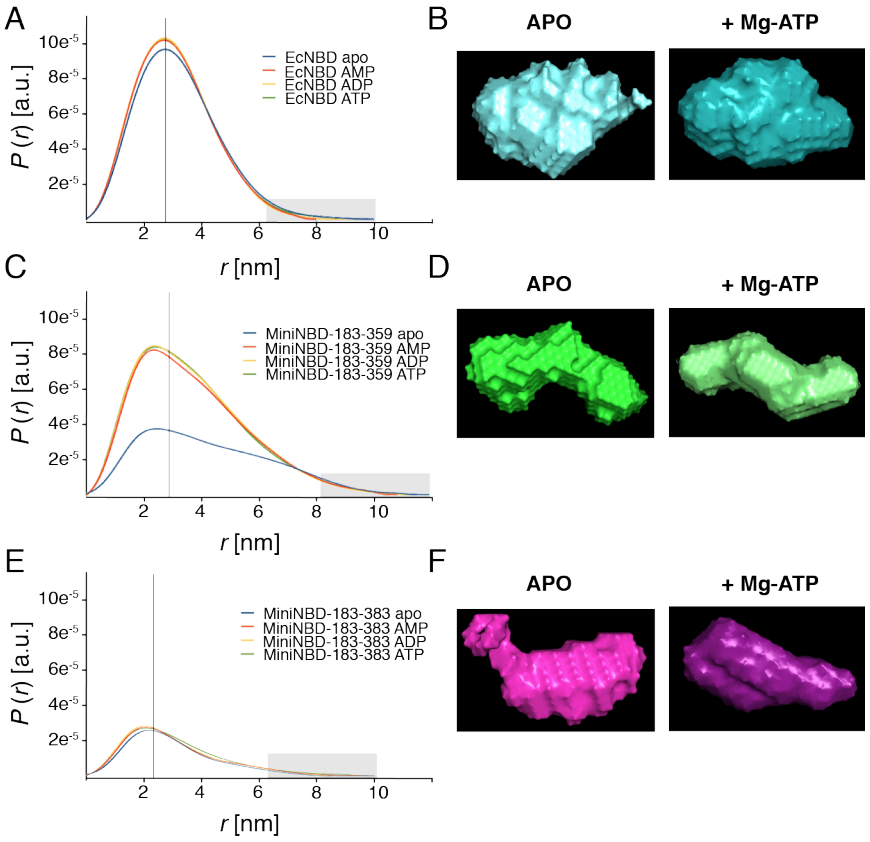


Figure 41: SAXS data and models of Mini-NBD-183-359 and Mini-NBD-183-383 apo, ATP, ADP or AMP bound forms. (A) (C) (E) show the density distribution of EcNBD, Mini-NBD-183-359, Mini-NBD-183-383 in apo, ATP, ADP and AMP bound form. The rigid body models are shown on the left in apo and Mg-ATP bound form (B), (D), (F).

To sum up, with the SAXS experiments the shape of apo EcNBD was determined to be globular and Mini-NBDs are shifted towards a slightly more elongated shape. All proteins undergo tiny changes upon ligand binding. In the ligand-bound form, the structures are more compact compared to the apo form. The density distributions do not show a significant difference between ATP, ADP or AMP. Hence, the overall shape does not reflect the hydrolysis state of the bound ligand. There is only a difference between ligand-bound (more compact) and unbound (less compact) state.

8.2.2 *Crystal structures of Mini-NBD-183-359 and Mini-NBD-183-383*

To obtain detailed high-resolution structural information on the Mini-NBD proteins we resolved in collaboration with Dr. Roman Peter Jakob [2] the crystal structures of both proteins with Mg-ADP, Mg-AMP or Mg-AMPPcP bound in the active center.

8.2.2.1 *The cyrstal structures of the Mini-NBDs*

Crystal structures for Mini-NBD-183-359 and Mini-NBD-183-383 in the ligand-bound form are very similar The obtained crystals for Mini-NBD-183-359 and Mini-NBD-183-383 are shown in Figure 42A and Figure 42E, respectively. Surprisingly, Mini-NBD-183-383 crystallizes in a different space group compared to Mini-NBD-183-359[21]. We were able to determine the crystal structures of Mini-NBD-183-359 in the ADP- (2.3 Å), AMP- (2.7 Å) and AMPPcP-bound (2.0 Å) form (Figure 42B-Figure 42D) as well as Mini-NBD-183-383 in the AMPPcP-bound (2.9 Å) form (Figure 42F). The crystal structure of both proteins in the AMPPcP-bound form, look very similar and overlap with an C_α root mean square deviation (RMSD) of 0.56 Å (Figure 42G). Furthermore, the structure of AMP and ADP bound Mini-NBD-183-359 are superimposing with an RMSD of 0.3 Å showing that the number of phosphates of the ligand does not influence the overall structure (Figure 42H) of lobe II.

Next, we were interested, if the different nucleotides are bound in the same conformation. For that the ATP bound (yellow) DnaK structure (PDB:4B9Q) is overlayed with the Mini-NBD-183-359 ADP (cyan) and AMP (blue) structures (Figure 42J). In the AMP bound structure, the first phosphate group is compared to ADP and ATP rotated. Maybe the missing lobe I interactions make it more difficult to position the phosphate groups in space.

[21] *Crystal systems can be classified as so-called space groups. These are derived by the combination of all available symmetry operations with the Bravais translation. This leads to 230 combinations or space groups.*

2 Biozentrum, Prof. Dr. Timm Maier, University of Basel, Switzerland

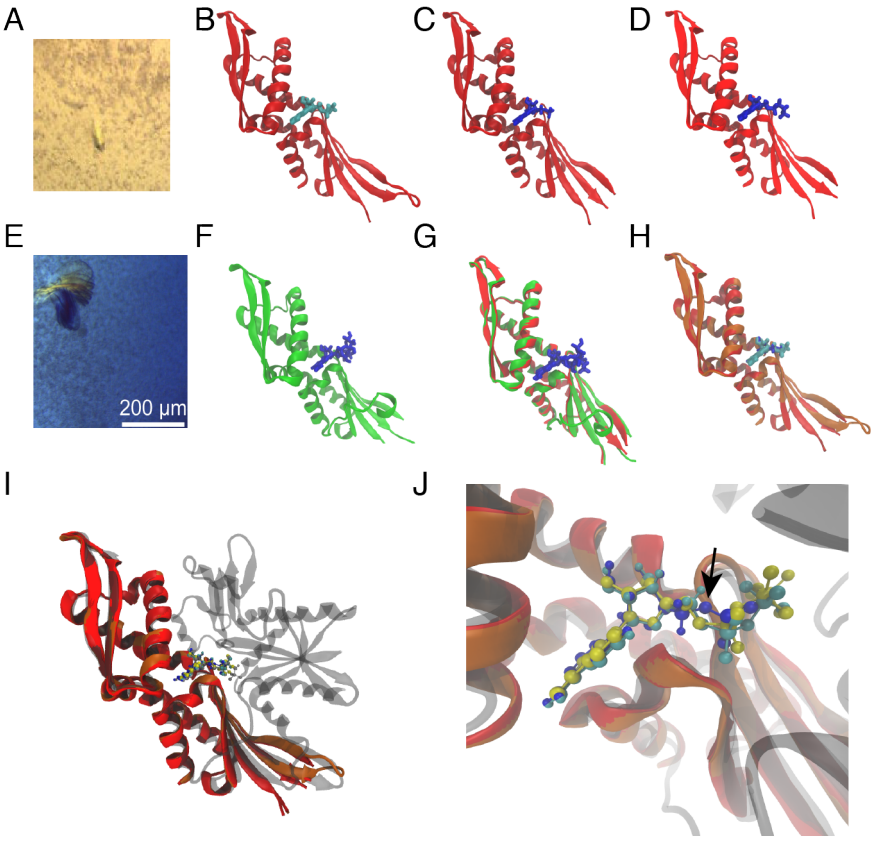


Figure 42: Crystal structures of Mini-NBD-183-359 and Mini-NBD-183-383 ADP, AMP or AMPPcP bound forms. (A) Crystal of Mini-NBD-183-359. (B) Crystal structure of Mini-NBD-183-359 ADP bound (2.3 Å), (C) AMP bound (2.7 Å) and (D) AMPPcP bound (2.0 Å) state. (E) Crystal of Mini-NBD-183-383 is crystallizes in a different space group and shows platelets. (F) Mini-NBD-183-383 structure in the AMPPcP bound state. (G) Superimposition of Mini-NBD-183-359 and Mini-NBD-183-383 in AMPPcP bound state shows an RMSD of 0.56 Å. (H) Superimposition of Mini-NBD-183-359 AMP and ADP bound form shows an RMSD of 0.28 Å. (I) Overlay of PDB:4B9Q, Mini-NBD-183-359 ADP bound and Mini-NBD-183-359 AMP bound structure. (J) AMP (blue) phosphate group is rotated compared to the bound ADP (cyan) or ATP (yellow).

8.2.2.2 *Analysis of Mini-NBD-183-383 cyrstal structure reveals a well positioned active center*

For further analysis, the crystal structure of Mini-NBD-183-383 AMPPcP-bound was used (resolution of 2.9 Å; with R_{work}/R_{free} values of 24.4/26.0%) (Figure 43A). This domain also showed better stability in optical tweezers data and shows no dimers in solution as the Mini-NBD-183-359 protein in SAXS data. Therefore, further analysis was performed using the Mini-NBD-183-383 structure.

The LigPlot in Figure 43C shows the residues involved in the coordination of the ligand. First, an overlay of the structure with the published PDB:4B9Q structure was done (Figure 43B) and the calculated RMSD plotted for each residue (Figure 43D). The structures overlay very well with an RMSD of 0.7 Å. A closer look at single substructures, reveals four regions which show a higher deviation in RMSD. These regions, marked with 1, 2, 3 and 4 in Figure 43B, adopt a different conformation in Mini-NBD-183-383 compared to EcNBD. Regions 1 and 4 are the termini of the proteins and show higher deviation due to the missing lobe I. The insertion and proper folding of the c-terminal helix is impossible when lobe I is missing (region 4). Region 2 and 3 are linker regions between secondary structural elements and are tilted compared to the PDB:4B9Q structure. This tilt might be due to the missing interaction with lobe I. Taken together, the main differences between the Mini-NBD-183-383 structure and the PDB:4B9Q is due to the missing interaction of lobe II with lobe I. Hence, we found two major regions (2 and 3), where lobe I has a profound influence on certain structural elements of lobe II.

Next, we wanted to investigate the active site of Mini-NBD-183-383. All residues involved in the nucleotide coordination are marked with a red arrow in Figure 43D. It is striking that all coordinating residues are very well positioned and the RMSD of each is below the mean RMSD. In summary, the determined structure of Mini-NBD-183-383 is very similar to the published structure of DnaK (PDB:4B9Q). The most significant altered regions are at the termini and two loop structures. Lobe I, which is not part of the Mini-NBDs, may interact with those flexible loop structures in the full-length protein and influence their position.

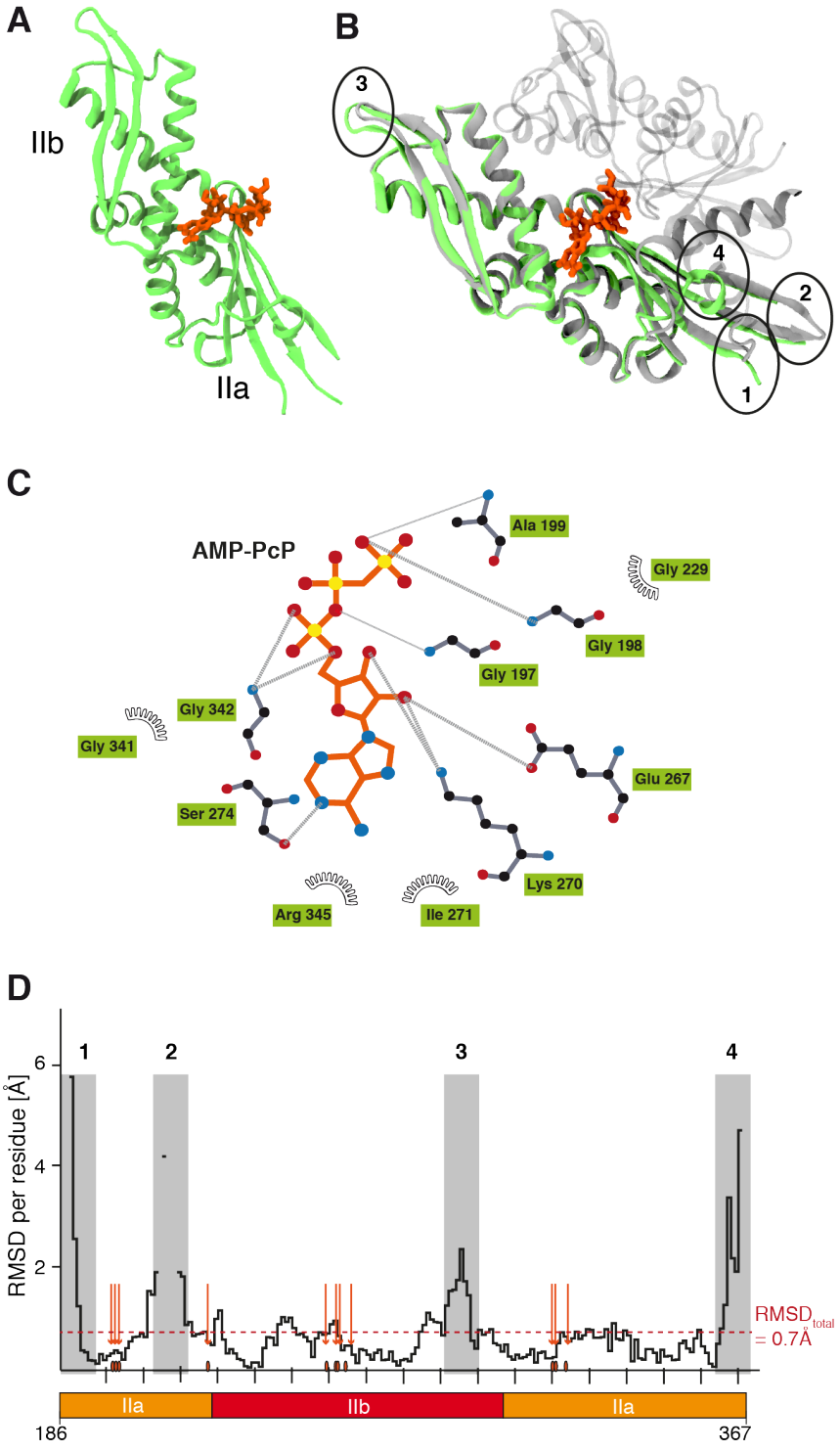


Figure 43: Crystal structures of Mini-NBD-183-383 AMPPcP bound form in comparison with EcNBD structure. (A) Crystal structure of Mini-NBD-183-383 AMPPcP bound solved with a resolution of 2.9 Å. (B) Overlay of PDB:4B9Q with Mini-NBD-183-383 AMPPcP bound structure shows four regions of higher deviation (1-4). (C) LigPlot of the coordinating residues. (D) RMSD for each residue. Regions of higher RMSD values are located at the termini and linker regions (gray boxes). Residues coordinating the ligand are very well preserved (red arrows and red dots) and have an RMSD lower than the overall RMSD of 0.7 Å.

8.3 SUMMARY OF THE OBTAINED RESULTS FOR THE MINI-NBDS

In summary, we can design a stably folded minimal ATP binding domain, which we found by single-molecule force spectroscopy. Furthermore, the specificity for adenosine is still incorporated in this lobe II domain, even in the absence of lobe I. Ligand-binding analysis reveals an ATP binding constant of $K_D = 15\,\mu M$ and a $k_{off} = 0.12\,s^{-1}$. The overall protein stability is slightly different between Mini-NBD-183-359 and Mini-NBD-183-383. The force unfolding curves showed that lobe IIa in Mini-NBD-183-359 is weakly folded in the absence but stably folded in the presence of ligand. For Mini-NBD-183-383, we showed that the additional 20 residues at the c-terminal part rescue the weak stability in the apo state. We were able to determine SAXS data as well as crystal structures of the Mini-NBD-183-359 and Mini-NBD-183-383 in the ligand-bound and unbound state. The SAXS data showed that upon ligand binding the domains become more compact and hence more stable. The crystal structures confirmed that both proteins are well structured, and the active site is well-positioned. This leads to an identical binding conformation of the ligand, compared to EcNBD. Furthermore, some flexible loop regions in the Mini domains are tilted, compared to the full-length structure. Hence, lobe I is influencing the positioning of certain structural elements in lobe II.

Part III

DISCUSSION

9

FUNCTIONAL IMPLICATIONS FROM FOLDING AND UNFOLDING PATHWAY ANALYSIS

9.1 THE HIGHLY CONSERVED C-TERMINAL HELIX IS KEY TO STRUCTURAL RIGIDITY

The c-terminal helix connects lobe I and lobe II. It is nearly buried in the structure and acts as a molecular glue. As a result, the protein has a high mechanical stability which results in high unfolding forces. In the n-/c-terminal pulling geometry, we found that lobe I or lobe II unfolding is preceded by the unraveling of the c-terminal helix. This helix, which is sequentially part of lobe II, is structurally inserted into lobe I. Experiments with a construct missing the c-terminal helix showed dramatically lower unfolding forces ($\approx$ 12 pN) and incomplete refolding of the protein [9, 79]. Intriguingly, a population of unfolding events with the same low forces are observed in the refolding path of EcNBD, arguing that the c-terminal helix is not in the native conformation. Indeed, the stability of this state is dramatically reduced to only one third compared to the natively folded EcNBD. Hence, proper positioning of the c-terminal helix is a critical step in the refolding process to the native fold.

The c-terminal helix and the proximate linker connect the NBD with the SBD. This conserved proximate linker region plays an essential role in the allosteric mechanism of Hsp70. In the ADP-bound form, the linker is flexible. Hence, NBD and SBD are in the closed conformation, where the lid domain covers the substrate-binding site. Conversely, in the ATP-bound form, the linker interacts with the SBD and NBD interface [33, 59, 92] and Hsp70 adopts the open conformation (substrate-binding site is open). Importantly, while ligand binding influences the linker region, it does not affect the c-terminal helix conformation. The c-terminal helix, therefore, acts as a stability element, which allows large movements of the Hsp70 domains but keeps the fold of NBD rigid.

Moreover, the c-terminal helix is highly conserved in the sugar-kinase family fold. It has been shown that the structural similarity between actin, hexokinase, and NBD is strikingly high [36, 37]. All of these proteins undergo quite large structural rearrangements induced by ligand-binding [2, 94]. The c-terminal helix is structurally highly conserved, and hence, we propose that the structural rigidity of these protein family members is to a significant extent a result of this c-terminal helix connecting lobe I and lobe II.

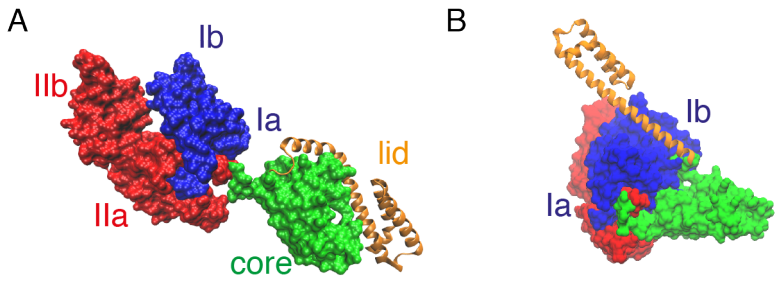


Figure 44: Crystal structure of the open and closed forms of DnaK. NBD lobe I in blue, NBD lobe II in red, SBD lid domain (orange) and SBD core domain (green). (A) Closed conformation of DnaK (apo or ADP-bound state, PDB: 2KHO). (B) Open conformation of DnaK (ATP-bound state, PDB: 4B9Q).

9.2 DISTINCT ROLES OF LOBE I AND LOBE II ARE NECESSARY FOR ALLOSTERIC FUNCTION

As discussed in the results sections (Chapter 4 and Chapter 5), most of the ATP-binding free energy is retained in lobe II ($-10.4\,k_BT$ for lobe II and $-14.2\,k_BT$ for whole EcNBD). Correspondingly, most of the ligand coordinating residues are located in lobe II (13 out of 18) and only 5 out of 18 in lobe I. However, many studies have shown that critical ligand *sensing* (as opposed to ligand *coordinating*) residues are mostly located in lobe I and are highly conserved among Hsp70 NBDs [10, 59, 76, 106, 123, 137, 139]. This shows the distinct role of each lobe. Lobe II is necessary for sufficient ligand binding and lobe I for sensing the nucleotide state in the active site.

Lobe I can transmit the signal to the substrate-binding domain, which results in allosteric coupling of NBD and SBD. In the apo or ADP state of NBD, the SBD adapts the so-called closed conformation (Figure 44A). Here, the α-helical part of SBD stays in contact with the β-sheet domain of SBD and hence covers the substrate binding site. The transition from the apo to the ATP-bound state is coupled to a wide structural motion and rearrangement of lobe I with respect to lobe II [58]. This motion moves lobe Ia closer to the SBD, and lobe Ib becomes positioned above the nucleotide binding cleft (Figure 44B). This led us to conclude that lobe I is less relevant for the binding of ligand itself; it is more important in signal progression to the SBD.

9.3 LOBE I AS A REGULATORY ELEMENT DURING OXIDATIVE CELL STRESS

Lobe II is a fast-folding unit and can therefore be a potential regulatory element for the folding state of NBD. During oxidative stress

conditions, lobe I is unfolded [133]. Winter *et al.* reported that upon low levels of ATP and oxidative stress, the tryptophan fluorescence of NBD (located in lobe I: Trp102) drops, arguing for an unfolding of this domain [133]. Another study showed that upon glutathionylation of Cys15, the ATP turnover decreases significantly and the structure is reversibly altered [135]. Our results with Mini-NBD showed that lobe II binds adenosine explicitly but is unable to hydrolyze ATP to ADP. Furthermore, lobe II forms rapidly and stably. Hence, we propose that under oxidative conditions, lobe II is still folded, while lobe I is unfolded due to the modification of Cys15. Lobe II does, in this scenario, guarantee fast refolding of the lobe I, after non-oxidative-stress conditions are re-established.

9.4 LOBE II AS A FOLDING SEED FOR NBD

The mtNBD protein is, in contrast to EcNBD, not able to refold after unfolding it in single-molecule force experiments. To investigate which part of mtNBD impairs the refolding, we designed chimeric variants of mtNBD. Grafting lobe IIb into the mtNBD sequence did not result in a natively folded protein. Based on a study which showed in *in vivo* mitochondrial import assays that a stretch of 11 amino acids from lobe IIa is enough to obtain a more soluble protein [12], we designed chimeric proteins with the same stretch of amino acids, with and without the lobe IIb region from EcNBD. We could express all chimeric proteins in *E.coli*. In our case, the co-expression of chaperones (pETDuet-1) was not necessary, as is required for Ssc1 (yeast Hsp70) [13, 104]. All proteins could be purified from the soluble fraction, except for mtNBDΔ^{236}NGVFEVKS^{243}, which formed aggregates. In disagreement with Blamowska *et al.*[12], the chimeric proteins, with less than the whole lobe II region of EcNBD, did not fold properly. An explanation for this discrepancy can be found in the different readouts. In *in vivo* import assays, the protein is expressed in the yeast cell and is imported. After import into the mitochondria, the aggregation state of mtHsp70 can be observed without Hep1 present. In the experiments presented in this thesis, the protein was directly purified from *E.coli* and investigated at the single-molecular level. Optical tweezers data are much more sensitive in detecting partially refolded proteins compared to the mitochondrial import assay.

It is indeed essential to graft the whole lobe II region of EcNBD into the mtNBD to enable refolding. This proposes EcNBD lobe II as a folding seed. Already the refolding data of EcNBD provide evidence that lobe II is necessary to act as a folding platform for lobe I. To make a compelling case for the assumption that the whole lobe II region is required to act as a folding seed for the NBD fold, the folding-incompetent mtNBD was made folding-competent by grafting the lobe

II region of EcNBD into the mtNBD sequence. These experiments also indicate that the impaired folding ability of lobe II in mtNBD is responsible for the folding incompetence of this protein and can be rescued when a folding-competent lobe II is present.

The interdomain linker between NBD and SBD is thought to be responsible for the aggregation propensity of mtNBD [11, 13]. Our experiments show that the interaction of the linker with NBD is not the only critical structural element accountable for aggregation. Also lobe II, when folding-incompetent, might lead to aggregation.

Lobe II as an identified folding seed for NBD may in the future help in the engineering of other folding-competent members of the sugar-kinase family, such as actin, which is notoriously difficult to purify recombinantly. This would enable the creation of mutant variants of actin and allow *in vitro* studies of the actin cytoskeleton.

9.5 BIOLOGICAL RELEVANCE OF A FOLDING INCOMPETENT MTNBD AND THE ROLE OF HEP1

Our data showed that EcNBD can refold readily. In contrast, the mtNBD variant from yeast is not able to refold independently of its co-chaperone Hep1. Why is the folding behavior of the mtNBD so different than EcNBD? The sequence homology of both is higher than 60% (emboss needle tool [95]). Both structures are highly conserved, and one might naively expect a similar folding propensity. Above, I discussed that lobe II is a folding seed for the NBD structure. Folding-incompetent NBD proteins can be a result of an impaired refolding ability of lobe II. The question I want to address now is why nature has designed a folding inefficient mtNBD protein?

The different folding timescales can be a result of their different subcellular localization and distinct functional requirements. The mtNBD protein is synthesized in the cytosol of the cell and has to be imported into the mitochondrial matrix. To direct the mtHsp70, it carries an n-terminal mitochondrial import sequence [27]. During this import process, mtHsp70 stays in an unfolded state. For the progression of the transport process it would be a disadvantage to have lobe II folded. In the mitochondrial matrix, Hep1, a chaperone for mtHsp70, associates with the translocated mtNBD and helps it adopt its native structure [12, 104]. In contrast, DnaK is also synthesized in the cytosol but is not transported through membranes. Therefore, DnaK has evolved to fold without the help of chaperones and can fold rapidly in the cytosol to become active and functional as an Hsp.

In the mitochondrial matrix, mtHsp70 folds with the help of Hep1. We propose that the action of Hep1 is to stabilize lobe II and therefore, enhance folding. Blamowska *et al.* showed that Hep1 interacts with the mtNBD and the interdomain linker sequence and hence supports

folding [13]. Furthermore, for the human Hsp70, it was reported that the co-chaperone BP1c binds to the lobe II region and forms a stable complex [103]. The stabilization of lobe II and the slow refolding of lobe I can indeed be a regulatory tool which is used in eukaryotes to allow the function of a protein only at its accurate localization.

In our measurements with mtNBD, we found several weakly folded portions. We propose that Hep1 binds to these partially collapsed states of lobe II, supports folding of the entire lobe II domain and thereby enables lobe I refolding. However, binding of Hep1 has surprisingly not been observed in our measurements [22], which could be due to low affinity or slow binding. By using the constant velocity mode for measurements, partially folded states get disrupted quickly, which interferes with slow Hep1 binding. However, Blamowska *et al.* showed in *in vitro* refolding of mtHsp70 that Hep1 is released upon nucleotide binding to mtHsp70 [11]. Our optical tweezers experiments suggest that lobe II may be a possible binding site for Hep1 action. Lobe II is, when correctly folded, able to bind Mg-ATP or Mg-ADP. In this scenario, after the ligand is bound, Hep1 is released, as proposed by Blamowska *et al.*, and lobe I can fold in the presence of folded lobe II.

[22] *Hep1 binding would likely lead to a stabilization of particular states.*

10

DESIGN OF A SPECIFIC MINIMAL ATP-BINDING DOMAIN BASED ON SINGLE-MOLECULE DATA

10.1 STRUCTURE AND FUNCTION ANALYSIS OF MINI-NBD

10.1.1 *Lobe II as an independent and stably folding domain*

We expressed and purified stably folded lobe II in *E.coli*. The design of this domain was based on single-molecule force spectroscopy data, which revealed this domain to be fast and natively forming under load. In a further analysis, we could show that lobe II in isolation specifically binds adenosine ligands and un- and refolds readily. Bound ligand also increases the stability of the domain significantly. Our results are consistent with findings by Shomura *et al.*, who reported the crystal structure of the human Hsp70-NBD lobe II in the AMP-PcP bound form and its cochaperone BP1c bound to lobe II (PDB:1XQS) [103] (Figure 45). In contrast to our approach, they purified the fragment after proteolytic digestion of the full-length NBD, which results in BP1c bound to lobe II. Nevertheless, also in this study, lobe II alone is stable and adopts a native fold.

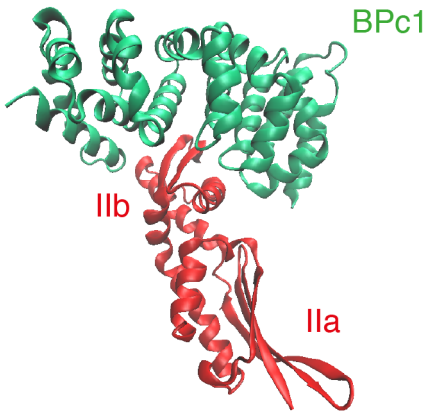


Figure 45: Crystal structure of human Hsp70 NBD lobe II (red) and BPc1 (green) bound (PBD:1XQS).

10.1.2 *Importance of lobe I for sub-structural parts of lobe II*

Our crystal structure shows that the Mini-NBD domains (Mini-NBD-183-359 and Mini-NBD-183-383) are correctly folded. The overlay with the published DnaK ATP-bound structure (PDB:4B9Q) reveals some differences in two loop regions (Figure 43D). These differences are due to the absence of lobe I and argue that the lobe domains interact with each other. Interestingly, the loop region in lobe IIb tilts away when lobe I is absent. Our structure of Mini-NBD shows that this region seems to be quite flexible and can adopt different conformations depending on the interaction with lobe I. It is known that this loop is an essential site for GrpE binding. Furthermore, GrpE influences the lobe-lobe interface of NBD by rotating lobe I against lobe II, which results in enhanced nucleotide release [15, 44]. It is possible that GrpE tilts the loop region in lobe IIb to disrupt the interaction with lobe I leading to a fine-tuned signal in the lobe-lobe interface.

The partly unresolved region in lobe IIa can be due to the missing interactions with the highly conserved helix in lobe Ia (residues 172 to 182). This region is important for DnaJ binding and might be repositioned by the linker movement upon adaption of the open conformation of DnaK allowing proper DnaJ interaction. Again, this region seems to be quite flexible and a tool for signal transduction in the DnaK cycle.

Furthermore, the crystal structure shows that the c-terminal helix is unstructured in the Mini-domain. This is in agreement with our optical tweezers data where the c-terminal helix is stretched with the DNA-handles and hence is already unfolded or loosely bound. From a structural point of view, the c-terminal helix is inserted into lobe I. Therefore, lobe I is needed to fold and position the c-terminal helix correctly. This again provides evidence that the c-terminal helix is mainly a stability element for the NBD fold.

10.1.3 *Importance of lobe I for nucleotide binding and hydrolysis*

Omitting lobe I leads to a catalytically inactive ATP-binding protein (Mini-NBD). ATP can still be bound, but the off-rate is 100-fold higher compared to the full-length NBD protein. Lobe I and lobe II form a binding cavity for ATP which leads to regulated nucleotide binding and release. The catalytic activity is significantly higher in the full-length protein which argues that lobe I can discriminate between the nucleotide-bound states (ATP or ADP bound). The hydrolysis of ATP is maintained by the in-line attack of a water molecule [36, 87, 129]. Residues K70 and E171 are the sensing elements of the nucleotide-bound state (Figure 46) and play an important role in the allosteric mechanism of DnaK [122]. K70 is positioned closely to the γ-phosphate

and has a stabilizing role during the transition state of the phosphate cleavage reaction. Its role is similar to the arginine finger in GTPases [99]. E171 is proposed to coordinate the attacking water and to serve as a proton acceptor. T199 is needed to guide the water molecule to the γ-phosphate [58]. Taken together, lobe I provides all residues involved in the ATP hydrolysis reaction and lobe II fulfills the role of a stable nucleotide binding element. Accordingly, the Mini-NBD domains can bind ATP with a micromolar affinity but are not able to hydrolyze ATP to ADP.

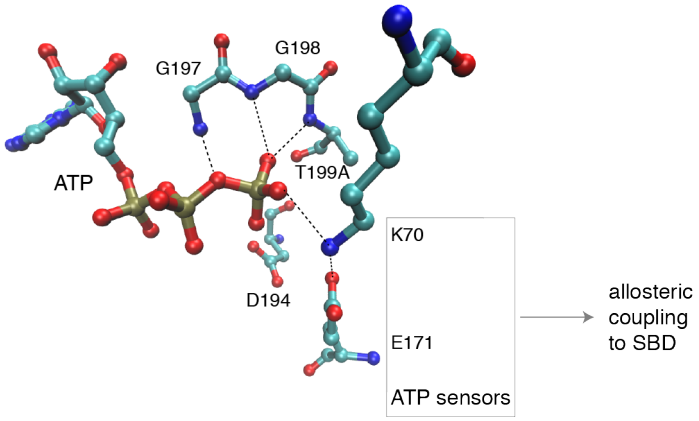


Figure 46: Residues involved in the ATP hydrolysis reaction. K70 and E171 build the ATP sensing element and transfer the ATP-bound signal to the SBD. T199 guides the catalytic water into the right position. G197, G198, and D194 are located in the highly conserved loop region and are involved in ATP coordination. (O (red), N (blue), C (cyan), P (gold))

10.2 EVOLUTION OF THE SUGAR-KINASE FAMILY

Hsp70 is a member of the sugar-kinase family. This family characterizes a highly conserved butterfly-like 3D structure. Structural characteristics are: the c-terminal helix, the lobe-lobe architecture, the ATP-binding site and the connection of lobe I with lobe II (Figure 7). Figure 47 shows, as an example, the crystal structure of actin and EcNBD in the ATP-bound form. The sequence alignment reveals a sequence identity of only 14% (Emboss needle tool [95]), but the structural conservation is strikingly high.

Flaherty *et al.* reported that the RMSD of C_α between Hsc70 and actin in apo is 2.3 Å [36]. The alignment of the nucleotide-bound form of both proteins is even better with an RMSD of 0.66 Å. This early study already showed that actin and Hsp70 proteins are structurally highly similar and hence likely evolutionary related. The same group proposed a common ancestral protein which can bind nucleotide and dimerize (Figure 48).

The highly conserved lobe-lobe architecture of sugar-kinase proteins argues for a domain duplication event, which occurred twice. The ability of the Mini-NBD to specifically bind adenosine nucleotides makes it a perfect candidate for such a common ancestral domain. In our experiments, one crystallization unit harbors two Mini domain proteins, which seem to form a dimer. As Figure 49 shows, Mini-NBD-183-359 (Figure 49A and Figure 49B) and Mini-NBD-183-383 (Figure 49C and Figure 49D) are arranged differently. In both conformations, the two lobes are tilted against each other and do not align next to each other. This difference in orientation may arise due to the already evolved lobe II, which has specialized for the interaction with lobe I. Maybe the common ancestral domain was similar to the lobe II we know today, but the interface has not been optimized for lobe I and hence dimerization between to common ancestral domains was possible.

In the proposed scenario, the duplicate of the common ancestral domain, namely lobe I, evolved in divergent-evolutionary scenarios according to its functional requirements. For example, lobe I in Hsp70 is the sensory element of the nucleotide state and transducer of the allosteric signal to the SBD. In actin, lobe I serves as the polymerization interface and allows large filaments to be assembled. Kawai *et al.* proposed a similar picture for the evolution of the glucose kinase family [54]. Taken together, our finding of lobe II being a stable nucleotide binding domain fits in the evolutionary scenario of the sugar-kinase family fold. We propose a lobe II-like domain as the common ancestral domain.

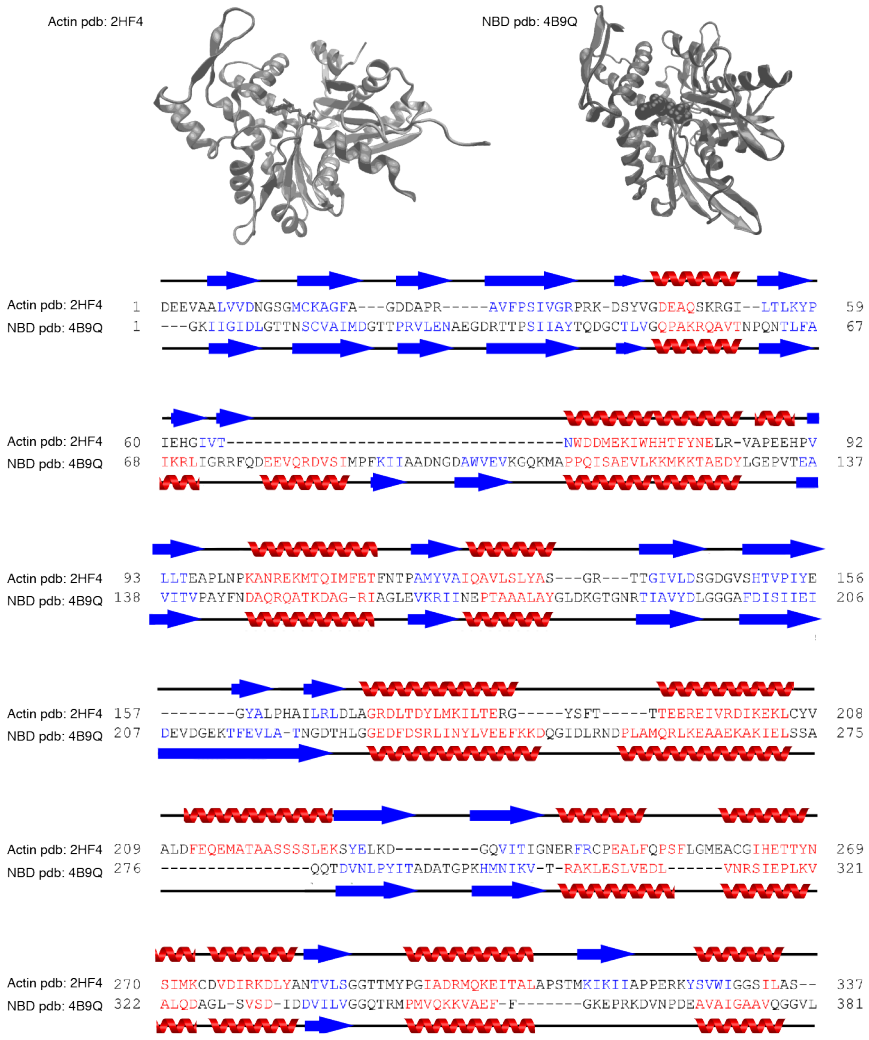


Figure 47: Sequence alignment of actin and EcNBD. The top shows the crystal structure of actin and EcNBD (pdb: 2HF4 and 4B9Q, respectively). The bottom shows the plot of the sequence alignment of actin and EcNBD. The respective secondary structural elements are color coded: β-sheets (blue) and α-helices (red). The alignment was performed with PROMALS3D [90].

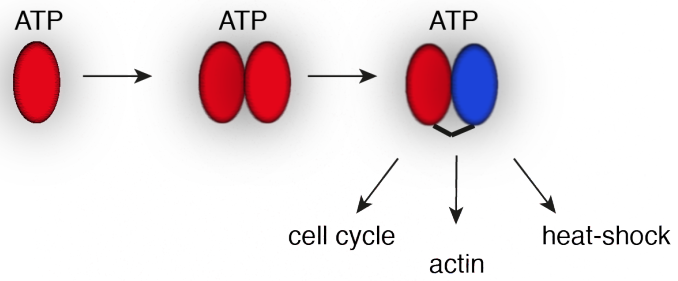


Figure 48: Evolutionary scenario of the sugar-kinase family. First, a lobe II-like common ancestral domain capable of ligand binding (red) existed. After gene duplication, the two similar domains were able to dimerize. Then, the two domains connected via a linker and one lobe evolved further in a divergent-evolution scenario. Finally, proteins that are known today are involved in cell cycle, cytoskeleton and heat shock processes. Adapted from [6].

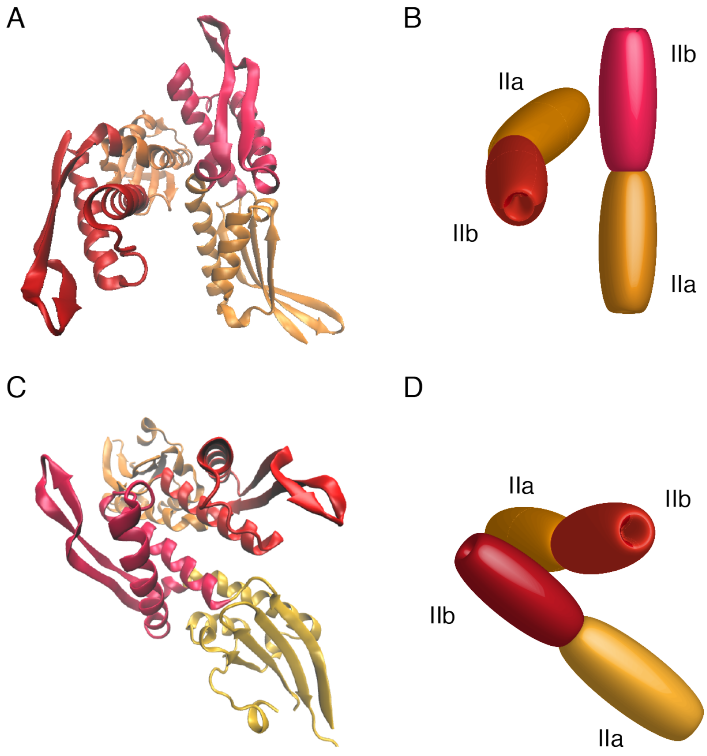


Figure 49: Dimerization of Mini-NBD-183-359 and Mini-NBD-183-383. (A) Crystal structure of Mini-NBD-183-359 in the ADP-bound form. (B) Graphical representation of the dimer formation of Mini-NBD-183-359 shown in A. (C) Crystal structure of Mini-NBD-183-383 in the AMPPcP-bound form. (D) Graphical representation of the Mini-NBD-183-383 dimer formation shown in C.

Part IV

OUTLOOK

11

THE LOBE-LOBE INTERFACE OF ECNBD AND THE INFLUENCE OF GRPE

In the presented experiments, the unfolding and refolding pathways of NBD have been characterized in great detail. The distinct roles of lobe I and lobe II have been addressed. The next step further, could include the aspects of how the communication between the lobes or the interactions of the whole NBD with the SBD is coordinated. It has already been shown that the dynamics of the individual subdomains in EcNBD are essential for the sensing of the nucleotide state and, therefore, the allosteric mechanism [10, 136].

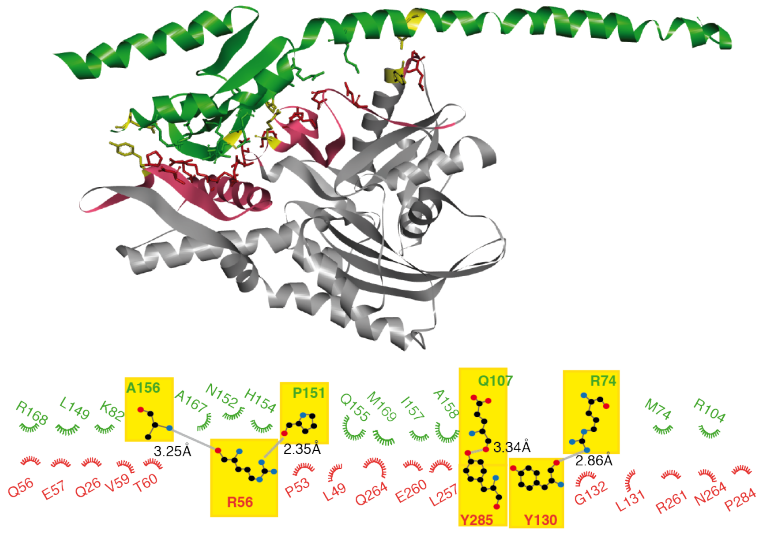


Figure 50: Crystal structure (PDB:1DKG) and LigPlot of GrpE bound to NBD. All residues involved in the interaction of GrpE and NBD are represented as sticks. The yellow shade corresponds to direct interacting residues between both domains. The LigPlot was performed using the free software tool [65].

In the case of nucleotide-exchange factors (NEFs), flexible structural parts and specific binding regions are used to influence the nucleotide state and the lobe-lobe interaction [15]. Small-angle rotations of lobe IIb against Ib upon nucleotide binding are one of the elements GrpE (NEF of Hsp70) takes advantage of [134]. GrpE uses this rotational flexibility to facilitate ADP release. Through this mechanism, GrpE accelerates the ATP turnover rate 1200 times [89]. Figure 50 shows closely interacting residues of GrpE with the NBD of DnaK. The mechanism of GrpE action involves structural elements of NBD, such as an ex-

posed and essential ligand binding loop region in lobe I. It is known that a point mutation within this region (G32D) dramatically reduces the GrpE binding ability [18]. Furthermore, two salt bridges are highly conserved in DnaK, and it is reported that these are involved in GrpE binding and action [15]. The questions which are still open are: (1) how stable is the lobe-lobe interface and how is it altered in the presence of nucleotides? (2) How does GrpE interfere with this stability to accelerate nucleotide release?

To answer these questions, the pulling geometry needs to be changed to probe the lobe-lobe interface. Therefore, an EcNBD protein was designed that can be pulled from the top of lobe I and lobe II, called EcNBD-EA[23]. In this active pulling geometry, the action of client proteins and nucleotide can be investigated in more details based on stability changes. First experiments showed that in this active pulling geometry, EcNBD can populate two unfolding pathways (coupled and uncoupled). In the coupled unfolding pathway, lobe I and lobe II interact with each other. In the uncoupled case, lobe I and lobe II unfold independently of each other. Maybe this pathway switch is connected to a different stabilization of lobe I and lobe II, but why precisely two pathways get populated and what the determining factor can be, is ongoing research.

[23] *Residues E47C and A290C are used as attachment points for the force application in optical tweezers experiments*

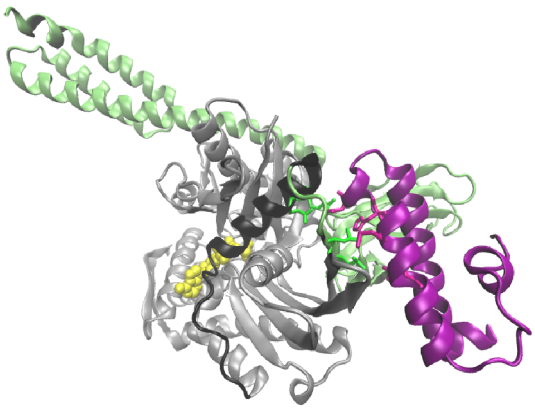


Figure 51: Crystal structure of DnaJ bound to DnaK (PDB:5NRO). The NBD domain is colored in gray, the c-terminal helix in black, the SBD in green, ATP in yellow and DnaJ in purple.

Preliminary experiments with the ligand and GrpE bound to EcNBD-EA look very promising. They can help to answer many mechanistic questions about the NBD lobe-lobe communication and the interaction of the ligand and nucleotide-exchange factor with NBD. Open questions which could be addressed are: (1) Why are two unfolding pathways populated? (2) What is the determining structural feature that decides what path will be populated? Maybe the shift in paths is one element to communicate the ligand-bound state to the SBD. Furthermore, assumptions about the interface of lobe I and lobe II can

be made in the absence and presence of ligand and GrpE. These insights would help to further understand the mechanism of GrpE and the translation of the nucleotide state to critical allosteric elements.

DnaJ, another cochaperone of DnaK, is a member of the J-domain family. The $\approx$ 65 residue long DnaJ forms a helical hairpin structure [57]. These J-domain proteins extend the substrate spectrum of DnaK by shuttling unfolded protein chains to DnaK. Both substrate and DnaJ stimulate the ATPase activity of DnaK synergistically [62]. Recently, the atomic structure of DnaJ interacting with DnaK could be solved in the ATP-bound open conformation of DnaK [57] (PDB:5NRO Figure 51). DnaJ interacts with the NBD at the lower part of lobe IIa and Ia in a linker region. It would be interesting to see if DnaJ binds to isolated NBD and what the effects on the stability of individual subdomains of NBD are.

These proposed experiments will pave the way for extended studies to analyze the whole cycle of chaperone activity of Hsp70 and its clients at the single-molecular level.

Part V

MATERIAL AND METHODS

12 MATERIAL

In this chapter, all used materials are listed and described.

12.1 USED COMMERCIAL KITS

- QIAprep Spin Miniprep Kit
- STRATAGENE QuikChange Lightning Multi Site-Directed Mutagenesis Kits
- QIA quick PCR Purification Kit

12.2 GROWTH MEDIA FOR BACTERIA

All media were prepared in double distilled water and had afterward been autoclaved. For the preparation of media for agar plates the LB-medium was supplemented with 2 % of bacto-agar. To prepare media or plates with antibiotics, it is worth mentioning that the antibiotic has to be added after the autoclavation step.

- 1 l LB_0-medium (Lysogeny broth Luria): 10 g Bacto Trypton (AppliChem), 5 g yeast extract (AppliChem), 5g NaCl (AppliChem)
- 1 l LB_0-medium with carbenicilline: 20 µg mL^{-1} carbenicilline (AppliChem) in LB_0-Media
- SOC-medium: 2 % (w/v) bacto-tryptone, 0.5 % (w/v) yeast extract, 8.56 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose

12.3 BUFFERS

- 1 l of 50 mM NaP-buffer pH 8.0: 172 ml Na_2HPO_4 (Merck), 28 ml NaH_2PO_4 (Merck)
- Buffers for Ni-NTA column purification
  - buffer A: 100 mM NaP-Puffer, 10 mM Imidazole (AppliChem), 2 mM TCEP (Sigma Aldrich)
  - buffer B: 50 mM NaP-Puffer, 250 mM Imidazole (AppliChem), 2 mM TCEP (Sigma Aldrich), 100 mM NaCl
  - buffer C: 100 mM NaP-Puffer, 10 mM Imidazole (AppliChem), 2 mM TCEP (Sigma Aldrich), 7 M Guanidiniumchloride (AppliChem)

  - buffer D: 100 mM NaP-Puffer, 500 mM Imidazole (AppliChem), 2 mM TCEP (Sigma Aldrich), 7 M Guanidiniumchloride (AppliChem)

- PBS-buffer: PBS buffer tablets ordered from Sigma Aldrich
- 50TKM-buffer: 50 mM TRIS pH 7.5, 50 mM KCl, 5 mM $MgCl_2$
- Buffer for hydrophobic interaction chromatography
  - 50TKM: 50 mM TRIS pH 7.5, 50 mM KCl, 5 mM $MgCl_2$
  - 50TKM-AS: 50 mM TRIS pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, 1 M ammonium sulfate
- Buffer for agarosegel
  - 50x TAE-buffer: 2 mM Tris/Acetat pH 8.0 (AppliChem), 2 mM EDTA (AppliChem)
- Buffer for SDS-gel
  - 1x Laemmli-buffer: 300 mM TRIS-HCl (AppliChem) pH=6.8, 15 mM EDTA (AppliChem), 12 % SDS, 30 % glycerol (AppliChem), 0.06 % bromphenole-blue
  - 10x running buffer for SDS-gel: 25 mM TRIS-HCl (AppliChem), 0.1 % SDS, 192 mM glycine (Sigma)
- Buffer for ATPase assay
  - Assaybuffer: 40 mM HEPES pH 7.6, 50 mM KCl, 11 mM $Mg(OAc)_2$
- Buffer for limited proteolysis assay
  - protease assay buffer (50 mM HEPES pH 7.5, 150 mM KCl, 1 mM $CaCl_2$, 5 mM $MgCl_2$)

12.4 USED ENZYMES

- Antarctic phosphatase (New England BioLabs)
- Phusion polymerase and GC-Phusion buffer (Finnzymes)
- Restriction enzyme NdeI, BsaI, BamHI with buffer CutSmart (New England BioLabs)
- Crimson Taq DNA Polymerase and Crimson Taq buffer (New England BioLabs)
- Quick Ligase und Quick Ligase buffer (New England BioLabs)
- Lactate dehydrogenase pyruvate kinase Mix (Sigma Aldrich)
- Catalase (Sigma Aldrich)
- Glucose oxidase (Sigma Aldrich)

12.5 LIST OF USED BACTERIA STRAINS

- XL1-Blue competent cells (Agilent Technologies)
- BL21-CodonPlus (DE3)-RIPL competent cells (Agilent Technologies)
- NEB Turbo Competent *E.coli* (C2984H) (New England Bioloabs)
- T7 Express lysY/lq Competent *E.coli* (High Efficiency) (C3013I) (New England Biolabs)

12.6 INSTRUMENTS

Instrument	**Model**	**Company**
Absorption spectrometer	DU 730	Beckman Coulter
Analytical Scales	XR 125 SM	Precisa
Circular Dichroism Spectrometer	J-815	Jasco
Electrophoresis Power Supply	Power Source	VWR
Fluorescence spectrometer	FP8500	Jasco
FPLC system		Bio-Rad
High Pressure Cell Disruptor	Benchtop	Constant Systems
HPLC system		Jasco
Large fixed angle centrifuge	Avanti J-E	Beckman Coulter
Microcentrifuge	Microfuge 18	Beckman Coulter
Microvolume spectrometer	NanoDrop ND-1000	PEQLAB
PCR cycler	pe STAR	PEQLAB
Peristalitc pump	Pumpdrive 5206	Heidolph
pH meter	CG 840	Schott
SDS-PAGE cell	Mini-PROTEAN 3	Bio-Rad
Swinging bucket centrifuge	Rotanta 460R	Hettich
UV/VIS spectrometer	Nanodrop	Thermo Fisher Scientific

Table 5: List of used instruments

12.7 COLUMNS FOR PROTEIN PURIFICATION

Type	Product name	Company
Buffer exchange column	illustra NAP-5, NAP-10, NAP-25	GE Healthcare Life Sciences
High pressure size exclusion	YMC -Pack Diol-300	YMC
High pressure size exclusion	Yarra 3 µm SEC-3000	Phenomenex
Ni-NTA affinity colunm	HisTrap HP 1mL	GE Healthcare Life Science
Size Exclusion Chromatography (SEC)	Superdex 200 10/300 GL (17-5175-01)	GE Healthcare Life Science
Hydrophobic Interaction Chromatography (HIC)	HiTrap Phenyl HP	GE Healthcare Life Science

Table 6: List of used purification columns

12.8 OTHER MATERIALS

Type	Product name	Company
Cover slips	High Precision Cover Slips 0.17 mm	Carl Roth
Sealing film	Parafilm	Bemis
Streptavidin functionalized beads	Streptavidin Coated Microspheres	Bang Laboratories
Silica beads	One micron Silica-Micropsheres	Bang Laboratories

Table 7: List of used materials

13 METHODS

13.1 CLONING OF DIFFERENT NBD PROTEIN CONSTRUCTS

13.1.1 *Cloning of EcNBD, loop-insertion variants and lobe II pulling variant*

The NBD gene was cloned in pET11a (Novagen) using the NdeI and BamHI restriction sites, was kindly provided by Dr. Gabriel Zoldak (nucleotide sequence can be found in Section A.2.1). This construct already contained the n- and c-terminal cysteines and spacers necessary for optical tweezers experiments and a c-terminal His_6-tag for protein purification (this construct is further referred as EcNBD). Based on this construct all other constructs were cloned using quick-change mutagenesis (Agilent) and specific primers. Dr. Gabriel Žoldák kindly provided the mtNBD protein and mtNBD optical tweezers raw data.

13.1.2 *Cloning of Mini-NBDs*

The Mini EcNBD constructs were amplified from the EcNBD gene sequence using primers amplifying BsaI-(183-383)-BamHI (Mini-NBD) or BsaI-(183-359)-BamHI (Mini-NBD 183-359). The Sumo protein was amplified using NdeI-(His_6-Sumo)-BsaI. Both fragments were digested using the respective restriction enzymes and ligated in a one-step ligation into a pre-digested pET11a vector (restriction sites: NdeI and BamHI). The ligation product was purified using the Quiagen PCR purification kit. The purified vector was transformed in *E.coli* XL1 blue and platted on carbenicillin agar plates. Colony polymerase chain reaction (PCR) verified if a colony carried the plasmid of interest. From positive clones, the plasmid was purified using the plasmid purification kit from Quiagen and send for sequencing (GATC Biotech AG). In the last step the plasmid was transformed into *E.coli* BL21 codon+ for expression.

13.1.3 *Agarose gel electrophoresis of DNA*

All amplified deoxyribonucleic acid (DNA) fragments and plasmids were analyzed by agarose gel electrophoresis to ensure the right size and purity. Agarose gels were prepared using 1x TAE buffer mixed with 1 % of agarose and cooked until the agarose was resolved. Next, SYTOX green was added according to the manufactures instructions

Name of construct	Purpose
Ins45	Investigation of the un-/refolding pathway probing for lobe Ib
Ins183	Investigation of the un-/refolding pathway probing for lobe-lobe connection
Ins290	Investigation of the un-/refolding pathway probing for lobe IIb
Ins364	Investigation of the un-/refolding pathway probing for c-terminal helix
NBD 183-383	Unfolding of lobe II only in presence of lobe I
Mini-NBD-183-359	Minimal ATP-binding domain without c-terminal helix
Mini-NBD-183-383	Minimal ATP-binding domain with c-terminal helix
50%mtNBD	Refolding competent mtNBD
75%mtNBD	Lobe IIb from EcNBD proof for lobe II as folding seed
75%mtNBDΔ^{236}NGVFEVKS^{243}	Lobe IIb and aa stretch 236-243 from EcNBD proof for lobe II as folding seed
mtNBDΔ^{236}NGVFEVKS^{243}	Amino acids 236-243 from EcNBD proof for lobe II as folding seed
NBD-E47C-A290C	Pulling top of lobe I and II GrpE and lobe-lobe interaction NBD variant
NBD-D45C-A290C	Pulling top of lobe I and II GrpE and lobe-lobe interaction NBD variant
NBD-G107C-A290C	Pulling top of lobe I and II GrpE and lobe-lobe interaction NBD variant
NBD-E47C-A290C-	Pulling top of lobe I and II with loop insert
Ins183	GrpE and lobe-lobe interaction NBD variant

Table 8: List of used protein constructs.

(Molecular Probes, peq green) and the solution was filled in a precasted container (Biorad) to solidify to a gel. The DNA samples and the DNA standard were supplemented with loading dye (Thermo Fisher) and loaded into the gel pockets. The gel ran in 1xTAE buffer at 120 V. Finally, the DNA was visualized under UV light.

13.1.4 *Bacteria cultivation conditions*

All *E.coli* bacteria strains (*E.coli* BL21 codon+ and XL1 blue) were cultivated in LB-media with carbenicilline. A single colony or a few microliter of a glycerol-stock were used to set up a 20 ml overnight culture. This culture was incubated at 37 °C with a shaking speed of 130 RPM.

13.1.5 *Transformation and expression of EcNBD and EcNBD variants*

All NBD protein variants were transformed in *E.coli* BL21 DE codon+ competent cells. 1 ng of plasmid was incubated with 250 ml of competent cells heat shocked for 30 s at 42 °C and cultivated for 1 h at 42 °C and 160 RPM. Next, the bacteria suspension was stroked out on carbenicillin containing agar plates. After overnight incubation at 37 °C colonies were picked and a pre-culture was inoculated. The next day a glycerol stock from the pre-culture was prepared and a day-culture for the expression test or actual expression was inoculated 1:50 in fresh LB-medium with carbenicillin. The expression of the protein was performed using IPTG induction at OD_{600} = 0.6 and 3 h of expression time. After expression, the cells were harvested by centrifugation and the cell pellet was frozen at −20 °C.

13.2 PROTEIN PURIFICATION

13.2.1 *Purification of EcNBD, loop-insertion variants, lobe II pulling variant*

13.2.1.1 *Cell disruption and centrifugation*

The frozen cell pellet was defrosted and resuspended in buffer A with 10 mM $MgCl_2$. After resuspension, the cells were broken up using the French press system. To digest the cell membrane of *E.coli* and host DNA, lysozyme and DNAse I were added to the lysed cells and incubated for 1 h on ice. Next, the suspension was centrifugated at 17 000 RPM for 45 min. The supernatant containing the protein of interest was filtrated through a 0.2 nm filter and loaded onto the equilibrated (buffer A) Ni-NTA column.

13.2.1.2 *Affinity chromatography: Ni-NTA column purification*

For the Ni-NTA purification, the following protocol was used. First, the column was equilibrated with buffer A. Then, the protein was loaded onto the equilibrated column. Next, a washing step with 20 ml buffer A, followed by a denaturing step of the protein on the column using 10 ml of buffer C and 15 min incubation time, was performed. This step is necessary to remove bound proteins and get nucleotide-free protein. The guanidinium chloride was removed, flushing with 20 ml of buffer A. Last, the protein was eluted using buffer B. To remove the imidazole, the buffer was exchanged to 50TKM or phosphate-buffered saline (PBS) using NAP-illustra columns (GE Healthcare).

13.2.1.3 *Size-exclusion chromatography*

The final purification was performed using size-exclusion chromatography (SEC). The protein solution was loaded onto a Superdex 200 10/300 GL column (GE Healthcare), and as the mobile phase, either 50TKM buffer or PBS was used. The purity of the protein fractions were analyzed by SDS-PAGE. The pure NBD protein eluted after 30 min. Afterwards, the protein fractions were up concentrated, frozen in liquid nitrogen and stored at −80 °C.

13.2.2 *Purification of Mini domains*

The Mini EcNBD proteins were expressed as His_6-SUMO-tagged proteins, to ensure solubility. Cell disruption, affinity chromatography (Ni-NTA) and SEC were performed as described for EcNBD. After these steps the His_6-SUMO-tag was cleaved off, using the Senp2 catalytic fragment (self-prepared stock[24]) in a 1:50 molar ration and incubated overnight on ice. The next day a second Ni-NTA column was performed. Here, the Mini-NBD proteins were in the flow through because the c-terminal His_6-SUMO-tag is cleaved from the Mini-NBD and is bound to the column, as well as His_6-tagged Senp2. After that step, some SUMO-protein was still non-covalently bound to Mini-NBD. To remove the bound SUMO tag, a hydrophobic column was used.

[24] *Senp2 was expressed as a Hi−6-tagged protein. After NiNTA purification the protein needs to be in a high salt buffer (20 mM TRIS pH8, 300 mM NaCl, 2 mM TCEP). The second purification step, to obtain pure protein, was SEC.*

13.2.2.1 *Hydrophobic interaction chromatography*

First, the protein solution is precipitated using 1 M ammonium sulfate and incubated for 2 h at 4 °C on a shaker. The precipitated protein was filtered and loaded onto the equilibrated phenyl column (50TKM 1M ammonium sulfate buffer (50TKM-AS)). The column was washed with 10 column volume (CV) of buffer 50TKM-AS. Then the gradient was set to start with 100 % of 50TKM-AS and go down to 0 % 50TKM-AS in 20 CV. UV/VIS signals at 260 and 280 nm wavelength monitored the elution from the column. 1 ml fractions were collected during the

elution step and analyzed using SDS-PAGE. The fractions containing pure Mini-NBD protein were frozen in liquid nitrogen and stored at –80 °C.

13.2.3 *SDS-PAGE*

For all experiments 4- 20 % SDS-PAGE (Biorad) were used. The protein samples (with a final concentration of $\sim$ 20 µM) were diluted 1:1 with 2x Laemmli-buffer and incubated for 5 min at 95 °C. Protein samples and protein standard ladder were loaded and the electrophoresis was carried out for 45 min at 40 mA. The gel was stained with Coomassie Brilliant-blue solution and destained using a mixture of 80%water, 15%acetic acid and 5%methanol.

13.3 BIOCHEMICAL CHARACTERIZATION

13.3.1 *Circular dichroism*

To investigate the overall structure of a protein, far-UV light can be used to determine if a protein has a high content of α-helices, β-sheets or is in a random coil conformation. Polarized light is differently adsorbed by these structural elements and result in a different CD-spectrum.

13.3.1.1 *CD spectra measurement*

The protein was dialyzed into 2 mM Na-P buffer pH7.5 using NAP-5 columns and diluted to a final concentration of 2 µM. The exact concentration was obtained by Bradford assay (triplicates). Spectrum measurements were carried out at the CD-spectrometer (JASCO) at 20 °C. A 1 mm quartz cuvette (Hellma) was used and a total sample volume of 300 µl. The following parameters were set: band-width of the incident light path 1 nm, scanning speed set of 100 nm s^{-1}, D.I.T of 2 s and data pitch of 1 nm. Each spectrum was measured in the rage of 185–260 nm (wavelength) and accumulated for 30 times. The measured ellipticity θ_{obs} can be calculated to the molar ellipticity per peptide bond $[\Theta]_{molar}$ to obtained the standardized CD-curve:

$$[\Theta]_{molar} = \frac{\theta_{obs}(\lambda)}{N \times l \times c_{protein}} \tag{1}$$

with N the number of peptide bonds, l the length of the cuvette and c the protein concentration.

13.3.1.2 *Protein stability test using thermal denaturation*

The nucleotide binding domain has an overall α-helical CD-spectra. Therefore, the loss of signal due to denaturation at a wavelength of 222 nm can be recorded in so-called thermal denaturation experiments. The set parameters include: sampling rate of 0.3 s^{-1}, start temperature of 6 °C, target temperature of 70 °C, temperature ramp of 1 °C min^{-1} and a bandwidth of 3 nm. With this settings, the thermal denaturation of the proteins could be recorded and the stability of certain regions determined. With rising temperature the probability to populate the unfolded state increases. At a specific temperature, the so-called melting temperature (T_m), the denatured and natively folded state are in equilibrium. Beyond that temperature point, the protein predominantly populates the unfolded state. This picture can be extended from a simple two-state process to a multiple-step denaturation process with more than one transition. The recorded data was fitted to the following equation to obtain important thermodynamic parameters:

$$\theta = \frac{(\theta_N^0 + m_N \times T) + (\theta_U^0 + m_U \times T) \times \exp\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right)\right]}{1 + \exp\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right)\right]} \quad (2)$$

θ: observed ellipticity
θ_N^0 and θ_U^0: ellipticities of the native and the unfolded state
m_N and m_U: slope of the native and the unfolded branch
ΔH_{vH}: apparent van't Hoff enthalpy
R: universal gas constant
T: temperature in Kelvin

13.3.2 *Limited proteolysis assay*

The three-dimensional structure of proteins can be protected from proteolysis because cleavage sites are buried within the protein structure [17]. This assay can be used to probe if a protein structure is folded correctly or not. This so-called proteolysis assay was performed with the EcNBD and Mini domains. The variants were incubated with trypsin and the progression of digestion over time was analyzed using SDS-PAGE. If the proteins are well folded, no digestion bands should be visible (Figure 52).

To conduct the assay, first, the trypsin protease was dissolved freshly in protease assay buffer: 1 mg protease powder in 500 µL. Secondly, the assay buffer was complemented with 10 µM protein and incubated with 0.1 µM dissolved trypsin protease. After different time steps (e.g.

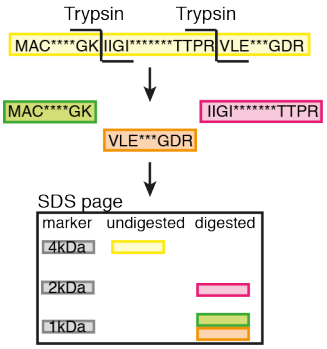


Figure 52: Limited proteolysis assay. The protease trypsin cleaves a peptide chain (yellow) after a specific sequence motif. The fragments (shown here in green, orange, and pink) can be visualized on a SDS-PAGE. Properly folded protein will appear on the gel as the undigested (yellow) band.

1, 5, 15, 30 and 60 min), 10 µL of the reaction where put in hot Laemmli-buffer (1:1 dilution) to stop the reaction. All samples were then analyzed on an SDS-PAGE. As a control EcNBD was used because its fold is resistant against trypsin digestion. To test the effect of nucleotides on the folded state we also incubated the proteins with 1 mM nucleotide (e.g., ATP, ADP, AMP) for 20 min in advance and proceeded as described.

13.3.3 *Activity assay: ATPase assay*

The ATPase rate of a protein can be determined using a coupled enzymatic assay. The hydrolysis of ATP is coupled to the oxidation of NADH to NAD^+ in an equimolar ratio. That means per ATP molecule one NADH gets oxidized. The assay was performed according to Montgomery *et al.* [83]. For the assay reaction ATPase buffer was used and complemented with 330 µM ATP (Sigma-Aldrich), 1 µM phosphoenolpyruvate (Sigma-Aldrich), 0.25 µM NADH (Sigma-Aldrich) and 2 µL of a 1:5 dilution of the enzyme mix (PK and LDH, Sigma-Aldrich). The reaction was started by adding 0.5 µM protein to the assay and the cuvette was placed immediately into the UV/VIS spectrometer. The reaction was recorded by monitoring the absorption at 340 nm each 10 s for 30 min at 25 °C. The rate of ADP production was calculated using the Lambert-Beer equation.

$$k_{\text{cat}} = \frac{\Delta c_{\text{ATP}}/\Delta t}{c_{\text{protein}}} = \frac{m_{\text{NADH}\rightarrow\text{NAD}^+}}{\epsilon_{\text{NADH}} \times d_{\text{cuvette}} \times c_{\text{protein}}} \tag{3}$$

with

$m_{\text{NADH}\rightarrow\text{NAD}^+}$: linear regression of the NADH oxidation over time (background corrected)

ϵ_{NADH}: extinction coefficient of NADH (6220 $\mathrm{M^{-1}\,cm^{-1}}$)
d_{cuvette}: path length of the cuvette (1 cm)

13.3.4 *Determination of off-rates using fluorescently labeled nucleotides*

Nucleotide off-rates were determined using fluorescently labeled ATPs (8-[(4-Amino)butyl]-amino-adenosine-5′-triphosphate (MABA)-ATP, Atto-488-ATP, Atto-532-ATP or Atto-647N-ATP each with a C6-linker). Fluorescently labeled ATP in solution has a lower fluorescent intensity compared to ATP bound to the nucleotide binding pocket. This fact can be used to determine the off-rate of the labeled ATP by supplanting with a high amount of unlabeled ATP. When the unlabeled ATP replaces labeled ATP, the fluorescent intensity drops down. This decay can be fitted with an exponential fit and the time constant corresponds to the off-rate of labeled ATP. For the assay 50TKM-buffer was used and complemented with 500 nM labeled ATP. Next, 0.5 or 10 µM protein was added and incubated until the fluorescent signal is constant. Then displacement of labeled ATP with a 200-molar excess of unlabeled ATP (100 µM) was induced. The kinetics was recorded for 2 min in 0.5 s increments and for each dye, the respective excitation and emission wavelength were set in the Jasco Fluorimeter time-course experiment mode.

13.3.5 *Determination of nucleotide dissociation constants*

For the determination of nucleotide dissociation constants again labeled nucleotide was used. When labeled nucleotide is bound to the protein the fluorescent intensity increases. We varied the protein concentration starting from 50 nM up to 50 µM in 50TKM buffer but with a constant ligand concentration of 0.5 µM of labeled ATP. Increasing the protein concentration leads to the availability of more ATP binding sites and, hence, the fluorescence intensity increases. The data was measured in the Jasco Fluorimeter system using the spectra measurement mode. The settings were; excitation wavelength (e.g. 490 nm for Atto488 ATP), emission rage (e.g. 500–700 nm), scanning speed of 200 nm $\mathrm{s^{-1}}$, excitation width of 2.5 nm and three accumulations of each concentration. The data was analyze by plotting the maximal fluorescent intensity over the corresponding protein concentration. This results in a typical binding curve, which can be fitted using the following equation to determine the dissociation constant (K_{D}).

$$F = F_{\mathrm{max}} \times \left(\frac{c_{\mathrm{protein}}}{K_{\mathrm{D}} + c_{\mathrm{protein}}} \right) \tag{4}$$

with
F and F_{max}: Fluorescent intensity and maximal fluorescent intensity

K_D: dissociation constant of the nucleotide

13.4 STRUCTURE DETERMINATION MINI-DOMAINS

13.4.1 *SAXS measurements*

For the small-angle x-ray scattering (SAXS) measurements, TKM50 buffer with 1 mMTCEP was used, and the final protein concentration was 10 mg mL^{-1}. To investigate the effect of nucleotides on the structure of the protein we recorded SAXS measurements in the absence and presence of 1 mM ATP, ADP or AMP. The SAXS experiments were done in collaboration with Dr. Tobias Madl (TUM, chemistry department).

13.4.2 *Crystal structure*

For crystallization experiments and analysis we collaborated with Dr. Roman Peter Jakob from the Biozentrum University of Basel from the chair of Prof. Dr. Timm Maier. All crystallization experiments were carried out with 10–15 mg mL^{-1} at room temperature in sitting-drop vapor diffusion experiments in a 1:1 ratio of protein and precipitant. Mini-NBD/AMP-PcP at a concentration of 10 mg mL^{-1} was crystallized at 4 °C with 25% PEG1500, 0.1 M Bis-Tris buffer pH 7.0. Crystals were cryo-preserved by increasing the PEG1500 concentration to 35% (v/v) and flash cooled in liquid nitrogen. Data were collected at the SLS beamline X06SA (Swiss Light Source, Paul Scherrer Institute, Switzerland) at 100 K and integrated, indexed and scaled using the XDS software [52]. The Structure was solved by molecular replacement using the nucleotide binding domain of DnaK (PDB-ID:4JN4 [92]) as search model with the program Phaser [78]. Model building and structure refinement were performed with Coot [32] and PHENIX [115]. Data collection and refinement statistics are summarized in Table 9. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under the accession code 5OOW.

	Mini-NBD-183-359
PDB ID	5OOW
Wavelength (Å)	1.00003
Resolution range (Å)	54.7 - 2.9 (3.0 - 2.9)
Space group	P 2_1
Unit cell	59.86 64.20 74.64
$\alpha, \beta, \gamma(\hat{}\circ)$	90 113.3 90
Unique reflections	1151729 (1846132)
Multiplicity	6.9 (7.0)
Completeness (%)	98.7 (99.4)
Mean I/sigma(I)	7712.4 (2.01.3)
Wilson B-factor	50.5
R-merge	0.255 (0.966)
CC1/2	0.97958 (0.730582)
R-work	0.239 (0.309)
R-free	0.24969 (0.391332)
Number of atoms	3103
macromolecules	2832
ligands	124
water	147
Protein residues	365
RMS (bonds)	0.006
RMS (angles)	1.09
Ramachandran favored (%)	97
Ramachandran outliers (%)	0
Clashscore	0.68
Average B-factor	57.0

Table 9: Data collection and refinement statistics of X-ray structure.

13.5 OPTICAL TWEEZERS APPLICATION IN PROTEIN FOLDING STUDIES

In the following sections, the applied methods and assays regarding single-molecule optical tweezers will be presented. First, the assay ingredients and their assembly, second, measurement modes and finally the analysis tools for optical tweezers data will be explained.

13.5.1 *Establishment of the dumbbell configuration*

The method of optical tweezers has already been successfully adapted in protein folding studies. Using a dual beam optical tweezers setup, the so-called dumbbell geometry can be established to exert force on a single-molecule (Figure 53). In this geometry, the protein of interest can be fused to DNA and be stretched by moving one trap apart from the other. The introduction of two cysteine residues allows to freely choose the point of force application [25]. The only limitation is the accessibility of the cysteine for the reaction with the oligonucleotide-maleimide. The cysteine reacts with functionalized oligonucleotides, which then hybridize with the DNA-handles. These DNA-handles carry either biotin or digoxigenin functionalization at the 5′-end. These 5′-ends can bind to streptavidin functionalized or anti-digoxigenin functionalized silica beads. The beads are 1 µm in diameter and have a refractive index of 1.47 in the visible range.

To establish a dumbbell, the two beads surfaces need to contact each other to allow binding of the DNA-protein-DNA construct. The binding to the bead is achieved by their surface functionalization (streptavidin or anti-digoxigenin).

In the following subsections I will explain, how the different reagents are prepared how the single reaction steps are carried out.

[25] *Its important to note when using this attachment approach that no other cysteines in the protein are accessible and interfere with the reaction. If native cysteines are present and cannot be mutated, other attachment approaches like ybbr-tag or unnatural amino acids are recommended.*

13.5.1.1 *DNA-handles preparation*

DNA handles serve as a spacer between bead surface and protein. They are 545 base pairs long fragments of the λ-phage DNA and can easily be prepared by PCR (the sequence and primers are listed in Section A.1). The handles carry either at the 3′ end triple biotin or digoxigenin functionalization Figure 54A. The 5′ end of the complementary strand carries an abasic side with a single-stranded overhang. This overhang is complementary to the maleimide-oligonucleotide fused to the protein (Section 13.5.1.3 and Figure 54B). For the DNA handle PCR the following protocol was used:

After the PCR programme was finished, the PCR product was purified using the PCR clean up kit from Qiagen. The amount of DNA handles after purification was determined spectroscopically.

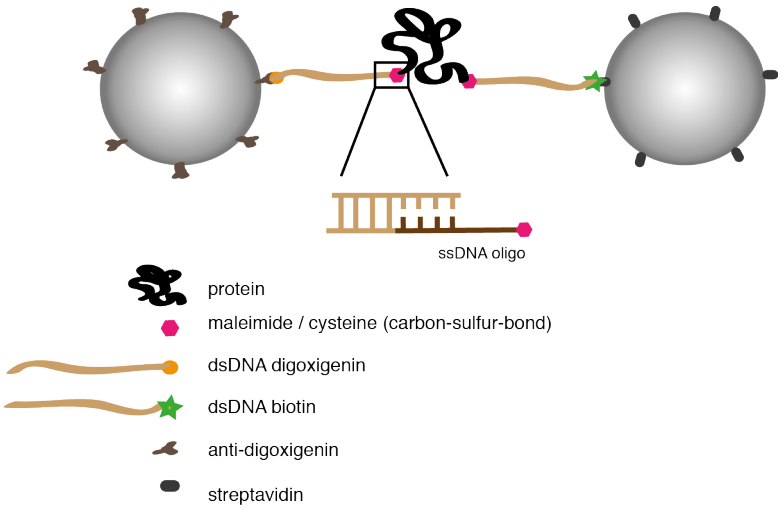


Figure 53: Optical tweezers assay: Dumbbell formation. In the assay two differently labeled silica beads (1 µm in diameter, anti-digoxigenin (brown) and streptavidin (black) labeling). The protein of interest carries two exposed cysteine residues at the force application sites. These cysteines can react with a ssDNA-maleimide compound and form a covalent thiol bond (pink). The ssDNA overhang can then hybridize with the dsDNA handles. These ds-DNA handles carry a 3′ ssDNA overhang complementary to the one fused to the protein. At the 5′-end either biotin (green) or digoxigenin (orange) label is introduced. These labels can bind to the corresponding beads (biotin-streptavidin and digoxigenin-anti-digoxigenin complexes). Thus, the protein is fused between two beads and force can be applied by moving one bead.

13.5.1.2 *Bead functionalization*

For the optical tweezers measurements silica beads (Bangs Laboratories No. SC03N) with 1 µm in diameter were used. The functionalization is based on the preparation of amino-reactive esters of carboxylate groups. The carboxy group of the bead is in this reaction activated with the carbodiimide EDC (Thermo Fisher Scientific) and Sulfo-NHS (Thermo Fisher Scientific). The resulting semi-stable amine-reactive ester forms a stable amide bond with the amino-terminus of the protein of interest (streptavidin or anti-digoxigenin). In our case we produced streptavidin coated beads and anti-digoxigenin coated beads using the described type of chemistry. The anti-digoxigenin beads are further labeled with TMR-BSA to be able to distinguish the two different bead sorts in our the optical tweezers setup.

Substance	**Volume** [µL]
ddH_2O	232
Thermopol buffer	30
dNTP	6
λ-phage DNA 500 µg mL^{-1}	1
primer 100 nM (forward, triple biotin)	5
primer 100 nM (forward, triple digoxigenine)	5
primer 100 nM (reverse, abasic side, ss-overhang)	10
Taq polymerase 25 U mL^{-1}	6

Step	**Temperature** [°C]	**Time[sec]**
1	94	300
2	94	15
3	60	15
4	68	40
5	cycle from to 2	37 times
6	68	300
7	4	inf

Table 10: DNA-handles preparation (PCR reaction).

13.5.1.3 *Oligonucleotide attachment using maleimide-cysteine functionalization*

For hybridization of maleimide-oligonucleotides to the protein, the artificially introduced cysteines were reduced with 2 mM Tris-(2-carboxyethyl)-phosphin (TCEP) for 1 h on ice. After buffer exchange, using SEC (YMC diol-120 column, YMC Europe, GmbH), the protein was incubated with an equimolar amount of maleimide-oligonucleotides for 2 h on ice. Protein-oligonucleotide hybrids were then purified by SEC (YMC diol-120 column, YMC Europe, GmbH). The SEC runs were performed at a JASCO HPLC system.

13.5.1.4 *Oxygen Scavenging system to prevent oxygen damage*

Biomolecules are very sensitive to reactive oxygen species. To enable long-lasting experiments, it is necessary to minimize these species. Therefore, the so-called GLOXY system was used. This system consists of 0.65% glucose (Sigma, Germany), 13 U mL^{-1} glucose oxidase (Sigma, Germany) and 8500 U mL^{-1} catalase (Calbiochem, Germany).

Figure 54: DNA handles design. (A) At the 5′ end of each handles are either three biotin or digoxigenine functionalizations (green boxes) included into the primer sequence (pink). At the 3′ end an abasic side is included (red X) and results in a single stranded DNA overhang (orange). (B) Example of one DNA handle. (C) Illustration of the whole dumbbell using the maleimide functionalized protein of interest with the complementary single-stranded overhang to the 3′ end of the DNA handle (blue and green).

Alternatively, if better pH control is needed, the POC (protocatechuic acid/protocatechuate-3,4-dioxygenase) system can be used instead of the glucose oxidase [1].

13.5.1.5 *Preparation of measurement chamber*

For the final measurement a two step incubation procedure was used. First, 10 µL of attached protein was incubated with 100 ng handles for 30 min on ice. Next, the protein-handle construct was incubated with a 1:10 dilution of the in-house functionalized anti-digoxigenine TMR-BSA labeled beads. After 5 min incubation time, ∼1.5 µL were added to the final mixture (total volume of 50 µL) including: 1 µL 1:3000 diluted streptavidine silica beads and the afore mentioned oxygen scavenger system. All experiments were performed in 50 mM TRIS pH 7.5, 50 mM KCl, 5 mM $MgCl_2$ (50TKM) buffer and if necessary 10–1000 µM ATP, ADP or AMP was added.

13.5.2 *Optical tweezers measurement modes*

In our setup, we bring together and separate the two beads multiple times with a constant velocity, until a dumbbell has formed (Figure 55A, left). The formation of a dumbbell is characterized by the signal of DNA stretching[26]. This signal is illustrated in Figure 55A on the right side (force versus extension plot). The DNA is stretched, and at a certain force, which is stochastically distributed, a sudden drop in force occurs. This drop corresponds to the unfolding or dissociation of the protein or a protein domain. In the example, in Figure 55, the protein unfolds in a single step. As the distance between the traps is decreased again, the force continuously decreases, and the protein can refold. After reaching zero force, a new cycle can begin by increasing the trap distances again. This pattern drives the protein constantly from the natively folded state to the unfolded state and back. A big advantage of this method is that it can be applied multiple times. This experimental assay is called constant velocity measurement mode. Another measurement mode that is used in optical tweezers experiments is passive mode (Figure 55B). Here, the distance of the mobile trap is set to a specific value and stays constant for a user controlled amount of time. In this mode, the fluctuations of the protein between different states and their energetics as well as their kinetics can be investigated. For the successful realization of this measurement, the protein must be able to populate states of interest at the selected trap distance on a reasonable timescale. Finally, the population of all states at different distances/forces can be plotted as force-lifetime distributions. The population of a state at a particular force depends on a certain force-dependent probability. With this information, the kinetic network and the energy landscape of the folding and unfolding process can be characterized.

[26] *One DNA handle consists of 545bp which corresponds to a contour length of* $0.338\,\mathrm{nm} \cdot 545 = 184.2\,\mathrm{nm}$. *In our assay, we have on each side of the protein one DNA-handle attached and therefore expect a contour length of about* 360 nm.

CONSTANT VELOCITY MEASUREMENTS Constant velocity measurements were performed by setting the pulling boundaries to 180 nm and 650 nm distance. This resulted in a force range from 0 pN to $\approx$ 35 pN. The waiting time at zero force was varied from 0 to 10 s and the pulling speed was set to 500, 200 or $20\,\mathrm{nm\,s^{-1}}$.

PASSIVE MODE EXPERIMENTS Passive mode experiments were used to investigate the force-dependent refolding of EcNBD and NBD variants as well as the binding of nucleotide to certain refolding intermediates. In this assay the force was constantly decreased in $\approx$ 0.5 pN steps, starting at 7 pN and going down to 2 pN. With decreasing force, the molecule can partially refold under load. For all force steps, force-dependent rates can be determined. The force dependent rates are calculated from the dwell time distribution of each state at a given force. The state assignment in the raw data is done using an HMM model.

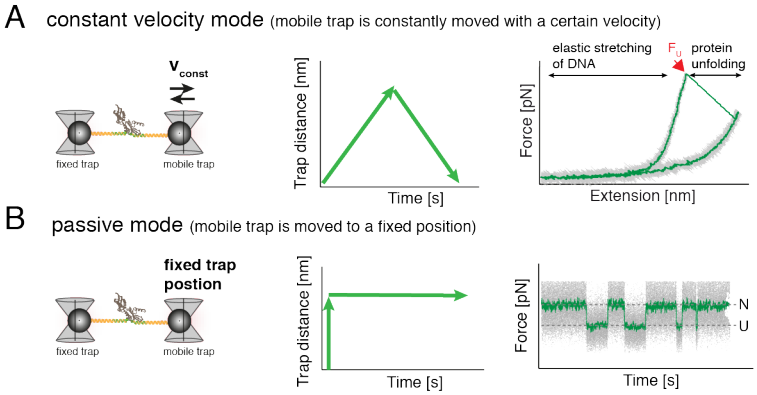


Figure 55: Optical tweezers assay: Measurement modes. (A) Constant velocity mode. Here, the mobile trap is moved away and back towards the other trap with a constant velocity (middle). Hence, a constant increase of the force is achieved, which is acting on the DNA-protein construct resulting in a force-extension curve (right). Here, the DNA gets stretched, then at a specific force the protein unfolds and results in a rupture (unfolding force: F_U) (red arrow). This rupture leads to a sudden decrease in force. In the end, the DNA and unfolded protein are fully stretched. By constantly decreasing the force the protein can refold again, and the next cycle can start. (B) Passive mode experiments. In this type of experiment, the trap distance is held constant (left, middle). In this case, the protein is constantly under load and can populate for example two states: unfolded (U) and native (N) (right).

Both procedures are described in more detail in the next section (Section 13.5.3). Plotting these force-dependent rates results in a so-called chevron plot. The force-independent rates can be extrapolated using the BS-model (Section 13.5.3.4) to fit the chevron plot. Analysis of optical tweezers data will in the following section described more intensively.

13.5.3 *Data analysis strategies of optical tweezers data*

13.5.3.1 *Elastic linker models*

Force-extension curves obtained in constant velocity mode were modeled with previously published polymer models. Before unfolding, the tether elasticity is governed by the stretching of the double-stranded deoxyribonucleic acid (dsDNA) handles, which were described by an extensible worm-like chain (eWLC) model:

$$F_{\mathrm{eWLC}}(x) = \frac{k_{\mathrm{B}}T}{p_{\mathrm{D}}}\left(\frac{1}{4\left(1-\frac{x}{L_{\mathrm{D}}}+\frac{F}{K}\right)^2}-\frac{1}{4}+\frac{x}{L_{\mathrm{D}}}-\frac{F}{K}\right) \tag{5}$$

with extension x, force F, thermal energy k_BT, DNA persistence length p_D, DNA contour length L_D and the elastic stretch modulus K. The force-extension behavior of an unfolded amino acid chain was described by a worm-like chain (WLC) model, in series with the ds-DNA polymer:

$$F_{WLC}(x) = \frac{k_BT}{p_p}\left(\frac{1}{4\left(1-\frac{x}{L_p}\right)^2} - \frac{1}{4} + \frac{x}{L_p}\right) \tag{6}$$

where L_p is the protein contour length and p_p the protein persistence length. Each pulling trace was fitted using the eWLC fit for the DNA stretching and a hybrid WLC-eWLC-series model to fit the individual unfolding events. The protein persistence length $p_p$was fixed at 0.7 nm and K was fixed to a value of 400 pN nm^{-1}. Typical values for the DNA fit were $p_D \approx 25$ nm and $L_D \sim 360$ nm. After a native unfolding event, the DNA parameters (K, p_D, L_D) were kept fixed for subsequent traces.

13.5.3.2 *Contour-length transformation*

Since the unfolding of the refolding intermediates occurs at low forces and the unfolding forces are stochastically distributed, the contour-length transformation is a valuable tool to compare length information between different pulling attempts and different molecules. The contour length of the unfolded polypeptide does not depend on additional experimental parameters, such as dsDNA linker elasticity or trap stiffness, and can, therefore, be used for unfolding pattern recognition. A detailed description can be found in Puchner *et al.* [25]. Briefly, to obtain trajectories in contour-length space, force extension traces (Figure 56A) are represented in the force-time domain (Figure 56B). These traces are then transformed into contour-length space using the polymer models described above, following. The result of the transformation is shown Figure 56C All points corresponding to a specific contour-length fall onto horizontal lines. In the case of EcNBD, the unfolding of the native protein can be analyzed easily. In contrast, the refolding of the protein resulted in partially folded intermediates. Hence, the DNA and an unfolded part of EcNBD got stretched. To account for this, the DNA parameters obtained for the native unfolding trace were used for the contour-length transformation analysis of the refolding intermediates (see first eWLC fit in Figure 56A).

13.5.3.3 *State detection using HMM*

Recording of passive mode experiments results in a force-dependent time trace, in which the protein can populate certain states. To iden-

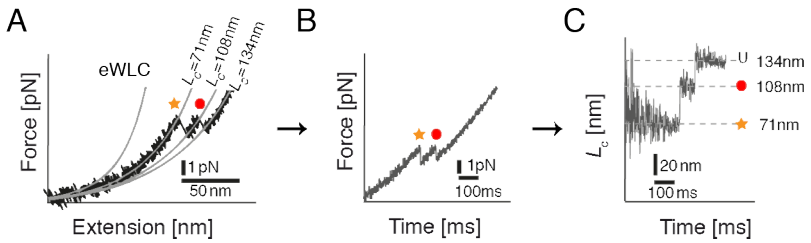


Figure 56: Contour-length transformation. (A) The force vs. extension trace is recorded with a pulling speed of 100 nm s^{-1} and 30 kHz acquisition speed. The data can be fitted using the worm-like chain model (WLC model). The first WLC-fit corresponds to the stretching of DNA. Each following fit corresponds to one unfolding event. The DNA fit parameters were determined from native EcNBD unfolding traces and are used in the analysis of the unfolding of the refolding intermediates as the DNA reference trace. Clearly, two unfolding events are visible and we obtained for the events 71 nm (star) and 108 nm (circle) unfolded contour length of the protein chain. (B) From force vs. extension traces, force-time traces can be calculated based on the known pulling velocity. That means, at each point of time we know the corresponding force because the trap-trap distance is increased linearly. (C) Force vs. time traces are further preceded to contour-length vs. time space. This contour-length transformed trace is now independent of any variations in experimental parameters and fluctuations. Which allows easy comparison of different molecules and measurements.

tify these states, a Hidden Markov model (HMM) based Algorithm, and the Viterbi Algorithm were used. In this model the probability p_i to populate a state x_i is memory-less and of continuous time and hence follows a Markovian process. At each point the transition from state x_i to x_j can happen with a certain transition probability $p_{i,j}$. The optimization of transition and emission probabilities happens via a so-called Forward-Backward Algorithm and the calculation of the likelihood. Both are procedures combined in the Baum-Welch Algorithm and results in the state probability estimation procedure. The final step is the assignment of states, which is done by the Viterbi Algorithm. Finally, each data point is assigned to one state and a maximal observable probability. From these force-dependent probabilities, force-dependent lifetimes and rates can be calculated. A detailed description of the whole analysis can be found in Stigler *et al.* [108].

13.5.3.4 *Shifting the energy landscape by force*

The simplest reaction model comprises two states with different free energies (2-state model). Both states are connected via an energy barrier (E_A) which has to be overcome to switch from state 1 to state 2. The transition state position[27] of each reaction is fixed. The position of the transition state is characterized by the distance to state 1 or state 2

[27] *peak of the energy barrier*

(Δx [nm]) (Figure 57A). A first description of this process can be done by using the Arrhenius equation

$$k_{\text{Arrhenius}} = Ae^{-\frac{E_A}{k_B T}} \tag{7}$$

with the transition frequency A. This equation shows also the reaction rate dependence on the temperature of the surrounding. A more elaborated model is the description by Kramer who extended the *transition state theory, TST* to give a more detailed description of A and E_A. The picture in this model is Langevin particle diffusion in a two well energy potential. The probability of finding a particle at a specific point x is, in equilibrium, time-independent and follows a Boltzmann process. Taken together, the transition rates of a reaction process highly depends on the shape of the energy potential. In force spectroscopy, this energy landscape of a protein, for example from the folded to the unfolded state, can be probed. The application of force to a protein acts as a denaturant, similarly to temperature or chemicals. Denaturant shifts the probability distribution of the populated states. This shift can be illustrated in the energy landscape as shown in Figure 57B. Increasing force lowers the energy barrier (ΔG) and thus shifts the probability of occupying state 1 or 2. At high forces, the occupancy of state 2 is favored, and at low forces, state 1. The *Bell-Evans model* can phenomenologically describe this tilting of an energy landscape. Originally, the Bell model described the process of force-dependent association and dissociation using the example of a receptor-ligand interaction. Evans *et al.* extended this theory by incorporating Kramers' Theory and other extensions, to describe the tilting of the energy landscape by an external force potential [34, 124]. As a result, the force-dependent rates change exponentially, and the unfolding rate (k_{unf}) can be written as:

$$k_{\text{unf}}(F) = k_0 e^{\frac{F\Delta x}{k_B T}} \tag{8}$$

where F is the force applied to the system, $k_B T$ is the Boltzmann constant, and T is the temperature in Kelvin.

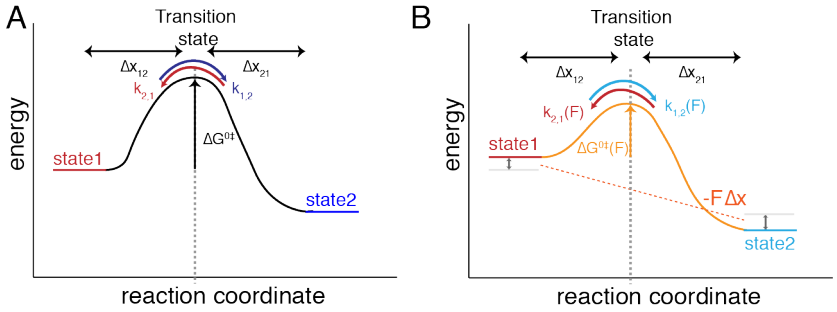


Figure 57: Energy landscape is influenced by force. Left: Undisturbed energy landscape of a 2-state system. In this case the energy barrier is so high that most molecules will populate state 1. Right: Force is applied on the system and the potential -FΔx shifts the whole energy landscape. Under this condition the probability to populate state 2 has increased.

FITTING OF FORCE-DEPENDENT RATES The folding and unfolding rates obtained in this thesis were described using a model that takes into account the energy needed to stretch the elastic linkers. The model assumes that, in the case of folding, the unfolded polypeptide of contour length L_i contracts up to a contour length $L_T = L_i - \Delta L$ at the transition state T from whence folding ensues [38, 100]. The force-dependence is given by

$$k(F) = k_0 e^{\left(\frac{-\Delta G_{iT}^{sys}(F, F_T)}{k_B T}\right)} \tag{9}$$

An equivalent model was used for unfolding.

FOLDING FREE ENERGIES The Gibbs free energy of the protein-dumbbell system when the protein is in state i at force (F_i) is given by

$$G_i(F_i) = G_i^0 + G_i^{sys}(F_i) = G_i^0 + G^{bead}(F_i) + G^{DNA}(F_i) + G_i^{prot}(F_i) \tag{10}$$

where G_i^0 is the free energy of the protein in state i and $G_i^{sys}(F_i) = G^{bead}(F_i) + G^{DNA}(F_i) + G_i^{prot}(F_i)$ is the mechanical energy stored in deflecting the beads from the trap centers (G^{bead}) as well as stretching the dsDNA linkers (G^{DNA}) and the unfolded polypeptide (G_i^{prot}). G^{DNA} and G_i^{prot} were calculated as integrals over eWLC and WLC models, respectively.

$$G^{DNA} = \int_{x_0}^{x_{eWLC}(F)} F_{eWLC}(x^D) dx^D \tag{11}$$

$$G^{\text{prot}} = \int_{x_0}^{x_{\text{WLC}}(F)} F_{\text{WLC}}(x^p)dx^p \tag{12}$$

G^{bead} was calculated using a Hookean spring model

$$G^{\text{bead}}(F) = \frac{1}{2}k_{eff}^{-1}F^2 \tag{13}$$

with $k_{eff}^{-1} = k_{\text{mobile, trap}}^{-1} + k_{\text{fixed, trap}}^{-1}$.

The determination of equilibrium free energies of a two state system in single molecule traces is given by the Boltzmann equation

$$\frac{P_jF_j}{P_iF_i} = e^{-\frac{\Delta G_{i,j}(F_i,F_j)}{k_BT}}. \tag{14}$$

The free energy difference $\Delta G_{ij}^0 = G_j^0 - G_i^0$ between states i and j can be determined by measuring the force-dependent state occupancies $P_i(F)$, and fitting them globally to an equilibrium model:

$$P_i(F_i) = \left(1 + \sum_{j\neq i} e^{\left(-\frac{\Delta G_{ij}^0 + \Delta G_{ij}^{\text{sys}}(F_i,F_j)}{k_BT}\right)}\right)^{-1} \tag{15}$$

13.5.3.5 *ATP-binding model*

TRAJECTORY AND DWELL TIME SIMULATIONS Trajectory simulations were based on an eight-state kinetic scheme involving RFI2, RFI2f, RFI1 and U in apo and ATP-bound forms (Figure 58A). Linker parameters were set to typical experimental values, and the trap distance was set to place RFI1 at the desired force level. For a selected trap distance, the folding/unfolding transition rates were then determined from force-dependent rate models based on experimental data. Binding/unbinding rates were set by pre-selecting the ATP on-rate , e.g. to $10^6\,\text{M}^{-1}\,\text{s}^{-1}$, and finding the corresponding off-rate ($k_{\text{off}} = K_D \times k_{\text{on}}$), where the ATP equilibrium constant was chosen to be 10 M for states RFI2f, RFI1 and U (i.e. effectively non-binding), and the experimentally determined value of 33 µM for RFI2. A state-sequence trajectory was then simulated based on this kinetic scheme Figure 58B. The noise was introduced by integrating the Langevin equation of the bead-DNA-protein-system, as described previously [72, 109]. Trajectories were boxcar-smoothed as experimental data and color-coded, based on the simulated state-sequence (Figure 58C and Figure 58D). The dwell times for each experimentally observable kinetic state (by merging binding/unbinding transitions) were then collected and displayed

as scatter plots (Figure 58E and Figure 58G) or as probability density functions (Figure 58F and Figure 58H).

SIMPLIFIED ATP BINDING MODEL In the limiting case of fast ATP exchange kinetics, ATP binding occurs in quasi-equilibrium and the unfolding kinetics of the transition RFI2→RFI1 can be described with a simplified model (Figure 58):

$$k_{\mathrm{RFI2ATP}\rightarrow\ \mathrm{RFI1}} = \frac{k_{\mathrm{RFI2apo}}}{\left(1+\frac{k_{\mathrm{on}}[ATP]}{k_{\mathrm{off}}}\right)} \tag{16}$$

SIGNAL-TO-NOISE CUTOFFS Short-lived transitions from one state to another will only be detected in our experiments when the excursion is large in magnitude (i.e., large force difference) or when the excursion lasts for a long enough time. To estimate the detectable events, we determined the standard deviations of dwells of length, where is the number of data points in a dwell and is the inverse sampling rate. Dwells will only be experimentally detected when they lie outside of the area described by

$$\mu \pm S/N \times \frac{\sigma}{\sqrt{\frac{\tau}{\Delta t}+1}} \tag{17}$$

where μ is the average force level for a state, σ is the standard deviation of the noise for a state and S/N is an empirical factor describing the signal-to-noise ratio. Figure 58E shows the cutoff area for excursions from state RFI1 for different numbers of S/N (red shaded). This signal-to-noise estimation gives a rationale why many dwells of state U are not detected experimentally at 4 pN, even though U has a lifetime that is similar to RFI2f.

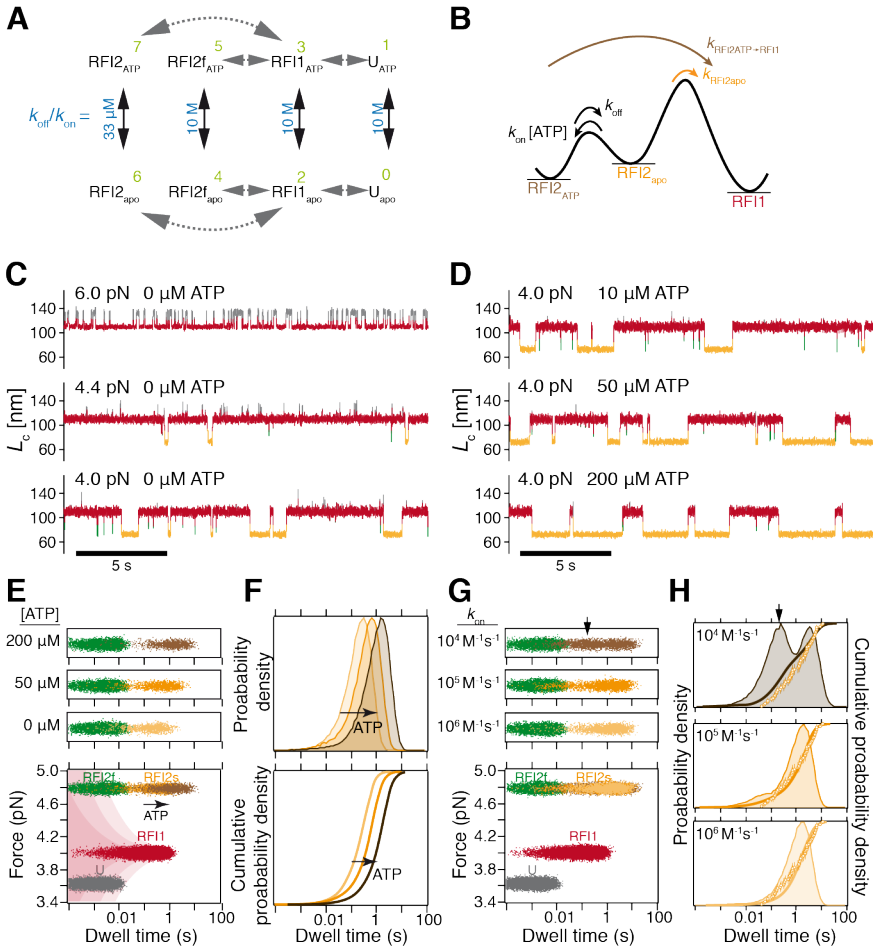


Figure 58: ATP-binding model. (A) Kinetic scheme of the binding reaction of ligand to the refolded portions of NBD. (B) Illustrated binding model, where ATP is only binding to RFI2. (C) Simulated passive mode trajectories at different forces (4, 4.4 and 6 pN). (D) Simulated passive mode trajectories in the presence of 10, 50 or 200 µM MgATP. (E) Dwell-time distribution or probability density distribution (F) of each state without ligand. (G) Simulation of the dwell-time distribution using different on-rates. (H) Probability density distribution of dwell-times (lines) with different on-rates. Data points are measured values. Slow exchange kinetics ($k_{on} < 10^{4.1}\,M^{-1}\,s^{-1}$ are not compatible with the measured data ($p < 0.05$, Kolmogorov-Smirnov test), while fast exchange kinetics are.

Part VI

APPENDIX

A

SUPPLEMENTARY INFORMATION

A.1 DNA HANDLES SEQUENCE

GGC GAT CTG GTC GTT GAT TTG AGT CTG GAT GCG GCC AGA TTT GAC GAG CAG ATG GCC AGA GTC AGG CGT CAT TTT TCT GGT ACG GAA AGT GAT GCG AAA AAA ACA GCG GCA GTC GTT GAA CAG TCG CTG AGC CGA CAG GCG CTG GCT GCA CAG AAA GCG GGG ATT TCC GTC GGG CAG TAT AAA GCC GCC ATG CGT ATG CTG CCT GCA CAG TTC ACC GAC GTG GCC ACG CAG CTT GCA GGC GGG CAA AGT CCG TGG CTG ATC CTG CTG CAA CAG GGG GGG CAG GTG AAG GAC TCC TTC GGC GGG ATG ATC CCC ATG TTC AGG GGG CTT GCC GGT GCG ATC ACC CTG CCG ATG GTG GGG GCC ACC TCG CTG GCG GTG GCG ACC GGT GCG CTG GCG TAT GCC TGG TAT CAG GGC AAC TCA ACC CTG TCC GAT TTC AAC AAA ACG CTG GTC CTT TCC GGC AAT CAG GCG GGA CTG ACG GCA GAT CGT ATG CTG GTC CTG TCC AGA GCC GGG CAG GGG CAG GGC TGA CGT TCA ACC AGA CCA GCG AGT CG

forward primer 3 bio
5′-bio-GGCGA T-bio CTGGT-bio CGTTGATTTG-3′
forward primer 3 dig
5′-dig-GGCGA T-dig CTGGT-dig CGTTGATTTG-3′
reverse primer abasic side
5′-CGA CTC GCT GGT CTG GTT GAA CGT CAG CCC TGC C**X** CCT GCC CGG CTC TGG ACA GG-3′

A.2 NUCLEOTIDE AND AMINO ACID SEQUENCE OF ALL USED CONSTRUCTS

A.2.1 *Protein sequences of NBD proteins from E. coli*

Pseudo wild-type NBD sequence (EcNBD) (1-393) for optical trapping (two cysteines, his tag, C15A (bold) and spacers (underlined)
Nucleotide sequence

CAT ATG GCT TGT GGT ACA GGT GGT GGA GAA CTG GGC AAA ATC ATC GGA ATC GAT CTG GGT ACA ACC AAT AGC GCT GTG GCG ATC ATG GAT GGC ACT ACT CCT CGT GTG CTG GAA AAT GCC GAA GGT GAC CGT ACT ACG CCG AGC ATC ATT GCC

TAT ACA CAG GAC GGT GAA ACC CTG GTA GGA CAG CCT GCT AAA CGT CAG GCA GTA ACG AAT CCT CAG AAC ACA CTG TTT GCT ATC AAA CGT CTG ATT GGC CGT CGT TTT CAA GAT GAA GAA GTT CAG CGT GAT GTG AGT ATC ATG CCG TTC AAA ATC ATT GCC GCC GAT AAT GGA GAT GCT TGG GTG GAA GTG AAA GGC CAG AAA ATG GCA CCT CCT CAA ATT TCA GCC GAA GTC CTG AAA AAA ATG AAA AAA ACC GCC GAG GAC TAT CTG GGA GAA CCG GTT ACA GAA GCC GTT ATT ACC GTT CCG GCC TAT TTC AAC GAT GCC CAA CGT CAA GCG ACC AAA GAC GCC GGT CGT ATT GCT GGT CTG GAA GTC AAA CGT ATC ATC AAC GAA CCG ACA GCA GCA GCC CTG GCT TAT GGA CTG GAT AAA GGC ACC GGG AAT CGT ACA ATT GCC GTG TAT GAC CTG GGT GGC GGA ACT TTC GAT ATC AGC ATT ATT GAG ATT GAC GAA GTG GAC GGT GAG AAA ACC TTT GAA GTG CTG GCC ACC AAT GGC GAT ACA CAT CTG GGC GGT GAA GAT TTC GAC TCT CGC CTG ATC AAT TAT CTG GTG GAA GAG TTC AAA AAA GAC CAA GGC ATT GAC CTG CGT AAT GAT CCA CTG GCC ATG CAA CGC CTG AAA GAA GCA GCG GAA AAA GCG AAA ATC GAG CTG AGT TCT GCC CAA CAG ACC GAT GTT AAT CTG CCG TAT ATC ACG GCC GAT GCT ACT GGT CCT AAA CAC ATG AAC ATC AAA GTG ACC CGT GCC AAA CTG GAA TCA CTG GTG GAG GAT CTG GTT AAT CGC TCC ATC GAA CCT CTG AAA GTG GCT CTG CAG GAT GCC GGA CTG TCA GTT TCG GAT ATT GAC GAC GTT ATC CTG GTG GGT GGT CAG ACT CGT ATG CCA ATG GTC CAA AAA AAA GTC GCC GAA TTT TTC GGT AAA GAG CCT CGT AAA GAC GTC AAC CCG GAT GAA GCT GTA GCA ATT GGT GCC GCC GTT CAA GGT GGT GTA CTG ACC GGA GAT GTT AAA GAT GTC CTG CTG CTG GAT CGT GGT GGA AAA TGT CTG GAG CAC CAT CAT CAC CAT CAC TGA GGA TCC

Amino acid sequence

MA**C**G TGGG ELGK IIGI DLGT TNS**A** VAIM DGTT PRVL ENAE GDRT TPSI IAYT QDGE TLVG QPAK RQAV TNPQ NTLF AIKR LIGR RFQD EEVQ RDVS IMPF KIIA ADNG DAWV EVKG QKMA PPQI SAEV LKKM KKTA EDYL GEPV TEAV ITVP AYFN DAQR QATK DAGR IAGL EVKR IINE PTAA ALAY GLDK GTGN RTIA VYDL GGGT FDIS IIEI DEVD GEKT FEVL ATNG DTHL GGED FDSR LINY LVEE FKKD QGID LRND PLAM QRLK EAAE KAKI ELSS AQQT DVNL PYIT ADAT GPKH MNIK VTRA KLES LVED LVNR SIEP LKVA LQDA GLSV SDID DVIL VGGQ TRMP MVQK KVAE FFGK EPRK DVNP DEAV AIGA AVQG GVLT GDVK DVLL LDRG GKCL E**HHH HHH**

A.2.2 *EcNBD Ins183*

Pseudo wild-type NBD sequence with additional 20 amino acid long linker (bold)

MACG TGGG ELGK IIGI DLGT TNSA VAIM DGTT PRVL ENAE GDRT TPSI IAYT QDGE TLVG QPAK RQAV TNPQ NTLF AIKR LIGR RFQD EEVQ RDVS IMPF KIIA ADNG DAWV EVKG QKMA PPQI SAEV LKKM KKTA EDYL GEPV TEAV ITVP AYFN DAQR QATK DAGR IAGL EVKR IINE PTAA ALAY GLDK **GGSS GEGS SGEG SSGE SSGG** GTGN RTIA VYDL GGGT FDIS IIEI DEVD GEKT FEVL ATNG DTHL GGED FDSR LINY LVEE FKKD QGID LRND PLAM QRLK EAAE KAKI ELSS AQQT DVNL PYIT ADAT GPKH MNIK VTRA KLES LVED LVNR SIEP LKVA LQDA GLSV SDID DVIL VGGQ TRMP MVQK KVAE FFGK EPRK DVNP DEAV AIGA AVQG GVLT GDVK DLLL DRGG KCLE HHHH HH

A.2.3 *EcNBD Ins290*

Pseudo wild-type NBD sequence with additional 20 amino acid long linker (bold)

MACG TGGG ELGK IIGI DLGT TNSA VAIM DGTT PRVL ENAE GDRT TPSI IAYT QDGE TLVG QPAK RQAV TNPQ NTLF AIKR LIGR RFQD EEVQ RDVS IMPF KIIA ADNG DAWV EVKG QKMA PPQI SAEV LKKM KKTA EDYL GEPV TEAV ITVP AYFN DAQR QATK DAGR IAGL EVKR IINE PTAA ALAY GLDK GTGN RTIA VYDL GGGT FDIS IIEI DEVD GEKT FEVL ATNG DTHL GGED FDSR LINY LVEE FKKD QGID LRND PLAM QRLK EAAE KAKI ELSS AQQT DVNL PYIT ADA**G GSSG EGSS GEGS SGES SGG**T GPKH MNIK VTRA KLES LVED LVNR SIEP LKVA LQDA GLSV SDID DVIL VGGQ TRMP MVQK KVAE FFGK EPRK DVNP DEAV AIGA AVQG GVLT GDVK DVLL LDRG GKCL EHHH HHH

A.2.4 *EcNBD Ins364*

Pseudo wild-type NBD sequence with additional 20 amino acid long linker (bold)

MACG TGGG ELGK IIGI DLGT TNSA VAIM DGTT PRVL ENAE GDRT TPSI IAYT QDGE TLVG QPAK RQAV TNPQ NTLF AIKR LIGR RFQD EEVQ RDVS IMPF KIIA ADNG DAWV EVKG QKMA PPQI SAEV LKKM KKTA EDYL GEPV TEAV ITVP AYFN DAQR QATK DAGR IAGL EVKR IINE PTAA ALAY GLDK GTGN RTIA VYDL GGGT FDIS IIEI DEVD GEKT FEVL ATNG DTHL GGED FDSR LINY LVEE FKKD QGID LRND PLAM QRLK EAAE KAKI ELSS AQQT

DVNL PYIT ADAT GPKH MNIK VTRA KLES LVED LVNR SIEP LKVA LQDA GLSV SDID DVIL VGGQ TRMP MVQK KVAE FFGK EPRK D**GGS SGEG SGEG SSGE SSGG** VNPD EAVA IGAA VQGG VLTG DVKD VLLL DRGG KCLE HHHH HH

A.3 PROTEIN SEQUENCE OF MTNBD AND CHIMERIC PROTEINS

A.3.1 *Sequence of mtNBD*

MACK GSSG GQST KVQG SVIG IDLG TTNS AVAI MEGK VPKI IENA EGSR TTPS VVAF TKEG ERLV GIPA KRQA VVNP ENTL FATK RLIG RRFE DAEV QRDI KQVP YKIV KHSN GDAW VEAR GQTY SPAQ IGGF VLNK MKET AEAY LGKP VKNA VVTV PAYF NDSQ RQAT KDAG QIVG LNVL RVVN EPTA AALA YGLE KSDS KVVA VFDL GGGT FDIS ILDI DNGV FEVK STNG DTHL GGED FDIY LLRE IVSR FKTE TGID LEND RMAI QRIR EAAE KAKI ELSS TVST EINL PFIT ADAS GPKH INMK FSRA QFET LTAP LVKR TVDP VKKA LKDA GLST SDIS EVLL VGGM SRMP KVVE TVKS LFGK DPSK AVNP DEAV AIGA AVQG AVLS GEVT DVLL LSSG AKCL

A.3.2 *Chimera 50%mtNBD*

Lobe II EcNBD (bold) and Lobe I mtNBD

MACK GSSG GQST KVQG SVIG IDLG TTNS AVAI MEGK VPKI IENA EGSR TTPS VVAF TKEG ERLV GIPA KRQA VVNP ENTL FATK RLIG RRFE DAEV QRDI KQVP YKIV KHSN GDAW VEAR GQTY SPAQ IGGF VLNK MKET AEAY LGKP VKNA VVTV PAYF NDSQ RQAT KDAG QIVG LNVL RVVN EPTA AALA YGLE KS**GN RTIA VYDL GGGT FDIS IIEI DEVD GEKT FEVL ATNG DTHL GGED FDSR LINY LVEE FKKD QGID LRND PLAM QRLK EAAE KAKI ELSS AQQT DVNL PYIT ADAT GPKH MNIK VTRA KLES LVED LVNR SIEP LKVA LQDA GLSV SDID DVIL VGGQ TRMP MVQK KVAE FFGK EPRK** AVNP DEAV AIGA AVQG AVLS GEVT DVLL LSSG AKCL EHHH HHH

A.3.3 *Chimera 75%mtNBD*

Lobe IIb EcNBD (bold), Lobe IIa and I mtNBD

MACK GSSG GQST KVQG SVIG IDLG TTNS AVAI MEGK VPKI IENA EGSR TTPS VVAF TKEG ERLV GIPA KRQA VVNP ENTL FATK RLIG RRFE DAEV QRDI KQVP YKIV KHSN GDAW VEAR GQTY SPAQ IGGF VLNK MKET AEAY LGKP VKNA VVTV PAYF NDSQ RQAT KDAG QIVG LNVL RVVN EPTA AALA YGLE KSDS KVVA VFDL GGGT FDIS ILDI DNGV FEVK STNG DTHL GGED F**DSR LINY LVEE FKKD QGID LRND PLAM QRLK EAAE KAKI ELSS AQQT DVNL PYIT ADAT GPKH MNIK VTRA KLES LVED LVNR SIEP** VKKA LKDA GLST SDIS EVLL VGGM SRMP KVVE TVKS LFGK DPSK AVNP DEAV AIGA AVQG AVLS GEVT DVLL LSSG AKCL EHHH HHH

A.3.4 *mtNBDΔ^{236}NGVFEVKS^{243} NBD*

mtNBD^{236}NGVFEVKS^{243} replaced with the EcNBD sequence ^{209}*EVDGEKTFEVLA*^{220} (bold)

MACK GSSG GQST KVQG SVIG IDLG TTNS AVAI MEGK VPKI IENA EGSR TTPS VVAF TKEG ERLV GIPA KRQA VVNP ENTL FATK RLIG RRFE DAEV QRDI KQVP YKIV KHSN GDAW VEAR GQTY SPAQ IGGF VLNK MKET AEAY LGKP VKNA VVTV PAYF NDSQ RQAT KDAG QIVG LNVL RVVN EPTA AALA YGLE KSDS KVVA VFDL GGGT FDIS ILDI D**EVD GEKT FEVL A**TNG DTHL GGED FDIY LLRE IVSR FKTE TGID LEND RMAI QRIR EAAE KAKI ELSS TVST EINL PFIT ADAS GPKH INMK FSRA QFET LTAP LVKR TVDP VKKA LKDA GLST SDIS EVLL VGGM SRMP KVVE TVKS LFGK DPSK AVNP DEAV AIGA AVQG AVLS GEVT DVLL LSSG AKCL EHHH HHH

A.3.5 *75%mtNBDΔ^{236}NGVFEVKS^{243} NBD*

mtNBD (lobe IIa, lobe I) ^{236}NGVFEVKS^{243} replaced with the EcNBD sequence ^{209}*EVDGEKTFEVLA*^{220} (bold) and lobe IIb (underlined)

MACK GSSG GQST KVQG SVIG IDLG TTNS AVAI MEGK VPKI IENA EGSR TTPS VVAF TKEG ERLV GIPA KRQA VVNP ENTL FATK RLIG RRFE DAEV QRDI KQVP YKIV KHSN GDAW VEAR GQTY SPAQ IGGF VLNK MKET AEAY LGKP VKNA VVTV PAYF NDSQ RQAT KDAG QIVG LNVL RVVN EPTA AALA YGLE KSDS KVVA VFDL GGGT FDIS ILDI D**EVD GEKT FEVL A**TNG DTHL GGED FDSR LINY LVEE FKKD QGID LRND PLAM QRLK EAAE KAKI ELSS AQQT DVNL PYIT ADAT GPKH MNIK VTRA KLES LVED LVNR SIEP VKKA LKDA GLST SDIS EVLL VGGM SRMP KVVE TVKS LFGK DPSK AVNP DEAV AIGA AVQG AVLS GEVT DVLL LSSG AKCL EHHH HHH

A.4 MINIMAL ATP-BINDING DOMAIN (MINI-NBD)

A.4.1 *Sumo-Mini-NBD-183-359*

Sequence with c-terminal His6-tag (bold), sumo sequence (italic) followed by EcNBD sequence (Mini-NBD-183-359)

MGSS **HHHH HH***GS GLVP RGSA SMSD SEVN QEAK PEVK PEVK PETH INLK VSDG SSEI FFKI KKTT PLRR LMEA FAKR QGKE MDSL RFLY DGIR IQAD QTPE DLDM EDND IIEA HREQ IGG*K GTGN RTIA VYDL GGGT FDIS IIEI DEVD GEKT FEVL ATNG DTHL GGED FDRL INYL VEEF KKDQ GIDL RNDP LAMQ RLKE AAEK AKIE LSSA QQTD VNLP YITA DATG PKHM NIKV TRAK LESL VEDL VNRS IEPL KVAL QDAG LSVS DIDD VILV GGQT RMPM VQKK VAEF FGK

A.4.2 *Sumo-Mini-NBD-183-383*

Sequence with C-terminal His_6-tag (bold), sumo sequence (italic) followed by EcNBD sequence (Mini-NBD-183-383)

MGSS **HHHH HH***GS GLVP RGSA SMSD SEVN QEAK PEVK PEVK PETH INLK VSDG SSEI FFKI KKTT PLRR LMEA FAKR QGKE MDSL RFLY DGIR IQAD QTPE DLDM EDND IIEA HREQ IGG*K GTGN RTIA VYDL GGGT FDIS IIEI DEVD GEKT FEVL ATNG DTHL GGED FDRL INYL VEEF KKDQ GIDL RNDP LAMQ RLKE AAEK AKIE LSSA QQTD VNLP YITA DATG PKHM NIKV TRAK LESL VEDL VNRS IEPL KVAL QDAG LSVS DIDD VILV GGQT RMPM VQKK VAEF FGKE PRKD VNPD EAVA IGAA VQGG VLT

A.5 ECNBD-Δα

Sequence of EcNBD-Δα, with C-terminal His_6-tag

MGKI IGID LGTT NSCV AIMD GTTP RVLE NAEG DRTT PSII AYTQ DGETL VGQP AKRQ AVTN PQNT LFAI KRLI GRRF QDEE VQRD VSIM PFKI IAAD NGDA WVEV KGQK MAPP QISA EVLK KMKK TAED YLGE PVTE AVIT VPAY FNDA QRQA TKDA GRIA GLEV KRII NEPT AAAL AYGL DKGT GNRT IAVY DLGG GTFD ISII EIDE VDGE KTFE VLAT NGDT HLGG EDFD SRLI NYLV EEFK KDQG IDLR NDPL AMQR LKEA AEKA KIEL SSAQ QTDV NLPY ITAD ATGP KHMN IKVT RAKL ESLV EDLV NRSI EPLK VALQ DAGL SVSD IDDV ILVG GQTR MPMV QKKV AEFF GKEP RCLE HHHH HH

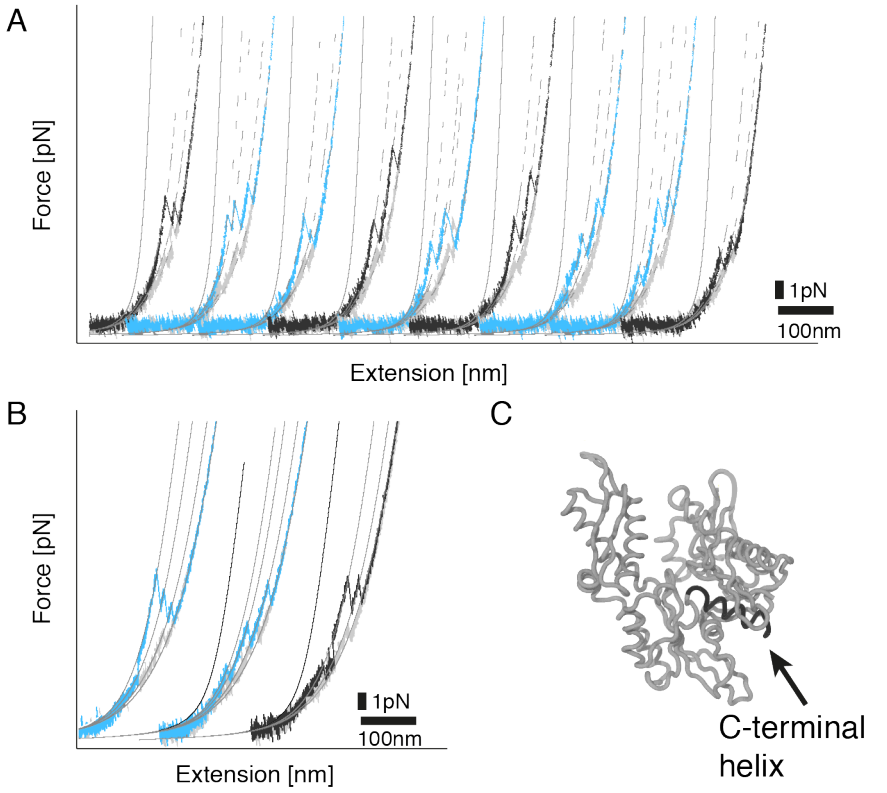


Figure 59: EcNBD-Δα constant velocity measurements apo and holo. Settings: Pulling speed of 200 nm s^{-1}, waiting time 2 s. (A) apo (B) 1 mM Mg-ATP (C) Structure of NBD (PDB:2KHO) in black the C-terminal helix is shown.

A.6 BIOCHEMICAL AND BIOPHYSICAL CHARACTERIZATION OF ECNBD AND MTNBD VARIANTS

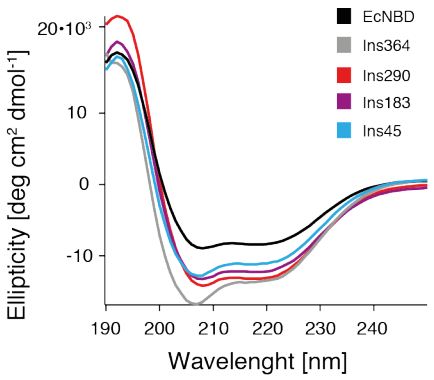


Figure 60: CD spectra of insert variants.

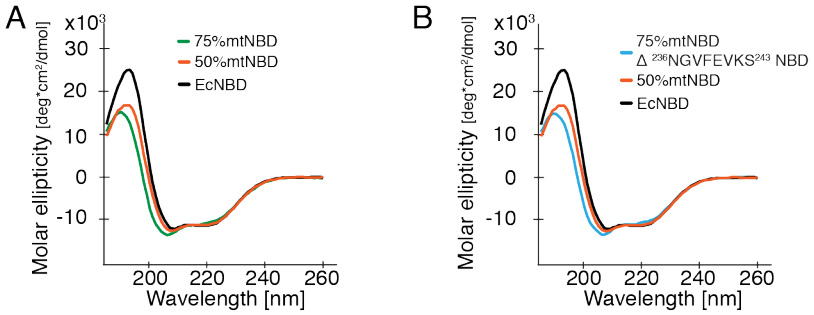


Figure 61: CD spectra of chimeric proteins.

Construct	$k_{cat}[s^{-1}]$	$k_{off}[s^{-1}]$
EcNBD-nc	$1.5 \pm 0.2\times10^{-3}$	$12 \pm 1.0\times10^{-3}$
literature	$1.5 \pm 0.7\times10^{-3}$	$6.7 \pm 1.2\times10^{-3}$
Ins364	$0.8 \pm 0.1\times10^{-3}$	$6.7 \pm 1.2\times10^{-3}$
Ins290	$1.1 \pm 0.1\times10^{-3}$	$11.0 \pm 0.8\times10^{-3}$
Ins183	$1.5 \pm 0.1\times10^{-3}$	$7.9 \pm 1.0\times10^{-3}$
Ins45	$0.6 \pm 0.2\times10^{-3}$	$10.0 \pm 1.1\times10^{-3}$
EcNBD-lobeII	$1.2\pm 0.1\times10^{-3}$	$8.2\pm 1.0\times10^{-3}$
50%mtNBD	$1.7\pm 0.9\times10^{-3}$	none
75%mtNBD	none	none
75%mtNBDΔ^{236}NGVFEVKS^{243}	none	none
mtNBDΔ^{236}NGVFEVKS^{243}	none	none

Table 11: Table of catalytic-rates and off-rates of all variants used. The literature value is taken from [83, 112, 117].

protein-ligand	R_G[nm]	Molecular weight [kDa]
EcNBD apo	2.47	46
EcNBD ATP	2.39	42
EcNBD ADP	2.37	41
EcNBD AMP	2.36	41
Mini-NBD-183-359apo	3.35	43
Mini-NBD-183-359ATP	2.9	37
Mini-NBD-183-359ADP	2.9	39
Mini-NBD-183-359AMP	2.93	40
Mini-NBD-183-383apo	2.32	25
Mini-NBD-183-383ATP	2.23	25
Mini-NBD-183-383ADP	2.32	19
Mini-NBD-183-383AMP	2.34	20

Table 12: Liste of radius of gyration and molecular weight obtained by SAXS. Listed are radius of gyration and molecular weight of apo and ligand bound EcNBD, Mini-NBD-183-359and Mini-NBD-183-383protein in 50 mMTRIS pH 7.5.

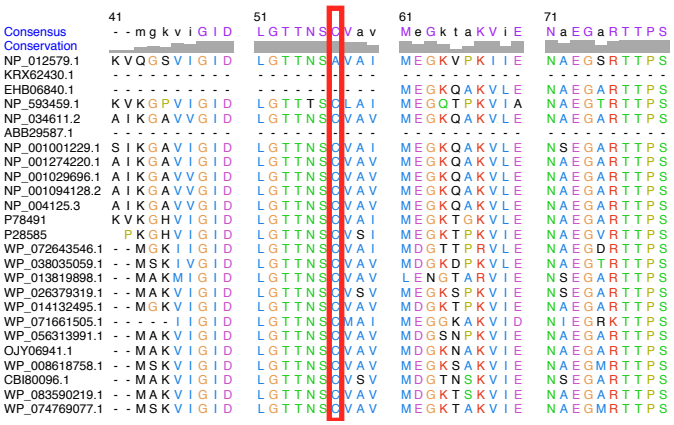


Figure 62: Multiple alignment of 28 different Hsp70 members. Code corresponds to the UniProt Database code. Red Box: conserved cysteine which is missing in Ssc1.

BIBLIOGRAPHY

[1] Colin Echeverría Aitken, R Andrew Marshall, and Joseph D Puglisi. "An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments." In: *Biophysical journal* 94.5 (Mar. 2008), pp. 1826–1835 (cit. on p. 114).

[2] Alexander E Aleshin, Chenbo Zeng, Hans D Bartunik, Herbert J Fromm, and Richard B Honzatko. "Regulation of hexokinase I: crystal structure of recombinant human brain hexokinase complexed with glucose and phosphate11." In: *Journal of Molecular Biology* (1998) (cit. on p. 77).

[3] C B Anfinsen. "The formation of the tertiary structure of proteins." In: *Harvey lectures* 61 (1967), pp. 95–116 (cit. on p. 5).

[4] A Ashkin. "Acceleration and Trapping of Particles by Radiation Pressure." In: *Physical Review Letters* 24.4 (Jan. 1970), pp. 156–159 (cit. on p. 17).

[5] A Ashkin and J M Dziedzic. "Optical Trapping and Manipulation of Viruses and Bacteria." In: *Science* 235.4795 (1987), pp. 1517–1520 (cit. on p. 17).

[6] P BORK, C Sander, and A Valencia. "An Atpase Domain Common to Prokaryotic Cell-Cycle Proteins, Sugar Kinases, Actin, and Hsp70 Heat-Shock Proteins." In: *Proceedings of the National Academy of Sciences* 89.16 (1992), pp. 7290–7294 (cit. on pp. 15, 88).

[7] David Baker and Andrej Sali. "Protein Structure Prediction and Structural Genomics." In: *Science* 294.5540 (Oct. 2001), pp. 93–96 (cit. on p. 17).

[8] David Balchin, Manajit K Hayer-Hartl, and F Ulrich Hartl. "In vivo aspects of protein folding and quality control." In: *Science* (2016) (cit. on pp. 4, 6).

[9] Daniela Bauer, Dale R Merz, Benjamin Pelz, Kelly E Theisen, Gail Yacyshyn, Dejana Mokranjac, Ruxandra I Dima, Matthias Rief, and Gabriel Žoldák. "Nucleotides regulate the mechanical hierarchy between subdomains of the nucleotide binding domain of the Hsp70 chaperone DnaK." In: *Proceedings of the National Academy of Sciences of the United States of America* 112.33 (Aug. 2015), pp. 10389–10394 (cit. on pp. 18, 23, 56, 77).

[10] Akash Bhattacharya, Alexander V Kurochkin, Grover N B Yip, Yongbo Zhang, Eric B Bertelsen, and Erik R P Zuiderweg. "Allostery in Hsp70 chaperones is transduced by subdomain rotations." In: *Journal of Molecular Biology* 388.3 (May 2009), pp. 475–490 (cit. on pp. 78, 91).

[11] Marta Blamowska, Walter Neupert, and Kai Hell. "Biogenesis of the mitochondrial Hsp70 chaperone." In: *The Journal of Cell Biology* 199.1 (Oct. 2012), pp. 125–135 (cit. on pp. 14, 53, 80, 81).

[12] Marta Blamowska, Martin Sichting, Koyeli Mapa, Dejana Mokranjac, Walter Neupert, and Kai Hell. "ATPase Domain and Interdomain Linker Play a Key Role in Aggregation of Mitochondrial Hsp70 Chaperone Ssc1." In: *Journal of Biological Chemistry* 285.7 (Feb. 2010), pp. 4423–4431 (cit. on pp. 47, 79, 80).

[13] Marta Blamowska, Martin Sichting, Koyeli Mapa, Dejana Mokranjac, Walter Neupert, and Kai Hell. "ATPase domain and interdomain linker play a key role in aggregation of mitochondrial Hsp70 chaperone Ssc1." In: *The Journal of biological chemistry* 285.7 (Feb. 2010), pp. 4423–4431 (cit. on pp. 14, 52, 79–81).

[14] Andreas Bracher and Jacob Verghese. "The nucleotide exchange factors of Hsp70 molecular chaperones." In: *Frontiers in Molecular Biosciences* 2 (Apr. 2015), pp. 1–9 (cit. on p. 11).

[15] D Brehmer, S Rüdiger, C S Gässler, D Klostermeier, L Packschies, Jochen Reinstein, M P Mayer, and Bernd Bukau. "Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange." In: *Nature Structural Biology* (2001) (cit. on pp. 84, 91, 92).

[16] D Brehmer, S Rüdiger, C S Gässler, D Klostermeier, L Packschies, M P Mayer, and Bernd Bukau. "Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange." In: *Nature Structural Biology* 8.5 (May 2001), pp. 427–432 (cit. on p. 11).

[17] A Buchberger, H Theyssen, and H Schröder. "Nucleotide-induced conformational changes in the ATPase and substrate binding domains of the DnaK chaperone provide evidence for interdomain communication." In: *Journal of Biological* (1995) (cit. on p. 106).

[18] A Buchberger, H Schröder, M Büttner, A Valencia, and Bernd Bukau. "A conserved loop in the ATPase domain of the DnaK chaperone is essential for stable binding of GrpE." In: *Nature Structural Biology* 1.2 (Jan. 1994), pp. 95–101 (cit. on p. 92).

[19] Michael Bugiel, Aniruddha Mitra, Salvatore Girardo, Stefan Diez, and Erik Schäffer. "Measuring Microtubule Supertwist and Defects by Three-Dimensional-Force-Clamp Tracking of Single Kinesin-1 Motors." In: *Nano letters* (2018) (cit. on p. 18).

[20] Bernd Bukau. "Regulation of the Escherichia coli heat-shock response." In: *Molecular microbiology* 9.4 (Aug. 1993), pp. 671–680 (cit. on p. 13).

[21] Giulia Calloni, Taotao Chen, Sonya M Schermann, Hung-chun Chang, Pierre Genevaux, Federico Agostini, Gian Gaetano Tartaglia, Manajit Hayer-Hartl, and F Ulrich Hartl. "DnaK functions as a central hub in the E. coli chaperone network." In: *Cell reports* 1.3 (Mar. 2012), pp. 251–264 (cit. on p. 12).

[22] Lyra Chang, Andrea D Thompson, Peter Ung, Heather A Carlson, and Jason E Gestwicki. "Mutagenesis reveals the complex relationships between ATPase rate and the chaperone activities of Escherichia coli heat shock protein 70 (Hsp70/DnaK)." In: *The Journal of biological chemistry* 285.28 (July 2010), pp. 21282–21291 (cit. on p. 57).

[23] Aaron Chevalier et al. "Massively parallel de novo protein design for targeted therapeutics." In: *Nature Publishing Group* 550.7674 (Oct. 2017), pp. 74–79 (cit. on p. 17).

[24] Eugenia M Clerico, Joseph M Tilitsky, Wenli Meng, and Lila M Gierasch. "How Hsp70 Molecular Machines Interact with Their Substrates to Mediate Diverse Physiological Functions." In: *Journal of Molecular Biology* 427.7 (2015), pp. 1575–1588 (cit. on pp. 8, 12).

[25] "Comparing Proteins by Their Unfolding Pattern." In: *Biophysical journal* 95.1 (July 2008), pp. 426–434 (cit. on p. 117).

[26] E A Craig, J Kramer, and J Kosic-Smithers. "SSC1, a member of the 70-kDa heat shock protein multigene family of Saccharomyces cerevisiae, is essential for growth." In: *Proceedings of the National Academy of Sciences* 84.12 (June 1987), pp. 4156–4160 (cit. on p. 14).

[27] E A Craig, J Kramer, J Shilling, M Werner-Washburne, S Holmes, J Kosic-Smithers, and C M Nicolet. "SSC1, an essential member of the yeast HSP70 multigene family, encodes a mitochondrial protein." In: *Molecular and cellular biology* 9.7 (July 1989), pp. 3000–3008 (cit. on pp. 14, 80).

[28] Elizabeth A Craig and Carol A Gross. "Is hsp70 the cellular thermometer?" In: *Trends in biochemical sciences* 16 (Jan. 1991), pp. 135–140 (cit. on p. 13).

[29] Liza Dahal, Tristan O C Kwan, and Jane Clarke. "pKID Binds to KIX via an Unstructured Transition State with Nonnative Interactions." In: *Biophysical journal* 113.12 (Dec. 2017), pp. 2713–2722 (cit. on p. 3).

[30] P R Dores-Silva, K Minari, C H I Ramos, L R S Barbosa, and J C Borges. "Structural and stability studies of the human mtHsp70-escort protein 1: an essential mortalin co-chaperone." In: *International journal of biological macromolecules* 56 (May 2013), pp. 140–148 (cit. on p. 14).

[31] R John Ellis and Allen P Minton. "Protein aggregation in crowded environments." In: *Biological chemistry* (2006) (cit. on p. 6).

[32] P Emsley and K Cowtan. "Coot: model-building tools for molecular graphics." In: *Acta crystallographica. Section D, Biological crystallography* 60.Pt 12 Pt 1 (Dec. 2004), pp. 2126–2132 (cit. on p. 109).

[33] Charles A English, Woody Sherman, Wenli Meng, and Lila M Gierasch. "The Hsp70 interdomain linker is a dynamic switch that enables allosteric communication between two structured domains." In: *The Journal of biological chemistry* (2017) (cit. on p. 77).

[34] E Evans. "Dynamic strength of molecular adhesion bonds." In: *Biophysj* 72.4 (Apr. 1997), pp. 1541–1555 (cit. on p. 119).

[35] Flaherty, K M, DeLuca-Flaherty, C, and McKay, D B. "Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein." In: *Nature* 346.6285 (Aug. 1990), pp. 623–628 (cit. on pp. 8, 15).

[36] K M Flaherty, D B McKay, W Kabsch, and K C Holmes. "Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein." In: *Proceedings of the National Academy of Sciences* 88.11 (June 1991), pp. 5041–5045 (cit. on pp. 10, 15, 77, 84, 86).

[37] R J Fletterick, D J Bates, and T A Steitz. "The structure of a yeast hexokinase monomer and its complexes with substrates at 2.7-A resolution." In: *Proceedings of the National Academy of Sciences* 72.1 (Jan. 1975), pp. 38–42 (cit. on p. 77).

[38] Gebhardt, J Christof M, Bornschlögl, Thomas, and Rief, Matthias. "Full distance-resolved folding energy landscape of one single protein molecule." In: *Proceedings of the National Academy of Sciences of the United States of America* 107.5 (Feb. 2010), pp. 2013–2018 (cit. on p. 120).

[39] Costa Georgopoulos, Krzysztof Liberek, Maciej Zylicz, and Debbie Ang. "9 Properties of the Heat Shock Proteins of Escherichia coli and the Autoregulation of the Heat Shock Response." In: *Cold Spring Harbor Monograph Archive* 26.0 (1994) (cit. on p. 13).

[40] Viara Grantcharova, Eric J Alm, David Baker, and Arthur L Horwich. "Mechanisms of protein folding." In: *Current Opinion in Structural Biology* 11.1 (Feb. 2001), pp. 70–82 (cit. on p. 4).

[41] Marco Grison, Ulrich Merkel, Julius Kostan, Kristina Djinović-Carugo, and Matthias Rief. "α-Actinin/titin interaction: A dynamic and mechanically stable cluster of bonds in the muscle Z-disk." In: *Proceedings of the National Academy of Sciences of the United States of America* 114.5 (Jan. 2017), pp. 1015–1020 (cit. on p. 17).

[42] Carol Gross, David Straus, James Erickson, and Takashi Yura. "8 The Function and Regulation of Heat Shock Proteins in Escherichia coli." In: *Cold Spring Harbor Monograph Archive* 19.0 (1990) (cit. on p. 13).

[43] C J Harrison, M Hayer-Hartl, M Di Liberto, F Hartl, and J Kuriyan. "Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK." In: *Science* 276.5311 (Apr. 1997), pp. 431–435 (cit. on p. 11).

[44] C Harrison. "GrpE, a nucleotide exchange factor for DnaK." In: *Cell stress & chaperones* 8.3 (2003), pp. 218–224 (cit. on p. 84).

[45] F Ulrich Hartl. "Protein Misfolding Diseases." In: *Annual Review of Biochemistry* 86.1 (June 2017), pp. 21–26 (cit. on p. 6).

[46] F Ulrich Hartl and Manajit K Hayer-Hartl. "Converging concepts of protein folding in vitro and in vivo." In: *Nature Structural & Molecular Biology* (2009) (cit. on p. 9).

[47] T Hesterkamp and Bernd Bukau. "Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in E.coli." In: *The EMBO Journal* 17.16 (Aug. 1998), pp. 4818–4828 (cit. on p. 12).

[48] M Horst. "Sequential action of two hsp70 complexes during protein import into mitochondria." In: *The EMBO Journal* 16.8 (Apr. 1997), pp. 1842–1849 (cit. on p. 14).

[49] J H Hurley. "The sugar kinase/heat shock protein 70/actin superfamily: implications of conserved structure for mechanism." In: *Annual review of biophysics and biomolecular structure* 25.1 (1996), pp. 137–162 (cit. on p. 16).

[50] Markus Jahn, Alexandra Rehn, Benjamin Pelz, Björn Hellenkamp, Klaus Richter, Matthias Rief, Johannes Buchner, and Thorsten Hugel. "The charged linker of the molecular chaperone Hsp90 modulates domain contacts and biological function." In: *Proceedings of the National Academy of Sciences of the United States of America* 111.50 (Dec. 2014), pp. 17881–17886 (cit. on p. 17).

[51] Ursula Jakob, Wilson Muse, Markus Eser, and James C A Bardwell. "Chaperone Activity with a Redox Switch." In: *Cell* 96.3 (Feb. 1999), pp. 341–352 (cit. on p. 13).

[52] Wolfgang Kabsch. "Integration, scaling, space-group assignment and post-refinement." In: *Acta crystallographica. Section D, Biological crystallography* (2010) (cit. on p. 109).

[53] P J Kang, J Ostermann, J Shilling, W Neupert, E A Craig, and N Pfanner. "Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins." In: *Nature* 348.6297 (Nov. 1990), pp. 137–143 (cit. on p. 14).

[54] Shigeyuki Kawai, Takako Mukai, Shigetarou Mori, Bunzo Mikami, and Kousaku Murata. "Hypothesis: Structures, evolution, and ancestor of glucose kinases in the hexokinase family." In: *Journal of Bioscience and Bioengineering* 99.4 (Apr. 2005), pp. 320–330 (cit. on p. 86).

[55] M S Kellermayer, S B Smith, H L Granzier, and C Bustamante. "Folding-unfolding transitions in single titin molecules characterized with laser tweezers." In: *Science* 276.5315 (May 1997), pp. 1112–1116 (cit. on p. 17).

[56] Thomas Kiefhaber, Annett Bachmann, and Kristine Steen Jensen. "Dynamics and mechanisms of coupled protein folding and binding reactions." In: *Current Opinion in Structural Biology* 22.1 (Feb. 2012), pp. 21–29 (cit. on p. 3).

[57] Roman Kityk, Jürgen Kopp, and Matthias P Mayer. "Molecular Mechanism of J-Domain-Triggered ATP Hydrolysis by Hsp70 Chaperones." In: *Molecular Cell* (2017) (cit. on p. 93).

[58] Roman Kityk, Jürgen Kopp, Irmgard Sinning, and Matthias P Mayer. "Structure and Dynamics of the ATP-Bound Open Conformation of Hsp70 Chaperones." In: *Molecular Cell* 48.6 (Dec. 2012), pp. 863–874 (cit. on pp. 78, 85).

[59] Roman Kityk, Markus Vogel, Rainer Schlecht, Bernd Bukau, and Matthias P Mayer. "Pathways of allosteric regulation in Hsp70 chaperones." In: *Nature Communications* 6 (2015), p. 8308 (cit. on pp. 77, 78).

[60] Andrey Kosolapov and Carol Deutsch. "Tertiary interactions within the ribosomal exit tunnel." In: *Nature Structural & Molecular Biology* (2009) (cit. on p. 3).

[61] Sathish Kumar Lakshmipathy, Rashmi Gupta, Stefan Pinkert, Stephanie Anne Etchells, and F Ulrich Hartl. "Versatility of trigger factor interactions with ribosome-nascent chain complexes." In: *Journal of Biological Chemistry* 285.36 (Sept. 2010), pp. 27911–27923 (cit. on p. 7).

[62] Thomas Laufen, Matthias P Mayer, Christian Beisel, Dagmar Klostermeier, Axel Mogk, Jochen Reinstein, and Bernd Bukau. "Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones." In: *PNAS* 96.10 (May 1999), pp. 5452–5457 (cit. on p. 93).

[63] C Levinthal. "Are there pathways for protein folding." In: *J Chim phys* (1968) (cit. on p. 5).

[64] K Liberek, D Skowyra, M Zylicz, C Johnson, and C Georgopoulos. "The Escherichia coli DnaK chaperone, the 70-kDa heat shock protein eukaryotic equivalent, changes conformation upon ATP hydrolysis, thus triggering its dissociation from a bound target protein." In: *Journal of Biological Chemistry* 266.22 (Aug. 1991), pp. 14491–14496 (cit. on p. 61).

[65] "LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery." In: 51.10 (Oct. 2011), pp. 2778–2786 (cit. on pp. 8, 91).

[66] Magnus O Lindberg. "Malleability of protein folding pathways: a simple reason for complex behaviour." In: *Current Opinion in Structural Biology* 17.1 (Feb. 2007), pp. 21–29 (cit. on p. 5).

[67] Ana Lisica and Stephan W Grill. "Optical tweezers studies of transcription by eukaryotic RNA polymerases." In: *Biomolecular Concepts* 8.1 (Mar. 2017), pp. 1–11 (cit. on p. 17).

[68] Q Liu, J Krzewska, K Liberek, and E A Craig. "Mitochondrial Hsp70 Ssc1: role in protein folding." In: *Journal of Biological Chemistry* 276.9 (Mar. 2001), pp. 6112–6118 (cit. on p. 14).

[69] JC Maxwell Proceedings of the Royal Society of London and 1863. "A Dynamical Theory of the Electromagnetic Field." In: *JSTOR* (1864) (cit. on p. 17).

[70] H Maity, M Maity, MMG Krishna, L Mayne, and S Walter Englander. "Protein folding: The stepwise assembly of foldon units." In: *Proceedings of the National Academy of Sciences* 102.13 (2005), pp. 4741–4746 (cit. on p. 5).

[71] Haripada Maity, Jon N Rumbley, and S Walter Englander. "Functional role of a protein foldon–an Omega-loop foldon controls the alkaline transition in ferricytochrome c." In: *Proteins: Structure, Function, and Bioinformatics* 63.2 (May 2006), pp. 349–355 (cit. on p. 5).

[72] M Manosas, J D Wen, P T X Li, S B Smith, C Bustamante, I Tinoco Jr., and F Ritort. "Force Unfolding Kinetics of RNA using Optical Tweezers. II. Modeling Experiments." In: *Biophysical journal* 92.9 (May 2007), pp. 3010–3021 (cit. on p. 121).

[73] Mapa, K, Sikor, M, Kudryavtsev, V, and Waegemann, K. "The conformational dynamics of the mitochondrial Hsp70 chaperone." In: *Molecular Cell* (2010) (cit. on p. 14).

[74] M P Mayer and Bernd Bukau. "Hsp70 chaperones: Cellular functions and molecular mechanism." In: *Cellular and Molecular Life Sciences* 62.6 (Mar. 2005), pp. 670–684 (cit. on p. 11).

[75] Matthias P Mayer. "Gymnastics of Molecular Chaperones." In: *Molecular Cell* 39.3 (Aug. 2010), pp. 321–331 (cit. on p. 9).

[76] Matthias P Mayer and Roman Kityk. "Insights into the molecular mechanism of allostery in Hsp70s." In: *Frontiers in Molecular Biosciences* (2015) (cit. on p. 78).

[77] J S McCarty, A Buchberger, and Jochen Reinstein. "The role of ATP in the functional cycle of the DnaK chaperone system." In: *Journal of molecular biology* 249.1 (1995), pp. 126–137 (cit. on pp. 10, 11).

[78] Airlie J McCoy, Ralf W Grosse-Kunstleve, Paul D Adams, Martyn D Winn, Laurent C Storoni, and Randy J Read. "Phaser crystallographic software." In: *Journal of Applied Crystallography* 40.Pt 4 (Aug. 2007), pp. 658–674 (cit. on p. 109).

[79] Sarah Meinhold. "Analyse der NBD Untereinheiten von DnaK." In: *Bachelor Thesis TUM, Biophysik Prof. Dr. Matthias Rief* (2015) (cit. on p. 77).

[80] Sarah Meinhold. "Folding mechanics of the truncated nucleotide binding domain of DnaK (NBD$\Delta\alpha$) by single molecule force spectroscopy." In: *Laboratory Report TUM, Biophysik Prof. Dr. Matthias Rief* (2016) (cit. on p. 56).

[81] Bingjie Miao, Julie E Davis, and Elizabeth A Craig. "Mge1 functions as a nucleotide release factor for Ssc1, a mitochondrial Hsp70 of Saccharomyces cerevisiae." In: *Journal of Molecular Biology* 265.5 (Jan. 1997), pp. 541–552 (cit. on p. 14).

[82] Dejana Mokranjac and Walter Neupert. "Thirty years of protein translocation into mitochondria: Unexpectedly complex and still puzzling." In: *BBA - Molecular Cell Research* 1793.1 (Jan. 2009), pp. 33–41 (cit. on p. 14).

[83] Diana L Montgomery, Richard I Morimoto, and Lila M Gierasch. "Mutations in the substrate binding domain of the Escherichia coli 70 kda molecular chaperone, DnaK, which alter substrate affinity or interdomain coupling." In: *Journal of Molecular Biology* 286.3 (Feb. 1999), pp. 915–932 (cit. on pp. 43, 58, 107, 136).

[84] Walter Neupert and Michael Brunner. "The protein import motor of mitochondria." In: *Nature reviews. Molecular cell biology* 3.8 (Aug. 2002), pp. 555–565 (cit. on p. 14).

[85] Matthew P Nicholas, Florian Berger, Lu Rao, Sibylle Brenner, Carol Cho, and Arne Gennerich. "Cytoplasmic dynein regulates its attachment to microtubules via nucleotide state-switched mechanosensing at multiple AAA domains." In: *Proceedings of the National Academy of Sciences* 112.20 (2015), pp. 6371–6376 (cit. on pp. 17, 18).

[86] Edward P O'Brien, John Christodoulou, Michele Vendruscolo, and Christopher M Dobson. "New scenarios of protein folding can occur on the ribosome." In: *Journal of the American Chemical Society* (2011) (cit. on p. 3).

[87] M C O'Brien, K M Flaherty, and D B McKay. "Lysine 71 of the chaperone protein Hsc70 Is essential for ATP hydrolysis." In: *Journal of Biological Chemistry* 271.27 (July 1996), pp. 15874–15878 (cit. on p. 84).

[88] Melanie C O'Brien and David B McKay. "How Potassium Affects the Activity of the Molecular Chaperone Hsc70 I. POTASSIUM IS REQUIRED FOR OPTIMAL ATPase ACTIVITY." In: *Journal of Biological Chemistry* 270.5 (Feb. 1995), pp. 2247–2250 (cit. on p. 10).

[89] L Packschies, H Theyssen, A Buchberger, Bernd Bukau, and R S Goody. "GrpE accelerates nucleotide exchange of the molecular chaperone DnaK with an associative displacement mechanism." In: *Biochemistry* 36.12 (1997), pp. 3417–3422 (cit. on p. 91).

[90] Jimin Pei and Bong-Hyun Kim. "PROMALS3D: a tool for multiple protein sequence and structure alignments." In: *Nucleic Acids Research* 36.7 (Apr. 2008), pp. 2295–2300 (cit. on p. 87).

[91] Benjamin Pelz, Gabriel Žoldák, Fabian Zeller, Martin Zacharias, and Matthias Rief. "Subnanometre enzyme mechanics probed by single-molecule force spectroscopy." In: *Nature Communications* 7 (Feb. 2016), p. 10848 (cit. on p. 17).

[92] Ruifeng Qi et al. "Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP." In: *Nature Structural & Molecular Biology* 20.7 (July 2013), pp. 900–907 (cit. on pp. 77, 109).

[93] Lu Rao, Joshua D Dworkin, William E Nell, and Ulrich Bierbach. "Interactions of a platinum-modified perylene derivative with the human telomeric G-quadruplex." In: *The journal of physical chemistry. B* 115.46 (Nov. 2011), pp. 13701–13712 (cit. on p. 17).

[94] Robbie Reutzel, Craig Yoshioka, Lakshmanan Govindasamy, Elena G Yarmola, Mavis Agbandje-McKenna, Michael R Bubb, and Robert McKenna. "Actin crystal dynamics: structural implications for F-actin nucleation, polymerization, and branching mediated by the anti-parallel dimer." In: *Journal of Structural Biology* 146.3 (June 2004), pp. 291–301 (cit. on p. 77).

[95] P Rice, I Longden, and A Bleasby. "EMBOSS: the European Molecular Biology Open Software Suite." In: *Trends in genetics : TIG* 16.6 (June 2000), pp. 276–277 (cit. on pp. 47, 80, 86).

[96] Gabriel J Rocklin et al. "Global analysis of protein folding using massively parallel design, synthesis, and testing." In: *Science* 357.6347 (July 2017), pp. 168–175 (cit. on p. 17).

[97] Lorenz Rognoni, Johannes Stigler, Benjamin Pelz, Jari Ylänne, and Matthias Rief. "Dynamic force sensing of filamin revealed in single-molecule experiments." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.48 (Nov. 2012), pp. 19679–19684 (cit. on p. 17).

[98] Luiza K Sanjuán Szklarz et al. "Inactivation of the Mitochondrial Heat Shock Protein Zim17 Leads to Aggregation of Matrix Hsp70s Followed by Pleiotropic Effects on Morphology and Protein Biogenesis." In: *Journal of Molecular Biology* 351.1 (Aug. 2005), pp. 206–218 (cit. on p. 14).

[99] K Scheffzek, M R Ahmadian, W Kabsch, L Wiesmuller, A Lautwein, F Schmitz, and A Wittinghofer. "The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic Ras mutants." In: *Science* 277.5324 (1997), pp. 333–338 (cit. on p. 85).

[100] Michael Schlierf, Felix Berkemeier, and Matthias Rief. "Direct observation of active protein folding using lock-in force spectroscopy." In: *Biophysical journal* 93.11 (Dec. 2007), pp. 3989–3998 (cit. on p. 120).

[101] Michael Schlierf, Zu Thur Yew, Matthias Rief, and Emanuele Paci. "Complex unfolding kinetics of single-domain proteins in the presence of force." In: *Biophysical journal* 99.5 (Sept. 2010), pp. 1620–1627 (cit. on p. 17).

[102] Dennis J Selkoe. "Folding proteins in fatal ways." In: *Nature Publishing Group* 426.6968 (Dec. 2003), pp. 900–904 (cit. on p. 6).

[103] Yasuhito Shomura, Zdravko Dragovic, Hung-chun Chang, Nikolay Tzvetkov, Jason C Young, Jeffrey L Brodsky, Vince Guerriero, F Ulrich Hartl, and Andreas Bracher. "Regulation of Hsp70 function by HspBP1: structural analysis reveals an alternate mechanism for Hsp70 nucleotide exchange." In: *Molecular Cell* 17.3 (Feb. 2005), pp. 367–379 (cit. on pp. 61, 81, 83).

[104] Martin Sichting, Dejana Mokranjac, Abdussalam Azem, Walter Neupert, and Kai Hell. "Maintenance of structure and function of mitochondrial Hsp70 chaperones requires the chaperone Hep1." In: *The EMBO Journal* 24.5 (Mar. 2005), pp. 1046–1056 (cit. on pp. 14, 79, 80).

[105] H Sondermann, C Scheufler, C Schneider, J Hohfeld, F U Hartl, and I Moarefi. "Structure of a Bag/Hsc70 complex: Convergent functional evolution of Hsp70 nucleotide exchange factors." In: *Science* 291.5508 (2001), pp. 1553–1557 (cit. on p. 11).

[106] Gabrielle Stetz and Gennady M Verkhivker. "Computational Analysis of Residue Interaction Networks and Coevolutionary Relationships in the Hsp70 Chaperones: A Community-Hopping Model of Allosteric Regulation and Communication." In: *PLoS Computational Biology* 13.1 (Jan. 2017), e1005299–34 (cit. on p. 78).

[107] Johannes Stigler and Matthias Rief. "Calcium-dependent folding of single calmodulin molecules." In: *Proceedings of the National Academy of Sciences of the United States of America* (2012) (cit. on p. 17).

[108] Johannes Stigler and Matthias Rief. "Hidden Markov analysis of trajectories in single-molecule experiments and the effects of missed events." In: *ChemPhysChem* (2012) (cit. on p. 118).

[109] Johannes Stigler and Matthias Rief. "Ligand-Induced Changes of the Apparent Transition-State Position in Mechanical Protein Unfolding." In: *Biophysical journal* 109.2 (July 2015), pp. 365–372 (cit. on p. 121).

[110] Johannes Stigler, Fabian Ziegler, Anja Gieseke, J Christof M Gebhardt, and Matthias Rief. "The complex folding network of single calmodulin molecules." In: *Science* 334.6055 (Oct. 2011), pp. 512–516 (cit. on p. 17).

[111] Joanna F Swain and Lila M Gierasch. "A new twist for an Hsp70 chaperone." In: *Nature Structural Biology* 9.6 (June 2002), pp. 406–408 (cit. on p. 58).

[112] Joanna F Swain, Gizem Dinler, Renuka Sivendran, Diana L Montgomery, Mathias Stotz, and Lila M Gierasch. "Hsp70 Chaperone Ligands Control Domain Association via an Allosteric Mechanism Mediated by the Interdomain Linker." In: *Molecular Cell* 26.1 (Apr. 2007), pp. 27–39 (cit. on pp. 11, 66, 136).

[113] Stefka G Taneva, Fernando Moro, Adrián Velázquez-Campoy, and Arturo Muga. "Energetics of nucleotide-induced DnaK conformational states." In: *Biochemistry* 49.6 (Feb. 2010), pp. 1338–1345 (cit. on pp. 58, 66).

[114] C Tanford. "The hydrophobic effect and the organization of living matter." In: *Science* 200.4345 (June 1978), pp. 1012–1018 (cit. on p. 3).

[115] Thomas C Terwilliger, Ralf W Grosse-Kunstleve, Pavel V Afonine, Nigel W Moriarty, Paul D Adams, Randy J Read, Peter H Zwart, and Li-Wei Hung. "Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias." In: *Acta crystallographica. Section D, Biological crystallography* (2008) (cit. on p. 109).

[116] S A Teter, W A Houry, D Ang, T Tradler, D Rockabrand, G Fischer, P Blum, C Georgopoulos, and F Ulrich Hartl. "Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains." In: *Cell* 97.6 (June 1999), pp. 755–765 (cit. on p. 12).

[117] Holger Theyssen, Hans-Peter Schuster, Lars Packschies, Bernd Bukau, and Jochen Reinstein. "The Second Step of ATP Binding to DnaK Induces Peptide Release." In: *Journal of Molecular Biology* 263.5 (Nov. 1996), pp. 657–670 (cit. on pp. 9–11, 26, 43, 58, 66, 136).

[118] Ignacio Tinoco and Carlos Bustamante. "The ribosome modulates nascent protein folding." In: *Science* 334.6063 (Dec. 2011), pp. 1723–1727 (cit. on p. 17).

[119] Kaye N Truscott et al. "A J-protein is an essential subunit of the presequence translocase-associated protein import motor of mitochondria." In: *The Journal of Cell Biology* 163.4 (Nov. 2003), pp. 707–713 (cit. on p. 14).

[120] Peter Man-Un Ung, Andrea D Thompson, Lyra Chang, Jason E Gestwicki, and Heather A Carlson. "Identification of Key Hinge Residues Important for Nucleotide-Dependent Allostery in E. coli Hsp70/DnaK." In: *PLoS Computational Biology* 9.11 (Nov. 2013), e1003279–11 (cit. on p. 11).

[121] Vladimir N Uversky. "Functional roles of transiently and intrinsically disordered regions within proteins." In: *The FEBS journal* 282.7 (Apr. 2015), pp. 1182–1189 (cit. on p. 5).

[122] Markus Vogel, Bernd Bukau, and Matthias P Mayer. "Allosteric Regulation of Hsp70 Chaperones by a Proline Switch." In: *Molecular Cell* 21.3 (Feb. 2006), pp. 359–367 (cit. on p. 84).

[123] Markus Vogel, Matthias P Mayer, and Bernd Bukau. "Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker." In: *Journal of Biological Chemistry* 281.50 (Dec. 2006), pp. 38705–38711 (cit. on p. 78).

[124] Sam Walcott. "The load dependence of rate constants." In: *The Journal of Chemical Physics* (2008) (cit. on p. 119).

[125] G C Walker. "Mutations altering heat shock specific subunit of RNA polymerase suppress major cellular defects of E. coli mutants lacking the DnaK chaperone." In: *The EMBO Journal* 9.12 (Dec. 1990), pp. 4027–4036 (cit. on p. 12).

[126] A C Wallace, R A Laskowski, and J M Thornton. "LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions." In: *Protein engineering* 8.2 (Feb. 1995), pp. 127–134 (cit. on p. 8).

[127] Wang, M D, Schnitzer, M J, Yin, H, Landick, R, Gelles, J, and Block, S M. "Force and velocity measured for single molecules of RNA polymerase." In: *Science* 282.5390 (1998), pp. 902–907 (cit. on p. 17).

[128] M D Wang, H Yin, R Landick, J Gelles, and S M Block. "Stretching DNA with optical tweezers." In: *Biophysical journal* 72.3 (Mar. 1997), pp. 1335–1346 (cit. on p. 17).

[129] S M Wilbanks, C DELUCAFLAHERTY, and D B McKay. "Structural Basis of the 70-Kilodalton Heat-Shock Cognate Protein Atp Hydrolytic Activity .1. Kinetic Analyses of Active-Site Mutants." In: *Journal of Biological Chemistry* 269.17 (1994), pp. 12893–12898 (cit. on p. 84).

[130] S M Wilbanks and D B McKay. "How potassium affects the activity of the molecular chaperone Hsc70. II. Potassium binds specifically in the ATPase active site." In: *Journal of Biological Chemistry* 270.5 (Feb. 1995), pp. 2251–2257 (cit. on p. 10).

[131] J Wild, E Altman, T Yura, and C A Gross. "DnaK and DnaJ heat shock proteins participate in protein export in Escherichia coli." In: *Genes & development* 6.7 (July 1992), pp. 1165–1172 (cit. on p. 13).

[132] J Wild, P Rossmeissl, W A Walter, and C A Gross. "Involvement of the DnaK-DnaJ-GrpE chaperone team in protein secretion in Escherichia coli." In: *Journal of bacteriology* (1996) (cit. on p. 13).

[133] Jeannette Winter, Katrin Linke, Anna Jatzek, and Ursula Jakob. "Severe oxidative stress causes inactivation of DnaK and activation of the redox-regulated chaperone Hsp33." In: *Molecular Cell* 17.3 (Feb. 2005), pp. 381–392 (cit. on pp. 13, 79).

[134] Ching-Chung Wu, Vankadari Naveen, Chin-Hsiang Chien, Yi-Wei Chang, and Chwan-Deng Hsiao. "Crystal structure of DnaK protein complexed with nucleotide exchange factor GrpE in DnaK chaperone system: insight into intermolecular communication." In: *The Journal of biological chemistry* 287.25 (June 2012), pp. 21461–21470 (cit. on p. 91).

[135] Hong Zhang, Jie Yang, Si Wu, Weibin Gong, Chang Chen, and Sarah Perrett. "Glutathionylation of the Bacterial Hsp70 Chaperone DnaK Provides a Link between Oxidative Stress and the Heat Shock Response." In: *Journal of Biological Chemistry* 291.13 (Mar. 2016), pp. 6967–6981 (cit. on p. 79).

[136] Yongbo Zhang and Erik R P Zuiderweg. "The 70-kDa heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains." In: *Proceedings of the National Academy of Sciences* 101.28 (July 2004), pp. 10272–10277 (cit. on p. 91).

[137] Anastasia Zhuravleva and Lila M Gierasch. "Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones." In: *Proceedings of the National Academy of Sciences* 108.17 (2011), pp. 6987–6992 (cit. on p. 78).

[138] Gabriel Žoldák, Johannes Stigler, Benjamin Pelz, Hongbin Li, and Matthias Rief. "Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.45 (Nov. 2013), pp. 18156–18161 (cit. on pp. 17, 18).

[139] Erik R P Zuiderweg, Eric B Bertelsen, Aikaterini Rousaki, Matthias P Mayer, Jason E Gestwicki, and Atta Ahmad. "Allostery in the Hsp70 chaperone proteins." In: *Topics in current chemistry* (2013) (cit. on p. 78).

ACKNOWLEDGEMENTS

I want to thank those people who have contributed to the success of this work.

First, I want to thank Prof. Dr. Matthias Rief for giving me the opportunity and support to write this thesis. He gave me the freedom to try things out and supported new approaches. Thank you that your door was always open to discuss data and solve problems.

Second, I want to thank my mentor Dr. Gabriel Žoldák, who greatly supported me with his knowledge and contributed to the success of this thesis. He supported me to gain new knowledge by initiating collaborations or informing me about good workshops. Thank you very much, Gabriel!

Many thanks to Prof. Ruxandra Dima for the coarse-grained simulations, Dr. Tobias Madl and Gesa Richter for the SAXS measurements, Dr. Johannes Stigler for the ATP-model simulations and Dr. Roman Peter Jakob and Prof. Timm Maier for the crystal structure data. These people contributed to the success of this work by enriching the variety of methods applied to the EcNBD project. Thank you very much!

I want to thank the whole group of E22/E27 for the nice working atmosphere and helpfulness during my student and Ph.D. time.

Fabian Ziegler, Johannes Stigler, and Gabriel Žoldák who helped me a lot with IGOR.

A big thank you to my office mates, Andi, Uli, Anu, Gabriel, Kasia, and Mihai, who contributed to a relaxed time with lots of discussions, coffee, tea and cake supply.

Thanks to my students Matthias Brandl, Jonas Huber, Caroline Körösy, Sarah Meinhold, Moritz Strubl and Vera Wanka who help me a lot and for a great time in the lab.

I want to thank Ulrich Merkel, Andreas Weißl, Benjamin Pelz and Katarzyna Tych for the maintenance of the trap and advice at any time.

For all the technical support and help, I want to thank Rudi Lehrhuber, Monika Rusp, Gabi Chmel and Karin Vogt.

Elke Fehsenfeld and Nicole Mittermüller for their help in bureaucracy questions.

Last, I want to thank my family for their unlimited support during my study. Thank you for always standing behind me no matter what.

Anhang I

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung
Department für Physik

der TUM zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:
Single-molecule force spectroscopy reveals the domain hierarchy of DnaK NBD and identifies a folding seed and a minimal ATP-binding domain of DnaK

in Physik, Biophysik, E22, Prof. Rief
Fakultät, Institut, Lehrstuhl, Klinik, Krankenhaus, Abteilung

unter der Anleitung und Betreuung durch: Prof. Dr. Matthias Rief ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Ab. 6 und 7 Satz 2 angebotenen Hilfsmittel benutzt habe.

☒ Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

☒ Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

☐ Die vollständige Dissertation wurde in _____ veröffentlicht. Die promotionsführende Einrichtung

_____

hat der Veröffentlichung zugestimmt.

☒ Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

☐ Ich habe bereits am _____ bei der Fakultät für _____

_____

der Hochschule _____

unter Vorlage einer Dissertation mit dem Thema _____

_____

die Zulassung zur Promotion beantragt mit dem Ergebnis: _____

_____

Die öffentlich zugängliche Promotionsordnung der TUM ist mir bekannt, insbesondere habe ich die Bedeutung von § 28 (Nichtigkeit der Promotion) und § 29 (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Mit der Aufnahme meiner personenbezogenen Daten in die Alumni-Datei bei der TUM bin ich

☒ einverstanden, ☐ nicht einverstanden.

Garching, 24.07.2018, Unterschrift