MARKUS JAHN

HEAT SHOCK PROTEIN 90
STUDIED AT THE SINGLE MOLECULE LEVEL

DISSERTATION AT THE PHYSICS DEPARTMENT E22
TECHNISCHE UNIVERSITÄT MÜNCHEN

TECHNISCHE UNIVERSITÄT MÜNCHEN
DEPARTMENT E22

HEAT SHOCK PROTEIN 90 STUDIED AT THE SINGLE MOLECULE LEVEL

FROM STRUCTURE TO DYNAMICS

MARKUS JAHN

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
Prüfer der Dissertation:	1. Prof. Dr. Matthias Rief
	2. Prof. Dr. Thorsten Hugel
	3. Prof. Dr. Jan Lipfert

Die Dissertation wurde am 17.5.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 21.6.2016 angenommen.

ABSTRACT

Heat shock protein 90 is an abundant molecular chaperone essential for the maturation of a large set of client proteins. Structurally the large Hsp90 molecule is organized into three domains, where the C-terminal domain facilitates dimer formation. How such large systems can fold and assemble is investigated using optical tweezers. By manipulating one molecule at a time together with novel evaluation methods it is possible to dissect the multiple pathways and states of such a large system. While the C-terminal domain shows two-state folding behavior, the larger N-terminal and middle domains populate ensembles of misfolded states, which dominate and slow down the folding process. Linking the Hsp90 domains further impairs protein folding by the formation of cross-domain misfolds. These intra- and cross-domain misfolds decrease the effective folding time of Hsp90 from milliseconds to minutes and lead to the paradoxical effect that folding rates increase with a small applied force. Since the propensity of a protein to misfold increases with chain length, co-translational protein folding is important. After folding, Hsp90 assembles into its dimeric form. The kinetics are of special interest, because C-terminal dimerization may be an important feature in Hsp90's chaperone cycle. However, optical tweezers experiments show a stable dimerization with dissociation rates on the minute timescale.

The fully assembled Hsp90 system is very dynamic and linkers connecting Hsp90's domains facilitate the large conformational rearrangements needed for client chaperoning. A unique extended sequence between the N and M domains, called the Charged Linker, is investigated by an integrated approach, comprising optical tweezers, fluorescence and bulk experiments. Contrary to previous assumptions the Charged Linker is found not to be permanently unstructured but forms a structural motif by binding to the N domain. The stabilizing energy is only $1.1\ k_B T$, therefore the Charged Linker can easily dissociate, allowing conformational rearrangement of the N-terminal domains. The millisecond dynamics of the Charged Linker modulate the much slower N-terminal opening and closing kinetics of the dimer and the Charged Linker is necessary for reaching one important closed state of Hsp90. The N-terminal dynamics, which so far could only be resolved in single molecule FRET experiments, are also measured in optical tweezers. Contrary to other experiments a single compact closed state, which is only populated in the presence of ATP is found. These findings pave the way for measuring the complete cycle of Hsp90 with optical tweezers and, most important, understand its role in client chaperoning.

PUBLICATIONS

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

M. Jahn, J. Buchner, T. Hugel, and M. Rief. Folding and assembly of the large molecular machine hsp90 studied in single-molecule experiments. *Proc Natl Acad Sci U S A*, 113(5):1232–7, 2016.

M. Jahn, A. Rehn, B. Pelz, B. Hellenkamp, K. Richter, M. Rief, J. Buchner, and T. Hugel. The charged linker of the molecular chaperone hsp90 modulates domain contacts and biological function. *Proc Natl Acad Sci U S A*, 111(50):17881–6, 2014.

"It's still magic even if you know how it's done."

Terry Pratchett

ACKNOWLEDGMENTS

I express my sincere gratitude to my supervisors, Matthias Rief and Thorsten Hugel, for their continuous support and guidance. Their brilliant comments and suggestions have shaped this thesis and my notion of science. The way they manage to do cutting-edge research in a relaxed environment made my studies a very profitable and enjoyable part of my life.

I thank my fellow labmates for being great and awesome people, whose assistance was vital for the successful completion of this thesis. The fun we had inside and outside the lab made it possible to bear the downs and enjoy the ups during my studies.

A special thanks to Alena Dudarenka and Matthias Jahn for helping me a lot with artwork and creating shiny figures and illustrations.

Last but not the least, I would like to thank my family: my parents, my brothers and my girlfriend for supporting me spiritually throughout my studies and my life in general.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ADP Adenosine diphosphate

AMP-PNP Adenylyl-imidodiphosphate

ATP Adenosine triphosphate

ATPγS Adenosine 5-(γ-thio)triphosphate

BSA Bovine serum albumin

DNA Deoxyribonucleic acid

EDC 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide

eWLC Extensible worm-like chain

HMM Hidden Markov model

NHS N-Hydroxysuccinimide

NMR Nuclear magnetic resonance

PBS Phosphate-buffered saline

PCR Polymerase chain reaction

PEG Polyethylene glycol

RNA Ribonucleic acid

SD Standard deviation

SDS-PAGE Sodium dodecyl sulfate - Polyacrylamide gel electrophoresis

TCEP Tris(2-carboxyethyl)phosphin

TIR Total internal reflection

WLC Worm-like chain

WT Wild type

Part I

INTRODUCTION

In this part I provide the essential background information about protein folding and the molecular chaperone heat shock protein 90. I also discuss the advantages and disadvantages of using single molecule methods to study biomolecular systems and give a short outline.

1 THE PROTEIN FOLDING PROBLEM

All cellular functions, such as metabolism, cargo transport, signal transduction and movement are mediated by proteins. Therefore an understanding of how proteins work holds the key in understanding life itself. The human genome project identified about 20000 genes that encode proteins [51]. The number of unique proteins is even higher because genes are processed before translation and proteins can be modified afterwards. The translation from genomic information into proteins happens in the ribosome, a large molecular machine, which itself consists of different proteins. There the building blocks of proteins, the 20 different aminoacid residues, are covalently connected one after the other. The result is a linear polypeptide chain with a defined sequence of aminoacids. Before proteins can fulfill their functions in the cell they have to undergo a transition from their linear form to a unique three dimensional structure, called the *native state*. This process is called protein folding.

The native state is typically the *global free energy minimum* of the aminoacid chain in its solvent. Although the protein backbone has the same chemical composition for all proteins, the side chains of the aminoacids vary, leading to different folds from protein to protein. This means that the native state is defined by the aminoacid sequence, also called the primary structure. Folded proteins have similar substructures termed the secondary structure. The most prominent motifs are alpha helices and beta sheets, which appear for almost all aminoacid types, because they are stabilized by hydrogen bonds between the proteins' backbone residues. The tertiary structure is the final arrangement of the secondary structure elements that are connected by tight turns or flexible loops. Larger proteins often consist of autonomous folding units, called domains. After folding, identical or different proteins can assemble into larger complexes referred to as quaternary structures, which typically form the dynamic machines essential to accomplishing cellular functions.

How a protein can successfully fold from the linear unfolded state to its native state is a sophisticated problem. Typically a protein domain has an average length of 100 aminoacids [158] and the polypeptide chain and each residue have multiple degrees of freedom. The *Levinthal's paradox* illustrates the complexity of the conformational search for the native state [174], as in the following example: assuming a protein consists of 100 residues and each residue has only two degrees of freedom, this yields 2^{100} possible conformations. If a conformational change takes 1 ps, it would take more than 10^{10} years to

sample all possible conformations. This shows that a random search for the native state is not feasible. In experiments, folding rates range from microseconds to minutes and folding transitions are highly cooperative [155]. This led to the assumption that proteins fold along distinct pathways without accessing their total conformational space. Along these pathways, kinetically stable conformations referred to as *intermediate states* can exist. On-pathway intermediates are *en route* to the native state, while off-pathway intermediates involve non-native, misfolded structures.

The modern view of protein folding that reconciles the single pathways theory with the endless conformational possibilities is the description of the folding process as an energy minimization in a *funnel-shaped energy landscape* [99], as shown in Figure 1.1. The folding protein preferentially chooses one of multiple routes with minimal frustration. It may be trapped in intermediates that appear as local minima in the energy landscape. Multiple small barriers are interpreted as the roughness of the energy landscape and can decrease folding rates [88]. The funnel concept can describe the observed folding processes well, but the exact shape of the energy landscape is hard to map out experimentally. Nowadays it is possible, for small proteins, to reproduce the folding process in all-atom molecular dynamics computer simulations, advancing towards the ultimate goal of *de novo* structure prediction from the aminoacid sequence [73, 66].

Although nature has optimized folding, by adopting smooth energy landscapes with a kinetically accessible native state [155], the folding process in the crowded environment of a cell can be more complicated. In the cell, the interaction between partially folded proteins can form stable *aggregates*, that are not functional and implicated in neurodegenerative diseases [127]. To minimize this problem a special class of proteins called *chaperones* evolved. They actively assist folding or prevent aggregation by binding to misfolded species and keeping them in solution.

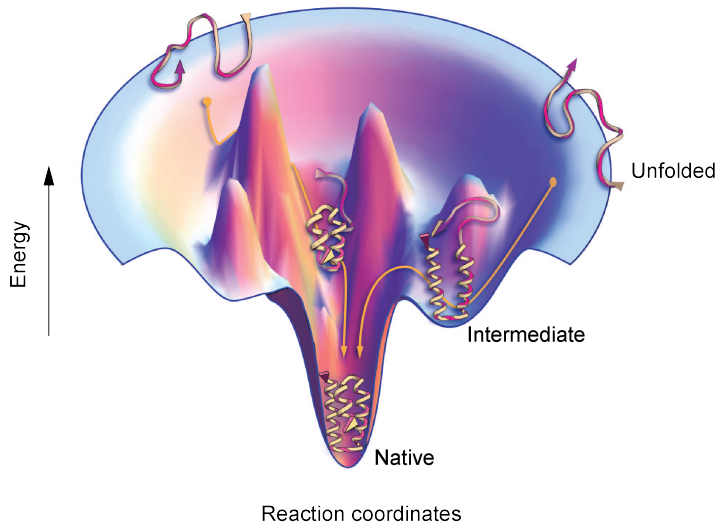

Figure 1.1: Protein folding in the energy landscape picture. The protein folding process from the unfolded state into the native state is driven by an energy minimization in an energy landscape. Typically all possible conformations are projected on one or a few reaction coordinates that describe the essence of the folding process. On the way to the native state local energy minima leading to intermediate states can occur. This figure is adapted from [27].

2

HEAT SHOCK PROTEIN 90

Heat shock proteins are a class of molecular chaperones that are up-regulated in cells upon stress induced by non-physiological conditions such as elevated temperature. Stress favors protein misfolding and therefore aggregation and heat shock proteins are needed to maintain proteostasis. The working mechanisms are different from chaperone to chaperone, as reviewed in [81]. Heat shock protein 70 binds stretches of hydrophobic aminoacids, preventing hydrophobic interactions and aggregation. Heat shock protein 60 protomers assemble into a large complex forming a cavity that can isolate misfolded proteins. Another chaperone, heat shock protein 100 forms rings that are able to unravel already misfolded or aggregated proteins.

The numbers refer to the molecular weight of the respective chaperone in kilodalton.

This work focuses on the *molecular chaperone* heat shock protein 90 (Hsp90). It is highly conserved from bacteria to human, is one of the most abundant proteins of the cell and is essential in eukaryotes [13]. Its clientele is functionally and structurally diverse, including kinases, transcription factors, ligases and telomerases [144, 92, 108]. In yeast Hsp90 interacts physically or genetically with 10% of the genome [170], thereby influencing a wide range of cellular functions. Unlike other chaperones Hsp90 is more selective choosing its clients maturing or regulating them at a very late step in the protein folding process [56].

The following sections give an introduction to the current state-of-the-art research into prokaryotic and eukaryotic Hsp90. They focus on structural organization, ATP hydrolysis, conformational changes, the interaction with co-chaperones and clients and the significance of Hsp90 in clinics.

2.1 STRUCTURE

Hsp90 proteins are organized as dimers, within which each monomer consists of *three domains*: the N-terminal domain (N), the middle domain (M) and the C-terminal domain (C). Figure 2.1 shows the crystal structure of the yeast Hsp90 homodimer [2].

The N domain consists of an eight-strand beta sheet, flanked by alpha helices. A resulting cleft has been identified as the ATP binding pocket [110], where ATP binds in an unusually kinked conformation. Upon binding, closing of the so-called ATP lid (residues 94 to 125 in yeast) locally changes the surface of the N domain [133, 28]. This evolutionary conserved fold with the special binding pocket makes Hsp90 a member of the GHKL super-family [31].

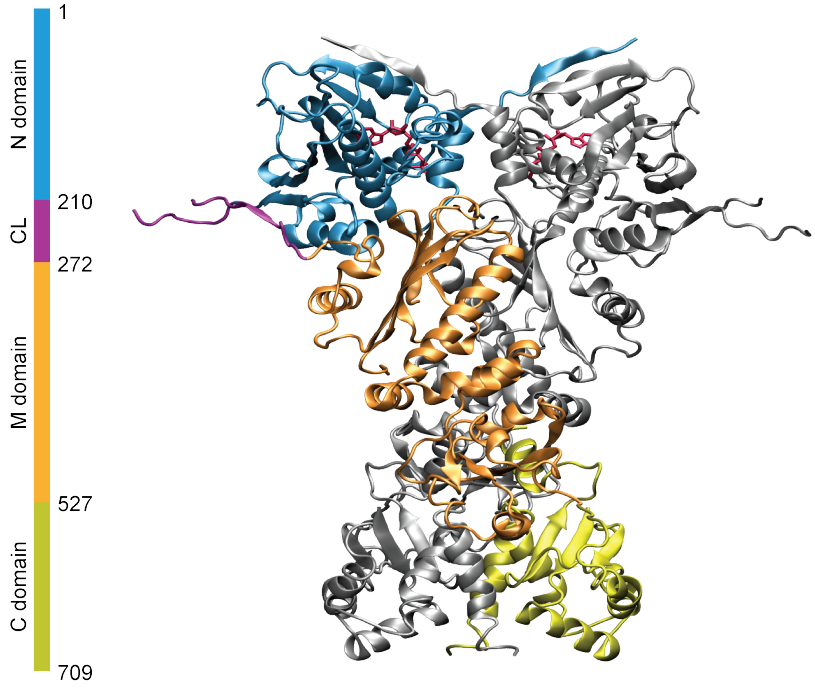

Figure 2.1: Crystal structure of the yeast Hsp90 homodimer. Hsp90's three domains are colored and domain sizes are indicated by the bar. The second identical protomer is shown in gray. The N and M domains are connected by an extended region, the so called Charged Linker (CL), which was partly deleted and is not fully resolved in the crystal structure. The same applies for the last 32 aminoacids at the C-termini. This crystal structure shows Hsp90 in a closed, compact conformation. It was co-crystallized with the ATP analog (AMP-PNP) shown as red sticks and the co-chaperone Sba1/p23, which is omitted for clarity.

The M domain can be divided into three subdomains, two alpha-beta-alpha subdomains connected by a three-alpha-helix subdomain. The N-terminal alpha-beta-alpha subdomain of the M domain comprises an important arginine (residue 380 in yeast) that is essential for ATP hydrolysis. It points towards the ATP binding pocket stabilizing the kinked ATP thereby enabling hydrolysis.

Recently the direct stabilization of the ATP by the arginine has been questioned [22].

The C domain, the smallest Hsp90 domain, is of globular shape consisting of alpha helices and beta sheets. One pair of alpha helices (residues 638 to 677 in yeast) of each protomer form a four helix bundle, responsible for stable *dimerization* [46]. The unstructured C-terminal part (last 32 residues in yeast) carries the MEEVD motif that can bind tetratricopeptide repeats, an important binding site for co-chaperones (see Section 2.3). The unstructured portion, as well as the binding motif, are absent in prokaryotes.

Dimerization is studied in Chapter 10.

The domains are connected by *linkers* enabling conformational rearrangements of Hsp90 as discussed in the following section. Most notable is the highly-charged linker connecting the N and M domains that is significantly extended in eukaryotes (62 residues in yeast). Its properties will be examined closely in Chapter 9.

2.2 CONFORMATIONAL DYNAMICS

Hsp90 undergoes *large conformational changes* between N-terminally unbound states (open) and N-terminally dimerized states (closed). Several crystal structures were able to trap Hsp90 in these different conformations. The closed, compact conformation, where Hsp90 protomers entwine each other was already shown in Figure 2.1. An *E. coli* Hsp90 structure reveals an open conformation where the N termini are far apart [133]. The crystal structure of Grp94, the Hsp90 analog in the endoplasmic reticulum, shows an open conformation where the N domains are additionally rotated [28]. These diverse structures indicate that Hsp90 can adopt different conformations and is a highly flexible molecule.

Earlier studies suggest that Hsp90's conformations are strictly controlled by *nucleotide binding and hydrolysis*, which are required for chaperone function *in vivo* [97, 100]. In the nucleotide-free (apo) and ADP states it populates open conformations, while in the ATP state it adopts closed conformations [120]. The aforementioned remodeling of the ATP binding pocket is believed to drive these conformational changes by exposing different residues. The ATP turnover is slow, in yeast about one ATP per minute is hydrolyzed, human Hsp90 is even slower by a factor of ten [122]. The reasoning behind the slow ATPase rate is that hydrolysis can only occur in a closed state [112], likely the state depicted in Figure 2.1, where the arginine residues of the M domain catalyze hydrolysis. In this model the rate limiting step for ATP

hydrolysis is the conformational change into a hydrolysis-competent, closed state [122].

More recent studies, however, show that eukaryotic Hsp90 can explore open and closed states in all nucleotide conditions [87, 115]. The conformational changes are faster than the observed ATPase rate, implying that they are not tightly coupled to one another. The emerging picture is that Hsp90 can access and fluctuate between multiple open and closed states under all nucleotide conditions until it reaches a state in which it is committed to hydrolysis [42]. In eukaryotes, these conformational states are controlled by other proteins presented in the next section.

N-terminal dynamics are studied in Chapter 10.

2.3 REGULATION BY CO-CHAPERONES

In eukaryotes, another effective layer of regulation assists the Hsp90 machinery. More than 20 proteins, the so called co-chaperones, interact with the Hsp90 dimer. Many of them stabilize distinct conformations and/or recruit clients, as reviewed in [126]. They can be divided into two groups: co-chaperones that bind directly to one or more domains and co-chaperones that bind to the C-terminal MEEVD amino-acid motif of Hsp90 via a tetratricopeptide (TPR) domain.

No Hsp90 co-chaperones have been found in bacteria.

The key chaperones that drive Hsp90's ATPase cycle are Sti1/Hop, Aha1, Sba1/p23 and Cdc37 [135]. Sti1/Hop binds through a TPR domain and stabilizes the open conformation therefore inhibiting the ATPase activity. It recruits clients together with Hsp70 as described in the following section. Another important recruiting co-chaperone specific for kinases is Cdc37, which binds directly to the M domains and stabilizes an open conformation. The co-chaperone Aha1 greatly increases ATPase activity by promoting closure of Hsp90 and aligning the arginine residue in the M domain [86]. Structurally the N-terminal domain of Aha1 binds to the M domains of Hsp90, while the C-terminal domain interacts with the N domains [119]. Sba1/p23 recognizes and binds to the compact closed state of Hsp90 between both N domains, stabilizing this conformation and inhibiting Hsp90's ATPase activity [123, 2]. It competes with Aha1 for a similar binding site. Figure 2.2, modified from [72], presents a possible *ATPase cycle* involving co-chaperone interactions and a theoretical client.

Co-chaperone names are given for both yeast and human if different.

The large set of co-chaperones that can bind simultaneously to the Hsp90 machinery or compete for different interaction sites leads to numerous possible complexes that are important for cycle progression [71, 32]. Since Hsp90 consists of two identical protomers, it can bind in principal two nucleotides and two identical co-chaperones. But most co-chaperones bind in a 1 to 2 *stochiometry*, leading to asymmetric complexes. This asymmetry seems to play a critical role in ATP hydrolysis and client binding [89, 68].

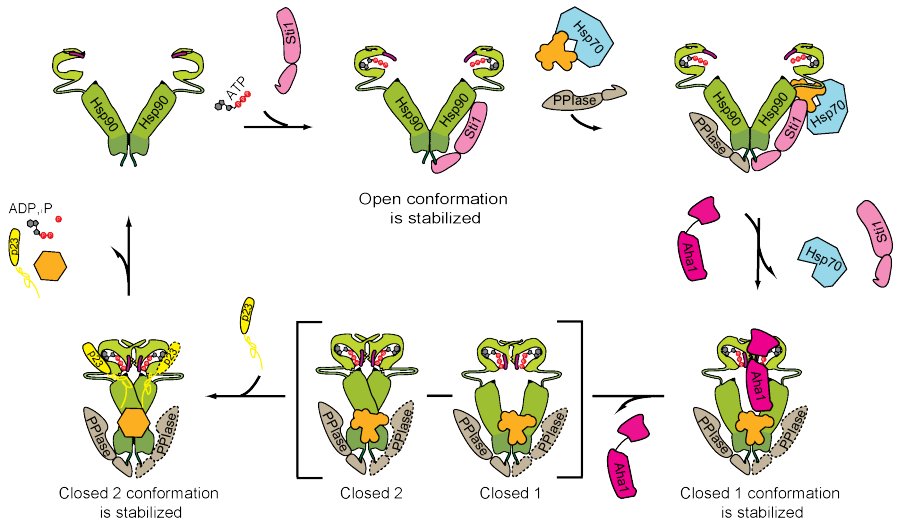

Figure 2.2: A possible ATPase cycle of Hsp90 deduced from multiple experiments involving different co-chaperones and a theoretical client [72]. Starting in the idle open conformation (left, top) Hsp90 binds ATP and Sti1/Hop. Sti1/Hop stabilizes the open conformation and recruits Hsp70 that has a potential client bound (orange blob). In the next step the client is transferred to Hsp90, Hsp70 and Sti1/Hop leave the complex and Aha1 promotes closure of Hsp90 (Closed 1). Then it adopts the closed compact state (Closed 2) that is necessary for ATP hydrolysis. This state is believed to be similar to the crystal structure shown in Figure 2.1 and is stabilized by Sba1/p23. It is assumed that at some point in this state client chaperoning occurs. After the release of co-chaperones and ADP, the machinery can start over again.

Another regulatory mechanism, recognized for Hsp90 and its co-chaperones in eukaryotes are *post-translational modifications*. Phosphorylation, acetylation and SUMOylation among others have been reported changing various features of the Hsp90 machinery (reviewed in [82]). Remarkably, phosphorylation and dephosphorylation has been shown to be sufficient to drive an Hsp90 chaperone cycle [167].

Prokaryotic Hsp90s lack co-chaperones and ATP binding and conformational changes are more tightly coupled [118]. It is assumed that the additional regulation of Hsp90 in eukaryotes is associated with an evolutionary gain of functions. This immediately implies that there are differences in the role of Hsp90 between organisms, although the aminoacid sequence is highly conserved [136].

2.4 CHAPERONING OF CLIENTS

Clients are also called substrates in the literature.

Little is known how Hsp90 interacts and chaperones its clients. The different clients lack structural similarity and no common binding motif has been identified. The only general feature is that clients show a propensity to be unstable proteins. It is a common notion that the broad range of states populated by the Hsp90 machinery is necessary to bind and chaperone its diverse clients. *In vivo* many clients depend on Hsp90, but *in vitro* reconstitution experiments are challenging, because of client instability and the *low affinity* of clients to isolated Hsp90. Whether the low affinities are a generic feature and other proteins are needed to keep Hsp90 in proximity, or Hsp90's binding conformations are not fully achieved in these experiments, is unknown. Recent NMR and hydrogen exchange experiments show that specific clients have distinct binding sites through different residues, often spanning multiple domains, as reviewed in [126, 82].

A patch of the M domain next to the cleft spanned by the Hsp90 dimer seems to be primarily involved in client binding.

Some clients are recruited actively via *different pathways*. One major pathway, also shown in Figure 2.2, is mediated by the co-chaperone Sti1/Hop and Hsp70. At first the misfolded or unfolded client is chaperoned by Hsp70. Then the adapter protein Sti1/Hop binds Hsp90 and Hsp70 via multiple TPR domains. At the same time Sti1/Hop promotes an open Hsp90 state that is competent for client binding. This fits into the picture that Hsp90 works in the late folding process [58]. That this cooperation of Hsp90 and Hsp70 is beneficial has been confirmed *in vitro* reactivating the artificial substrate luciferase [38]. Another recruiting pathway specific for kinases is via the co-chaperone Cdc37 [144].

One example of a client that is strictly dependent on Hsp90's function *in vivo* is the glucocorticoid receptor (GR). A reconstitution experiment investigating the hormone-bound ligand-binding domain of GR shows that the GR binds to a partially closed state of Hsp90. The affinity to the GR can be tuned by adding co-chaperones and the overall ATPase rate is reduced in respect to the client-free system [74].

Another GR study shows how Hsp90 and Hsp70 regulate GR activity by changing its stability [61].

While eukaryotic Hsp90s have many clients only very few proteins that interact with prokaryotic Hsp90s have been identified. Therefore client studies with *E. coli* Hsp90 exclusively used a partially folded model protein called Δ131Δ, a fragment of staphylococcal nuclease. This artificial client interacts mainly with two amphiphilic helices of the C domains situated in a cleft spanned by the M domains [46]. Here client binding induces conformational changes and ATPase activity is stimulated [140]. Interestingly, mutations of residues involved in client binding in *E. coli.* reduce the chaperone function of yeast Hsp90 *in vivo* and *in vitro* [39].

The mechanism by which Hsp90 activates and matures clients is far from understood. Already the two examples given above show differences suggesting that multiple chaperoning pathways are possible. How the ATPase cycle couples to the chaperone cycle and for what reason ATP is consumed, remain enigmatic.

2.5 CLINICAL RELEVANCE

Many clients of Hsp90 are involved in diseases making Hsp90 an attractive drug target. Most notable are the developments for cancer treatment that are already in clinical trials. Cancer cells often show highly elevated expression levels of Hsp90s. This is related to the stressed, abnormal environment in a cancer cell. Many oncogenic proteins that are important for proliferation of cancer cells are chaperoned by Hsp90. Prominent examples are the tyrosine kinase Src and the tumor suppressor protein p53, as reviewed in [160]. Since oncoproteins are often mutated and therefore more dependent on chaperoning, which is especially true in cancer cells, deactivating the Hsp90 machinery is more critical for cancer cells than for healthy cells. Specific inhibition of Hsp90 can be achieved by compounds binding competitively to the ATP binding pocket of Hsp90. Inhibition leads preferentially to ubiquitination of the client followed by proteasomal degradation. Many competitive inhibitors that deplete oncogenic protein levels are already in clinical trials and novel inhibitors with different targeting strategies are planned [147, 102].

Additionally, in neurogenerative diseases, like Alzheimer's and Parkinson's, Hsp90 maintains neuronal proteins of aberrant capacity, as reviewed in [76]. Hsp90 is also involved in viral replication and targeting Hsp90 together with antivirals is believed to prevent development of drug resistance, as reviewed in [172].

3

SINGLE MOLECULE EXPERIMENTS

Biomolecular systems are usually dynamic and complex, as presented in Chapters 1 and 2. To study their behavior at the nanometer scale, typically spectroscopic experiments using purified protein components are performed. In bulk assays millions of billions molecules are observed simultaneously. *Averaging* the measurement signals of the individual molecules gives a good signal-to-noise ratio, at the expense of losing information about the individual molecules. This ensemble averaging can lead to misinterpretations of experiments if the system investigated has multimodal state distributions. A simple illustration: a distance in a molecular system is measured. The system equally populates two states, one with a short distance and one with a long distance. Without further interpretation, the bulk experiment would reveal an intermediate distance that never occurs naturally.

Extracting kinetic information from bulk experiments can be tricky as well. To measure the transition rate from one state into another, the system first has to be prepared in one state, then the conditions must be changed in order to drive it to the other state. This relaxation can be followed over time and allows the determination of the rate. In protein folding experiments synchronization is typically achieved by changing from a buffer containing denaturant to a buffer without denaturant. In protein interaction experiments, ligands or other binding partners are added. In these cases, the exact rates can only be determined if the backward reaction is negligible. This shows that even for a simple two-state system, the determination of rates is not straightforward.

Single molecule (sm) experiments are powerful tools that can overcome these limitations. Following a single molecule trajectory allows the identification of different populated states and their lifetimes, especially rare or short lived states, as well as different states that return a similar measurement signal. The direct observation also allows the identification of the connectivity of the states involved in a system, showing possible and impossible pathways without any need for synchronization.

Depending on the goal of an experiment, different single molecule methods are at hand, mainly classifiable in two groups, *fluorescence- and manipulation-based methods*.

In sm fluorescence experiments, fluorophores are attached to the system. After laser excitation the fluorescent light is collected at a detector. Very sensitive cameras allow the localization of individual fluorophores and therefore sm particle tracking both *in vitro* and *in*

vivo [104]. A more detailed method used in this thesis is the single molecule Förster resonance energy transfer (smFRET). FRET is the radiation-free energy transfer between two fluorophores, termed donor and acceptor. The efficiency of the energy transfer can be measured by the fluorescence signals of the fluorophores, which is very sensitive to their distance making it a very useful tool to study conformational changes within proteins or protein-protein interactions. FRET experiments are either performed in a confocal or in a total internal reflection fluorescence (TIRF) setup. The first method measures freely diffusing molecules and provides detailed information about the population of different states although with limited kinetic information [70]. The latter allows observation of time trajectories of immobilized molecules [128].

For more details on FRET see Section B.3.

In manipulation-based experiments a single biomolecule is tethered via defined anchor points between two surfaces. By moving one of the surfaces force can be directly applied by stretching the molecule. The main experimental approaches are atomic force microscopy (AFM), optical tweezers and magnetic tweezers, which vary in their features and limitations, as reviewed in [93]. In AFM experiments, molecules can be manipulated using a soft cantilever with a sharp tip. The AFM was originally developed as a scanning probe microscope and is suited for assays at high forces ($10 - 10000\,\mathrm{pN}$). In optical tweezers experiments, force is applied via moveable beads held by laser traps. Optical tweezers are capable of measuring very low forces ($0.1 - 100\,\mathrm{pN}$) and they will be discussed in detail in Chapter 5. Another bead-based assay with superparamagnetic beads held by a magnetic field are magnetic tweezers experiments. Their unique feature is the possibility to apply and measure torque at the single molecule level.

In general, smFRET experiments are more powerful at acquiring rates and populations at equilibrium. The biggest limitations are the limited lifetimes and the sensitivity to the local environment of the fluorophores. Mechanical experiments are more reproducible and the active manipulation allows the population of states that are not accessible without external perturbation. A disadvantage is that the direct observation of a single molecule is only possible if force is applied. To extract information about the force-free parameters, that one is often interested in, models are necessary.

Although single molecule experiments are more demanding, their capability of revealing hidden and complex multi-state processes has greatly contributed to the understanding of molecular processes. They have been successfully used to characterize DNA, RNA, protein folding and unfolding, protein-protein interaction and molecular motors, as reviewed in [25].

4

OUTLINE

- Chapters 5 and 6 introduce the optical trap and give the necessary tools and developments to evaluate the measured data.
- Chapter 7 analyzes the unfolding pattern and the structure of yeast Hsp90 using optical tweezers.
- Chapter 8 gives detailed information how large proteins, such as yeast Hsp90, fold and what problems they encounter.
- Chapter 9 shows how linkers between domains can modulate intra-monomer conformations affecting the whole system using many different experimental approaches.
- Chapter 10 gives insights into the kinetics of the C-terminal dimerization and the nucleotide-dependent N-terminal dynamics using artificial dimers.
- Chapter D compares the *E.coli* to the yeast variant, highlighting important differences and similarities.

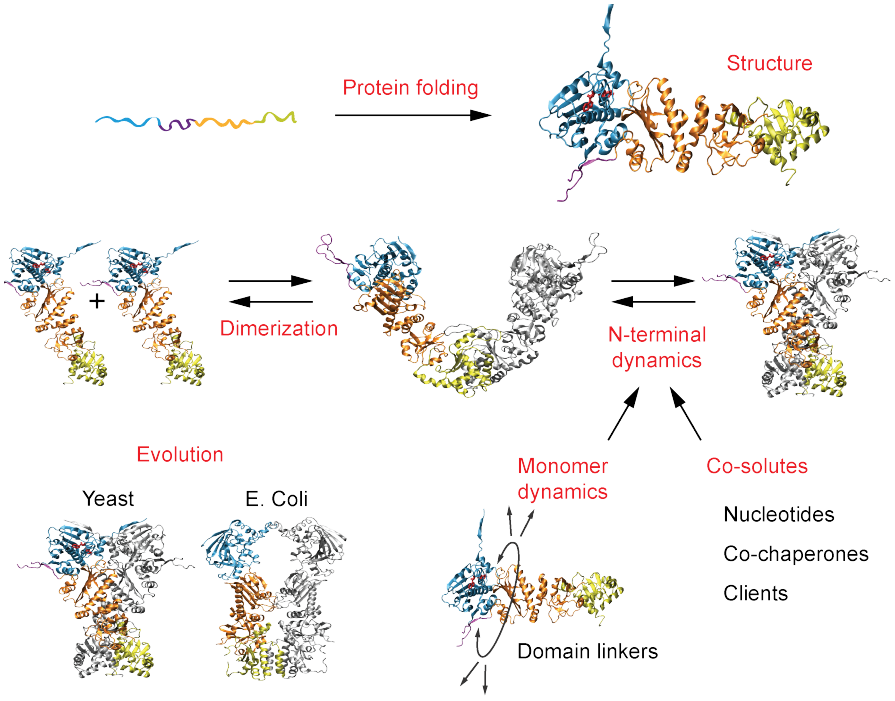

Figure 4.1: Graphical outline. Crystal structures are taken from [2, 133].

Part II

METHODS AND THEORY

In this part I explain the principles of how proteins can be manipulated and investigated using optical tweezers and how the physical properties of the investigated molecules can be determined by different theoretical approaches. I especially concentrate on the investigation of large proteins and developments made by myself.

5 OPTICAL TWEEZERS

Small solid objects can be spatially trapped by light. This was first shown by Ashkins in 1970, by capturing micron-sized latex beads using two opposing laser beams [4]. Several years later Ashkins and colleagues achieved trapping with a highly-focused, single-beam optical trap [6], laying the groundwork for modern optical tweezers experiments. Initial experiments on biological samples like viruses, bacteria and cell organelles were performed by directly capturing the sample inside the laser spot [5, 7]. Nowadays using micro-sized beads held by optical traps as force probes allows quantitative measurements of the force response of biological systems, such as the elastic properties of biopolymers like DNA [154], the folding [17, 139] and the ligand binding [137] of proteins. Of special interest are biological systems that generate forces *in vivo*, such as molecular motors [64, 16] or proteins that are naturally under force, such as force sensors [125].

5.1 OPTICAL TWEEZERS PRINCIPLES AND SETUP

A dielectric particle trapped in a focused laser spot experiences two predominant forces [143]. A *scattering force* acting in the laser beam direction and a *gradient force* pointing towards highest laser intensity. This gradient force is used for trapping and can be explained qualitatively by considering the momentum change of refracted beams by a transparent sphere using ray optics, as illustrated in Figure 5.1.

In fact the ray optics picture is only valid for particles whose radius is much larger than the wavelength of the laser. In the other limit, the Rayleigh regime, where the radius of the particle is much smaller than the wavelength, the particle can be considered as a single dipole in an electric field gradient. In this case the Lorentz force also drives the particle to the trap focus. In typical optical tweezers experiments the laser wavelength is similar to the particle size and calculation of the forces involved is difficult but can be achieved using generalized Lorenz-Mie theory [94].

To realize the traps and measure forces, a custom-built, high-resolution *dual-trap optical tweezers setup* with back-focal plane detection, designed by B. Pelz [152], is used. In this setup, light from the TEMoo mode of a Nd:YVO_4 solid state laser is split into its two polarization components, forming two beams. One of these beams can be steered using a piezo mirror stage. The beams are recombined and expanded to overfill the back aperture of the high-numerical-aperture trapping objective. Behind the specimen plane, the forward scattered light is

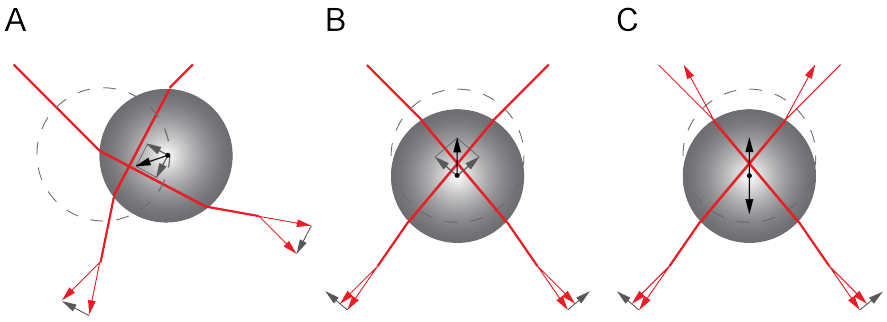

Figure 5.1: Optical trapping explained by ray optics. (A) A sphere is displaced from the trap center (dashed circle). If the sphere has a higher refection index than the surrounding medium, incoming rays are refracted in the same direction as the displacement, leading to a restoring force (black arrow) due to conservation of momentum. (B) The same restoring force arises if the sphere is displaced along the laser direction due to focusing or defocusing of the refracted light. (C) For a steep light gradient the scatter forces are lower than the gradient forces and the sphere can be stably-trapped shortly behind the laser focus.

collected by an identical objective, split into its polarization directions and monitored on two quadrant photodiodes.

In single-trap setups, one of the beads is held by a micropipette or one end of the sample is directly linked to the surface. In a dual-beam setup, the sample is clamped between two beads. Forces are applied by moving one bead and measured by the differential signal of both beads. The main advantage of this design is a reduction of noise and drift by mechanically decoupling the sample from the rest of the setup [90]. Additionally, external perturbations likely affect both traps similarly further reducing drift.

In this study spherical silica beads with a diameter of 1 µm and a refraction index of 1.47 are used. For small bead displacements from the trap center restoring forces are linear, like that of a Hookian spring. Prior to the measurements, trap stiffnesses and detector sensitivities are calibrated using a method that manages without assumptions on bead size and shape [146]. Dependent on the chosen laser intensity, trap stiffnesses of $0.2 - 0.4\,\mathrm{pN/nm}$ are used. Data are acquired at a sampling frequency of 100/150 kHz and downsampled to 20/30 kHz for analysis.

Only for certain corrections further assumptions about the system are necessary.

5.2 FORCE EXERTION ON A PROTEIN IN THE DUMBBELL ASSAY

To exert force on a protein in the dual-trap optical tweezers setup a symmetric sample with a shape similar to a dumbbell is constructed. Figure 5.2 illustrates this *bead-DNA-protein-DNA-bead design*. The sample assembly involves multiple stages: in a first step cysteine residues are genetically inserted into the protein (for details see Section B.1.1). They serve as defined attachment points for force application. The thiol groups of the cysteine residues are coupled to *maleimide-modified DNA oligonucleotides*. The nucleophile addition between the thiol and the maleimide groups is a faster and an easier-to-use reaction than the traditionally used disulphide bond formation (see Section C.1). Approximately 181 nm-long (545 basepairs) DNA handles are synthesized by PCR with a single-stranded overhang at one side, which is complementary to the DNA oligonucleotides attached to the protein, allowing stable hybridization. At the opposite side, the DNA handles have three biotin or three digoxigenin modifications. Amino-terminated silica beads coated via EDC/NHS chemistry with BSA and either streptavidin or anti-digoxigenin can bind to the biotin- or the digoxigenin-modified DNA handles, respectively. This yields a *symmetric dumbbell*, where the laser trapped beads are well separated from the protein by long DNA handles. The protocols used are detailed in the Section B.2.1.

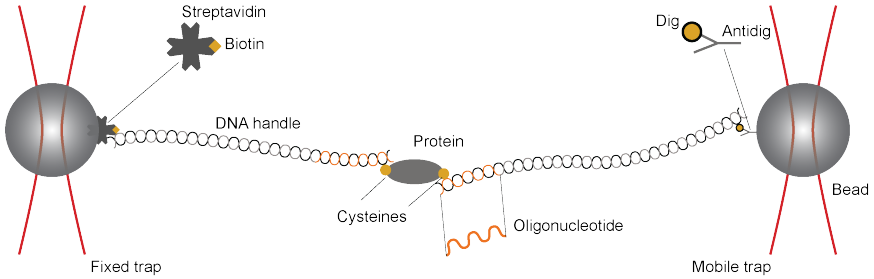

Figure 5.2: To exert force on a protein, DNA handles are attached at specific protein residues. The other ends of the handles bind via their biotin or digoxigenin modifications to streptavidin-coated (left) or anti-digoxigenin-coated (right) beads, which are held in laser traps.

It is also possible to synthesize the DNA handles as one piece, without using oligonucleotides and the single-stranded overhang [18], but the two-step protocol introduced by Gebhardt [139] and the use of maleimide oligonucleotides increase reaction efficiency (see Section C.1) and therefore experimental yield.

5.3 EXPERIMENTAL DRIVING PATTERNS

By means of optical tweezers proteins can be manipulated by changing the distance between the traps thereby applying different forces. Since the time course of the trap distance can be chosen almost freely, *different driving patterns* can be employed that are adapted to the problem investigated. Often repetitive patterns are used to investigate a property of the protein system multiple times and gather statistics. The driving patterns used in this thesis are summarized in Figure 5.3 and their applicability is described in the following.

These repetitive patterns are often called cycles.

In *constant velocity* cycles the trap distance is increased and decreased with a defined velocity, allowing proteins to be driven repeatedly from the fully folded state to the fully unfolded state and back. Force-extension traces are calculated that show all unfolding and refolding features of the protein over a large force range. These traces also reveal the sizes of the structural elements and their mechanical stabilities.

The native unfolding pattern is also called the fingerprint of a protein.

To gain insight into energetic and kinetic parameters, *constant distance* or passive mode experiments are the first choice. In these assays the trap distance is held constant and the force response is investigated. This approach is feasible if the system is populating and depopulating different states at equilibrium on a measurable timescale. It allows the full characterization of the energetics and kinetics of the system by the measurement of the force-dependent populations and lifetimes.

If the protein doesn't access all states of interest in constant distance experiments, *jump experiments* can help to increase transition rates, by lowering or increasing the force rapidly. Employing the jumps repeatedly makes it possible to collect force-dependent lifetimes of the involved states.

Jump experiments are also very useful to investigate proteins at low forces, where the force signal is lost.

For some questions it can be useful to combine jumps with ramps, which provides the benefits of both patterns like in *double-jump ramp experiments*. They will be motivated and discussed in detail in Section 8.4.

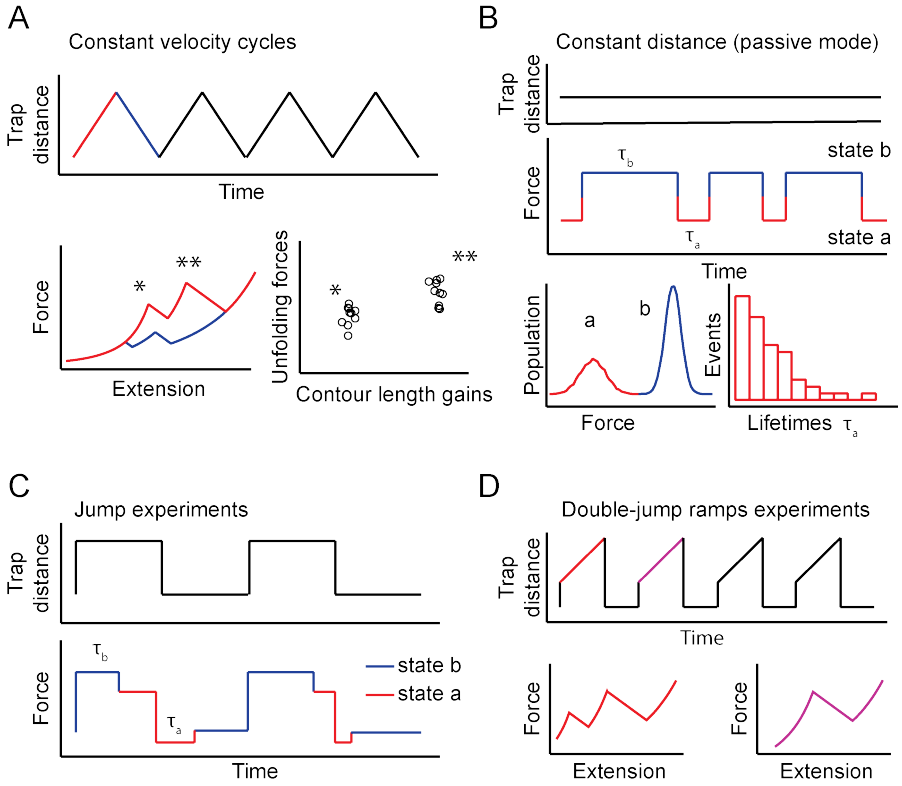

Figure 5.3: Main force exertion patterns employed in optical tweezers experiments: constant velocity (A), constant distance (B), jump (C) and double-jump ramp (D) patterns.
(A) In constant velocity experiments many unfolding (red) and refolding (blue) force-extension traces are obtained and thereof the stability and size of the unfolding (*,**) and refolding events, which are typically collected in scatter plots. Here an example of a protein with two structural elements is shown.
(B) Keeping the trap distance constant allows the force fluctuations between different states to be followed over time and their population probabilities and lifetimes to be extracted, as shown here for a two-state system (with states a and b).
(C) If the folding and the unfolding rates are too slow at equilibrium, jump assays enable the exploration of different force regimes, where rates are faster and lifetimes can be determined.
(D) In double-jump ramp experiments, an example of a more complex pattern, constant velocity unfolding traces probe the different states of the system after spending a defined time at low force.

6 DATA ANALYSIS

In this chapter, the necessary tools to evaluate the energetic and kinetic properties of protein systems measured in optical tweezers are provided and the underlying theories are discussed.

6.1 ELASTIC BEHAVIOR OF BIOPOLYMERS

Flexible polymers favorably adopt random coil conformations if equilibrated. Increasing the end-to-end distance of such polymers typically reduces the conformational entropy and results in a restoring force. The understanding of how such polymers respond to force is crucial for the correct interpretation of the optical tweezers experiments.

A model, called the *worm-like chain* (WLC) model, is successful in describing semi-flexible polymers. In contrast to other models that consider the polymer as a chain of stiff monomers that can rotate against each other, the WLC model describes the polymer as a continuous and isotropic rod that is bent by thermal energy. The stiffness or flexibility of the rod is defined by a material parameter called the persistence length, which is the decay length along the contour of the rod after which the tangent vector direction correlation drops below $1/e$. In other words two points that are separated further than the persistence length are uncorrelated.

It is possible to calculate the mean squared end-to-end distance (extension) of the WLC polymer, but no analytical expression of how force applied to the ends changes the end-to-end distance can be given. However, calculating the linear entropic behavior at the low force limit as well as the high force limit is feasible. A frequently used analytical expression derived by Marko and Siggia [78] that interpolates between these two limits, can describe the *force-extension* $F(x)$ behavior of an unfolded protein chain well. Using the contour length l_{prot} and the persistence length p_{prot} of the unfolded protein chain as well as the thermal energy k_BT, the interpolation formula can be written as:

Also single-stranded DNA or RNA are typically described by WLCs.

$$F_{WLC}(x) = \frac{k_BT}{p_{prot}} \left(\frac{1}{4} \left(1 - \frac{x}{l_{prot}} \right)^{-2} - \frac{1}{4} + \frac{x}{l_{prot}} \right) \tag{6.1}$$

At intermediate forces around k_BT/p_{prot}, the interpolation model overestimates the force by up to 10% compared to the numerical solution [154, 79]. Typical average persistence lengths for unfolded protein chains are around 0.3 to 0.75 nm, slightly higher than the bond

This is around 5 pN for proteins at room temperature.

length of an individual aminoacid residue. However, interactions of different neighboring aminoacids can also modulate the persistence length [19, 69], meaning that an aminoacid chain with an atypical average residue composition or extreme buffer conditions might show a different persistence length.

The WLC model characterizes unfolded protein chains well, but for double-stranded DNA the model fails at higher forces. The reason behind this is the double helix nature of the DNA. At high forces the double helix is distorted, leading to additional elastic contributions.

At forces around 70 pN DNA transitions into an overstretched state.

An extension of the WLC model, called the *extensible worm like chain model* (eWLC), can take this behavior into account by introducing an elastic stretch modulus K. The modified interpolation formula reads as [154]:

$$F_{eWLC}(x) = \frac{k_B T}{p_{DNA}} \left(\frac{1}{4} \left(1 - \frac{x}{l_{DNA}} + \frac{F}{K} \right)^{-2} - \frac{1}{4} + \frac{x}{l_{DNA}} - \frac{F}{K} \right) \quad (6.2)$$

B-DNA is the most common conformation of double-stranded DNA at physiological conditions.

The literature values of the persistence length p_{DNA} and the stretch modulus K for B-DNA are around 50 nm and 1100 pN, respectively [154]. In the dumbbell assay used (see Section 5.2), lower values for p_{DNA} and K are observed. There are multiple causes that explain the reduced stiffness: First, the finite DNA length, the free DNA ends, as well as rotational fluctuations of the beads attached to the DNA decrease the persistence length significantly, as shown in simulations [131]. Second, because of the biochemical nature of the dumbbell assay, there are nicks in the DNA handles and just one strand is attached to the beads, which further decreases DNA stiffness. Finally, ignoring unstructured regions in the protein construct can render the DNA parameters inaccurate, as shown in the following section.

6.1.1 *Fitting WLC models to complex proteins*

Folding, unfolding or other conformational transitions of proteins under force typically change the proteins' end-to-end distance. To evaluate these protein transitions and their contour lengths changes WLC models are fitted to force-extension traces (see Section 5.3). In addition they allow the determination of all DNA parameters that are necessary for further analysis.

The contour length refers to the contour length of the unfolded aminoacids.

Figure 6.1 A shows typical force-extension traces expected for a protein with two structural units. The first rising slope of the unfolding trace is due to stretching of the DNA handles alone. An eWLC fit (Equation 6.2) to this slope enables the determination of the DNA parameters (p_{DNA}, l_{DNA} and K) of the tether. The contour length gains of the following protein events are then evaluated one after the other by fitting WLCs (Equation 6.1) in series with the previously determined eWLC.

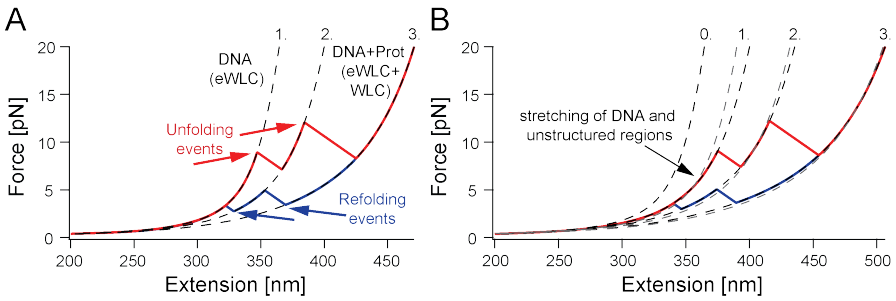

Figure 6.1: WLC fits (dashed lines) to unfolding (blue) and refolding (red) force-extension traces of a fictional two domain protein in the dumbbell assay.
(A) The first fit (1.) to the completely folded protein describes the stretching of pure DNA (eWLC). The second and third fit (2. and 3.) are WLCs in series with the previously determined eWLC, which yield the contour length gains due to protein unfolding.
(B) For a protein with unstructured regions, the first slope contains already unfolded peptide. In this case globally fitting all slopes allows the determination of all the DNA and protein parameters simultaneously (black WLC fits) as well as the consideration of unstructured regions (0.). If the unfolded regions are not accounted for and fitted as in Fig. A, poor fits and incorrect DNA and protein parameters are obtained (gray WLC fits).

Already small folded proteins may have small unstructured ends.

Large multi-domain proteins often contain *unstructured regions* and linkers connecting the domains. If the regions have significant length, the approach illustrated above is not practicable, because the first rising slope of the force-extension trace already contains stretched unstructured protein. For that reason a *global fit routine* that allows the simultaneous fitting of all slopes in a multi-domain unfolding pattern is implemented.

It is possible to include the unstructured parts only in the first fit at the expense of larger errors.

This provides two advantages: The first is that all slopes of the force-extension trace contribute to the evaluation of the DNA parameters, therefore increasing their accuracies. The second is that an additional contour length, which accounts for the unstructured parts can be easily considered as shown in Figure 6.1 B. Neglecting unstructured parts leads to an apparent lower DNA persistence length and longer DNA and protein contour length gains. These incorrect parameters may lead to misinterpretation of the involved structure and errors in the evaluation of energetic and kinetic properties, which will become obvious in the next sections.

In the example shown in Figure 6.1 B, ignoring 50 nm *of unstructured residues leads to a 50% error in DNA persistence length and a 10% error in protein contour length.*

The force-extension traces in this thesis are fitted explicitly using the global fit method described. The protein persistence length is fixed to 0.7 nm, which gives the best results for the proteins investigated. Expected average DNA persistence lengths of around 20 nm and DNA contour lengths of around 370 nm are found. For fitting, the elastic stretch modulus is fixed to values between 300 to 700 pN. To include unstructured regions typically a fixed contour length is

assumed, which is estimated from the structure or from traces of misfolded protein molecules. It is also important to note, that most reliable contour length gains between two WLCs are obtained if similar force regions are fitted. This reduces the errors dramatically since incorrect DNA parameters and the deviations by the WLC interpolation formula affect the fits similarly.

6.1.2 *Relating contour length gains to structure*

The contour length gains measured for unfolding or refolding events have to be related to the structural parts that unfold or refold. Calculating the number of aminoacids involved provides a *structural interpretation*, which is especially useful for identifying the native pattern of the protein as well as interpreting non-trivial events, such as the structure of intermediates. However, the calculation of the number of aminoacids from the observed contour length gain requires detailed knowledge of the end-to-end distance of the structure in its folded state. This information can be obtained from protein structures with atomic resolution, which allow the determination of the initial distance in the folded state $d_{ij,fold}$ between two residues i and j. Subtracting the initial distance from the average distance per aminoacid l_{aa} times the total number of aminoacids unfolded $n_{aa} = j - i$ yields the expected contour length gain l_{ij}:

These protein structures are typically obtained by NMR or X-ray scattering on protein crystals.

$$l_{ij} = n_{aa} l_{aa} - d_{ij,fold} \tag{6.3}$$

Comparing the theoretical with the measured contour length gains enables the validation or falsification of structural models. If multiple models fit the data, truncation mutants or different pulling directions help to identify the right one.

The exact value for the average contour length gain per aminoacid has to be chosen carefully, since the contour length gains obtained from WLC fits depend on the protein persistence length used. Therefore it is reasonable to determine the average contour length gain per aminoacid for a certain persistence length, from contour length gains with a clear structural interpretation. For example, from the contour length gain from the folded to the completely unfolded state. In this thesis, the length per aminoacid is assumed as 0.365 nm for a protein persistence length of 0.7 nm.

A consistency check for multi-domain proteins is comparing the size of the individual domains to the overall size.

6.2 STATE ASSIGNMENT WITH HIDDEN MARKOV MODELS

Trajectories of a single protein that populates multiple distinct states in equilibrium, for example folded and unfolded, can be recorded in constant distance experiments (see Section 5.3). These different states are typically separated by contour length and/or kinetics. To obtain

thermodynamic and kinetic information of the protein system, the individual points of the trajectory have to be assigned to the states observed. Especially at low forces, the amplitude of the thermal noise is much larger than the signal change and a simple threshold criterion to *discriminate the states* can not be employed without strongly smoothing and therefore impairing the data.

A powerful and elegant approach to assign the observed states to an unfiltered trajectory are *Hidden Markov Models* (HMM). They use probabilistic arguments to associate the individual data points of the trajectory $(o_1, o_2, ..., o_T)$ to hidden states x_i. The hidden states and their transition probabilities are defined by a Markov process and each hidden state has a certain emission probability to create a certain observable value.

After initialization, a *forward-backward algorithm* allows the determination of the most probable state of the hidden network for each point of the trajectory. It does so by recursively calculating the forward variable $\alpha_t(i, o_1...o_t)$, which tells the probability to be in a certain state i at data point t after having observed the measured sequence of observables $(o_1, ..., o_t)$. Similarly the backwards variable $\beta_t(i, o_t...o_T)$ starting from the last point of the trajectory is calculated. Using them together with Bayes' rule allows the determination of the probabilities P to be in the hidden states x_i at a certain time point t for the totality of all measured observables.

$$P(x_i|o_1, ..., o_T) \propto \alpha_t(i, o_1...o_t)\beta_t(i, o_t...o_T) \tag{6.4}$$

A maximum likelihood estimate can be implemented for the HMM by the *Baum-Welch algorithm*, which improves the values by maximization of the individual observation probabilities. The initial values for the transition and the emission probabilities can therefore be optimized by running the forward-backward and Baum-Welch algorithm iteratively.

For more details on the algorithms see [114].

To evaluate constant distance data, the traces are discretized into 50 to 200 emission force bins. The number of states and their initial emission probabilities are determined by fitting Gaussians to the force histogram of the evaluated trace. Transition probabilities between the hidden states are set to the timescale of the observed kinetics. The emission probabilities are then optimized by repeatedly running the forward-backward and Baum-Welch algorithms whilst keeping the transition probabilities fixed. After convergence, each data point is assigned to the state with the maximum observation probability. After assignment the force-dependent populations and lifetimes of each state can be extracted easily (see also Section 5.3).

To consider the given transition probabilities for the overall trajectory, the Verterbi algorithm can be used.

6.3 ENERGY CONTRIBUTIONS OF THE SYSTEM

If a protein changes its state, the free energy of the system is altered due to additional or lost enthalpic or enthropic contributions. Examples are the protein folding free energy, which is the free energy difference between the unfolded protein state and the folded one, or the binding free energy for a ligand-bound or ligand-free protein. Figure 6.2 A illustrates the energy difference of a simple two-state system.

In the dumbbell assay (see Section 5.2), the protein is examined by measuring the force via the displacement of the beads from their trap centers. To evaluate the energetic and kinetic properties of the protein states the consideration of all dumbbell components is important. Relaxation times of all components are fast, therefore the symmetric dumbbell system can be modeled as a single bead-DNA-protein system, with an effective trap stiffness and a single DNA handle with twice the length. The *total free energy of the dumbbell system* G_i at a force F is the sum of four components: First, the free energy of the protein E_i in the state i. Second, the free energy stored in the beads G^{beads}. Third, the free energy of the stretched DNA, G^{DNA}, and finally, the free energy of stretched unfolded protein G_i^{prot}, which is dependent on the protein state i.

To distinguish clearly between the total energy and the energy of the protein state the symbol E instead of G is used for the latter.

$$G_i(F) = E_i + G^{beads}(F) + G^{DNA}(F) + G_i^{prot}(F) \tag{6.5}$$

Under force, the beads are displaced in the harmonic potential of the traps. Because the same force F acts on both traps the sum displacements of the individual beads is $x(F) = (k_f^{-1} + k_m^{-1})F = k_{eff}^{-1}F$ with k_f and k_m being the spring constants of the fixed and the mobile trap. The expression can be simplified using the effective spring constant k_{eff} and the free energy of the beads can be written as:

$$G^{beads}(F) = \frac{1}{2}k_{eff}^{-1}F^2 \tag{6.6}$$

The energy stored in the stretched DNA handles can be calculated by integrating Equation 6.2:

$$G^{DNA}(F) = \int_0^{x_{eWLC}(F)} F_{eWLC}(x')dx' \tag{6.7}$$

If the protein in state i is partially or fully unfolded, the free energy of the stretched unfolded peptide using Equation 6.1 is

$$G_i^{prot}(F) = \int_0^{x_{WLC}(F)} F_{WLC}(x')dx' \tag{6.8}$$

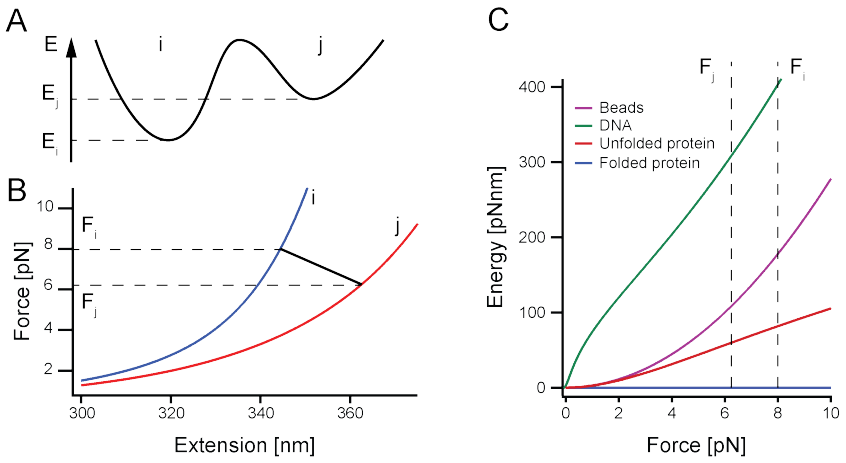

Figure 6.2: (A) A simple energy landscape with two states i and j (i.e. folded and unfolded) that are separated by the energy difference $E_j - E_i$ is shown. (B) Force-extension traces of an example protein with a folded state i and an unfolded state j. Assuming the protein unfolds at a force F_i the unfolded state will be at force F_j if the trap distance is held constant. (C) The energy contributions of the beads, the DNA and the unfolded protein dependent on force. If the protein transitions from state i at a force F_i to state j at a force F_j the energy contributions change. Together with the free energy $E_j - E_i$ of the protein the energy difference of the complete system is determined.

Figure 6.2 shows an example of the individual energy contributions for a simple protein with two states.

The energy difference of the complete system $G_{ij}(F_i, F_j) = G_j(F_j) - G_i(F_i)$ between two protein states i and j with corresponding forces F_i and F_j, that are separated by the free energy of E_{ij} can be expressed by the energy differences of the individual components:

At constant trap distance, the forces at different states (F_i and F_j) are related by k_{eff}, DNA and protein parameters.

$$G_{ij}(F_i, F_j) = E_{ij} + G_{ij}^{beads}(F_i, F_j) + G_{ij}^{DNA}(F_i, F_j) + G_{ij}^{prot}(F_i, F_j) \quad (6.9)$$

6.4 ANALYSIS OF ENERGIES FROM POPULATION PROBABILITIES

To determine the free energy between protein states E_{ij}, the experimentally accessible state populations are evaluated. At equilibrium the population ratio between two protein states $P_i(F_i)$ and $P_j(F_j)$ is related to the free energy difference of the complete dumbbell system (see Equation 6.9) by the *Boltzmann distribution*:

$$\frac{P_j(F_j)}{P_i(F_i)} = exp\left(-\frac{G_{ij}(F_i, F_j)}{k_B T}\right) \quad (6.10)$$

For multiple states the sum of all state probabilities is unity and the force-dependent population $P_i(F_i)$ can be expressed by summing up all Boltzmann factors:

$$P_i(F_i) = \frac{1}{1 + \sum_{i \neq j} \exp\left(-\frac{G_{ij}(F_i, F_j)}{k_B T}\right)} \tag{6.11}$$

Equilibrium state occupancies in this thesis are usually evaluated at fixed trap distances (see Section 5.3). This distance d_{trap} defines the force of each state $F_i(d_{trap})$. The free energy differences between the protein states E_{ij} are calculated from the state probabilities $P_i(F_i)$ by rearranging Equation 6.11 using the previously determined force-dependent energy contributions of the dumbbell. In principle, they can be calculated from one trajectory at one trap distance, but varying the trap distance enables the exploration of the force dependence of these probabilities. A global fit to all state probabilities at all measured forces using Equation 6.11 minimizes the errors of the energy calculation.

6.5 THEORY OF RATES

The timing of reactions is of great importance to understand the working mechanism of biological systems. Biological processes can often be thought of as a network of states with rates populating and depopulating these states. Two states are usually separated by an energy barrier, the height of it is also called *activation energy*, which must be overcome to transition from one state into the other (see also Figure 6.2 A).

Without an energy barrier, a rate description doesn't make sense.

The first description (for a first order reaction) of the relationship between the reaction rate k and the activation energy E_A was found by Arrhenius [3, 65].

$$k_{Arrhenius} = A e^{-\frac{E_A}{k_B T}} \tag{6.12}$$

This equation also describes the temperature dependence of reaction rates in a first approximation.

The preexponential factor A is called the attempt frequency and can be interpreted as the frequency of transition attempts, while the exponential factor discriminates the attempts that have enough energy to overcome the barrier.

A more elaborate model which relies more on physical principles, is the *transition state* theory (TST) [65]. The transition state is the peak of the energy barrier and TST assumes that the initial, ground state and the transition state are in quasi equilibrium. Once the transition state is reached, the system transitions to the final state. TST allows the quantitative description of the preexponential factor and can be extended to higher order reactions. The first order reaction rate from

The transition state is often called activated complex in case of chemical reactions.

a state i to j is, similarly to Arrhenius, defined by the free energy difference between the ground state i and the transition state T on the way to the final state j.

$$k_{TST,i\to j} = \frac{k_B T}{h} e^{-\frac{G_{iT}}{k_B T}} \tag{6.13}$$

Prefactors of similar scale can be obtained from collision theory.

The preexponential factor derived by TST is $\frac{k_B T}{h}$ with h being the Planck constant. It is valid for chemical reactions of small inorganic educts and is on the order of picoseconds. For *diffusive protein dynamics*, the prefactor is several orders of magnitude smaller. Typically an attempt frequency between 10^6 s and 10^9 s is assumed [67, 168, 107].

The approach that maybe best reflects the reality for transitions in proteins is the theory introduced by *Kramer* in 1940 [62, 45]. Here the transition rate is obtained by considering Langevin particles diffusing in a double well energy potential $U(x)$. Using the Fokker-Planck equation, the first passage times for a stationary particle flux from one dwell into the other are calculated. In the high friction (over-damped) limit the transition rate can be calculated considering the diffusion coefficient D of the particle:

$$\frac{1}{k_{i\to j}} = \frac{1}{D} \int_{-\inf}^{x_{max}} e^{\frac{U(x)}{k_B T}} \left[\int_{-\inf}^{x} e^{\frac{-U(x')}{k_B T}} dx' \right] dx \tag{6.14}$$

This expression is valid for any *energy landscape shape* $U(x)$. The first integrand mainly depends on the shape of the initial state while the second integrand mainly depends on the shape of the energy barrier. Expanding the potentials around the initial state and the transition state, integrating, and using the Einstein relation the Kramer's rate simplifies to [62]:

$$k_{i\to j} = \frac{\omega_i \omega_T}{2\pi\gamma} e^{-\frac{U(x_T)-U(x_i)}{k_B T}} = \frac{\omega_i \omega_T}{2\pi\gamma} e^{-\frac{G_{iT}}{k_B T}} \tag{6.15}$$

Although the formula is similar to the previously obtained results (see Equations 6.12 and 6.13), the new important features in Kramer's theory are that the preexponential factor depends on the shapes (curvatures) of the initial ω_i and the transition state ω_T and on the friction coefficient γ.

The friction dependence can be explored experimentally by changing the solvent viscosity [106].

The rate theories introduced above show which factors determine the speed of transitions. The form of the energy landscape is of great importance. With high-resolution single molecule experiments (see Chapter 3) it is possible to directly probe the energy landscape of a protein [163, 105]. Unfortunately, these experiments are typically limited to simple systems along one reaction coordinate and structural interpretation can be challenging [21]. Because protein chains

have many degrees of freedom and therefore high-dimensional energy landscapes, the projection of the whole energy landscape on one reaction coordinate may be an oversimplification, which should be kept in mind when analyzing protein transitions.

6.5.1 *Detailed balance*

For a system with two states i and j that are in equilibrium the rates are related to the energy difference G_{ij}.

$$\frac{k_{i\to j}}{k_{j\to i}} = e^{-\frac{G_{ij}}{k_B T}} \tag{6.16}$$

If a system is at equilibrium and a transition from state i to state j involves several different microscopic intermediate states, the reverse reaction has to populate the same intermediate states. This *microscopic reversibility* is called detailed balance. An important consequence is that the transition state of the forward rate $k_{i\to j}$ has to be on the same position for the backward rate $k_{j\to i}$.

6.6 RATES UNDER FORCE

In optical tweezers experiments the investigated systems are usually under load. Forces deform the energy landscape and as a consequence change the equilibrium populations and the transition rates between states. For biological systems the intrinsic rates of the unperturbed system, i.e. the rates at zero force, are often of interest. In this section models that describe the force dependence of rates and allow extrapolation to zero force are discussed.

6.6.1 *Bell-Evans model*

The first phenomenological description of force-dependent rates in a biological system was given by Bell in 1978 [11]. In this cell adhesion study the force-dependent receptor ligand dissociation between two cells is theoretically investigated. Later Evans *et al.* provided the theoretical framework and several extensions starting from Kramer's theory (see Section 6.5) [35].

In their model, an external force F perturbs the energy landscape $U(x)$ along the reaction coordinate x by a mechanical potential $-Fx$. This *tilts the energy landscape* as illustrated in Figure 6.3 A. In the Bell-Evans model, which can be seen as an expansion of TST (see Section 6.5), the transition state energy barrier is lowered by Fx_{iT} with F being the applied force and x_{iT} representing the distance between the

initial and the transition state. Therefore the force dependence of a transition rate is (compare to Equation 6.13):

$$k_{i\to j}(F) = Ae^{\frac{-(G_{iT}-Fx_{iT})}{k_BT}} = k_{i\to j}(F=0)e^{\frac{Fx_{iT}}{k_BT}} \tag{6.17}$$

Figure 6.3 shows the change of the energy barrier if force is applied for a dissociation (or an unfolding) event as well as the force-dependent rate in the Bell-Evans model. Measuring rates at different forces allows the extrapolation to zero force and the determination of the distance to the transition state.

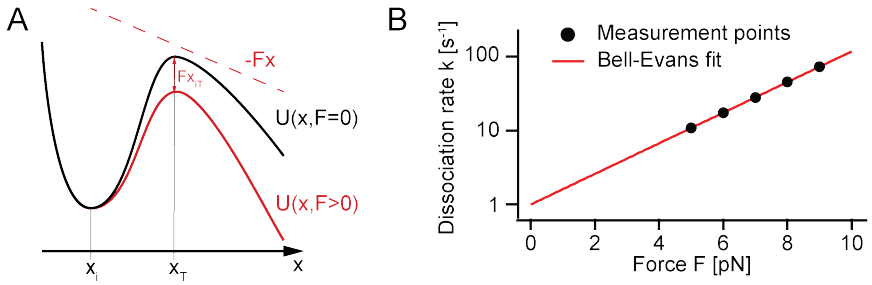

Figure 6.3: (A) The energy landscape for a dissociation process of a force-free system is shown in black. If force is applied to the system, the mechanical potential $-Fx$ (red dashed) tilts the energy landscape (red) lowering the energy barrier and favoring the transition. Here, the transition state position x_T is force independent. (B) In the Bell-Evans model the logarithms of the rates show a linear force-dependence (see Equation 6.17). Measuring the dissociation rates at different forces allows the determination of the rate at zero force and the distance to the transition state, which is defined by the slope.

This simple model has severe drawbacks. The distance to the transition state stays constant and doesn't change with force. If some protein state involves unfolded regions, this distance can be very force-dependent (see Section 6.1), which has to be considered (see also Section 6.6.3). Additionally, if the assay contains flexible tethers like DNA or moveable beads, such as the dumbbell system, the effective transition state distance of the whole system is measured rather than that of the protein system alone [33].

Same applies for AFM experiments, where the cantilever has to be considered [129].

These drawbacks lead to errors in extrapolating rates to zero force. Dissociation or unfolding rates are typically underestimated, while folding rates are overestimated. However, if the transition state is very close to the initial state, as is generally the case for dissociation or unfolding transitions, the Bell-Evans model can perform reasonably well.

6.6.2 *Dudko-Hummer-Szabo model*

A frequently used and more elaborate model is the one presented by Dudko, Hummer and Szabo [29]. Using Kramer's theory they derive a model that considers the shape of the potential in every point as well as the transition state position change due to the mechanical potential $-Fx$. For energy landscapes, that have a cusp-like or a linear-cubic form, analytical expressions for the force-dependent rates can be deduced.

The Bell-Evans model just depends on two parameters $k(F=0)$ and x_{iT} (see Equation 6.17). To account for the changes of the energy landscape the Dudko-Szabo-Hummer model additionally depends on the height of the energy barrier at zero force E_{iT} and a parameter accounting for the shape of the energy landscape. Therefore the measured rates have to be fitted with four parameters, which may lead to overfitting and inaccurate results. Nevertheless the models of the cusp-like and linear-cubic energy potential perform better than the Bell-Evans model and don't overestimate unfolding rates at zero force [29]. In a different publication they also provide an analytical correction term to consider the effect of flexible linkers [30].

6.6.3 *Berkemeier-Schlierf model*

The main model employed in this study is an energetic model to describe protein folding and unfolding rates developed in the Rief lab. It was introduced by Berkemeier and Schlierf for AFM experiments [129] and adapted by Gebhardt [37] for optical tweezers experiments. In this model, the force-dependent position of the transition state is described by a WLC separated from the initial state by a fixed contour length. This means that the actual transition state distance increases with force (see Figure 6.4 A) for both unfolding and refolding transition. The influence of force on the barrier height is calculated taking into account the energy difference of all components involved in the dumbbell between the initial state and the transition state $G_{iT}(F_i, F_T)$, with F_i and F_T being the forces at the initial state and the transition state, respectively.

F_T is related by the trap distance to F_i.

$$G_{iT}(F_i, F_T) = G_{iT}^{beads}(F_i, F_T) + G_{iT}^{DNA}(F_i, F_T) + G_{iT}^{prot}(F_i, F_T) \quad (6.18)$$

This expression is similar to the one described in Equation 6.9.

The individual contributions are explained in Section 6.3. The force-dependent rate constant is then:

$$k_{i\to j}(F) = k_{i\to j}(F=0)e^{-\frac{G_{iT}}{k_B T}} \quad (6.19)$$

Figures 6.4 B and C show an example of the force-induced energy increase or decrease of the transition state and the related force-

dependent rates for a folding and an unfolding transition of a simple two-state folder with typical system parameters.

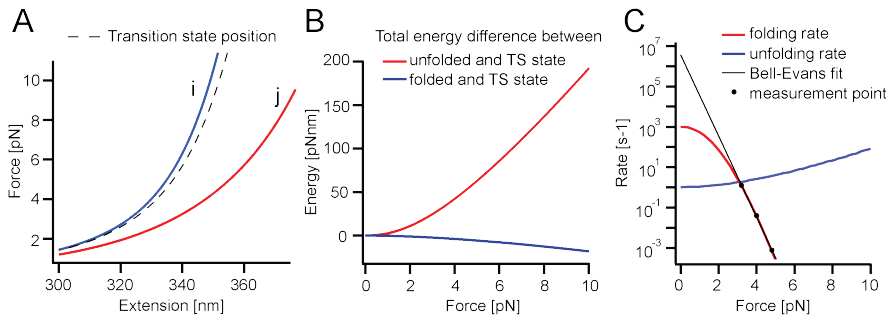

Figure 6.4: (A) The position of the transition state (TS) for a protein with a folded state i (blue) and an unfolded state j (red) is described by a WLC. Therefore the distance to the transition state depends on the force, where the transition happens. (B) The energy barrier is changed by the energy difference between the initial state and the transition state of the complete system. The force-dependent contributions expected for unfolding (blue) and refolding (red) are described by Equation 6.18. (C) The non-linearity of the energy differences shown in Fig. B lead to force-dependent unfolding and refolding rates that deviate from the exponential behavior. To compare this model to the Bell-Evans model, hypothetical measurement points are fitted with the Bell-Evans model. The inadequacy of the Bell-Evens model leads to large errors, especially for refolding rates.

In this approach, the *elasticity of the complete system* and the different distances of the transition state due to the force-dependent protein extension are fully considered. This leads to significant deviations from the Bell model (see Figure 6.4 C), especially if rates are measured under force and extrapolated to zero force.

Like the Bell-Evans model, this model only depends on two unknown parameters, namely the contour length of the transition state position x_T and the intrinsic rate $k(F=0)$. All other parameters that are needed to calculate the energy contributions such as the trap stiffness, DNA and protein parameters can be obtained beforehand from calibration and WLC fits. In this thesis exclusively this model is used to describe the force-dependent transition rates.

6.7 ANALYSIS OF RATES

Different methods exist to extract rates in optical tweezers experiments, depending on the investigated system and the applied measurement procedure. The following three subsections are devoted to this problem.

6.7.1 *Rates from constant distance or jump experiments*

If the protein states of interest are populated at a constant trap distance (see Section 5.3), HMM analysis as described in Section 6.2 is a good choice to read out the lifetimes of different protein states directly. These lifetimes are pooled into histograms and the off-rates of all populated states are determined by fitting single exponentials.

Deviation of the measured lifetimes from the single exponential fit is a strong indication that states are not correctly classified and multiple states mix.

Most systems in this thesis just populate two states at the given pre-tension. In this case the off-rates of both states are the on-rates into the respective other state. If more than two states are present, the off-rate of a certain state k_i is the sum of all off-rates from this state into all other states $k_i = \sum_{m \neq i} k_{i \to m}$. The individual transition rates $k_{i \to j}$ can be calculated by Equation 6.20 using the probabilities p_{ij} to go from states i to states j. They are determined by counting the number of transitions into each state.

Alternatively, lifetime histograms for all transitions can be fitted individually.

$$k_{i \to j} = k_i p_{ij} \tag{6.20}$$

If the distribution of lifetimes is close to the temporal resolution of the experimental setup (around 1 ms), missed events have to be considered. This is important because missing many transitions affects both rates significantly. They can be taken into account by introducing cutoffs in the lifetimes fits, which are then used in a simulation of the complete transition network to correct the transition rates for missed events [138].

An auto-correlation-based method can measure rates down to 10 µs *[171].*

Changing the trap distance allows the measurement of the rates at different forces and extrapolation to zero force (see Section 6.6). If the states don't exchange rapidly in constant distance experiments, they can be evaluated similarly in jump experiments (see Section 5.3).

6.7.2 *Rates from constant velocity experiments*

Kinetic information can also be extracted from unfolding and refolding force-extension traces. The unfolding or refolding force measured is determined by the respective force-dependent rate. *Force distributions* $p(F)$ of many unfolding or refolding events are easily obtained by recording multiple force-extension cycles. These force distributions approximately have a skewed Gaussian shape, which is skewed to low forces for unfolding events and high forces for refolding events,

because rates change with force. Force-dependent rates $k(F)$ can be calculated from force distributions using [30]:

$$k^{-1}(F) = \frac{\int_F^\infty p(f)df}{\dot{F}(F)p(F)} \tag{6.21}$$

A critical parameter is the force-loading rate $\dot{F}(F)$, i.e. the change of force with time, which depends on the pulling speed. In assays without flexible elements, the force-loading rate is linearly proportional to the pulling speed, but for tethered assays the force-loading rate is a function of the force itself.

By changing the pulling speed, different force regions can be explored and rates can then be fitted with one of the models described in Section 6.6. Alternatively, analytical expressions that already consider the force dependence of the rates in some way can be used. Then the rate at zero force is directly obtained from a fit to a force distribution or from the force maxima at different loading rates.

E.g., for the Bell-Evans model transition rates depend on the logarithm of the loading rate.

A very similar model that avoids the complication of calculating the force-loading rate is a method presented by Oberbarnscheidt *et al.* [95]. In this approach the force-extension traces are divided into small force bins. For each force bin ΔF the occurrence of the transition as well as the time spent in the bin is evaluated. Taking multiple traces allows the determination of the force-dependent rates for each bin $k(\Delta F)$ by the number of transitions observed $n(\Delta F)$ divided by the total time spent at that force $\Delta t(\Delta F)$, when the system has not transitioned yet.

$$k(\Delta F) = \frac{n(\Delta F)}{\Delta t(\Delta F)} \tag{6.22}$$

This approach can be seen as a constant distance experiment, where the constant distance trace at a certain force is reconstructed from small parts of multiple constant velocity traces in the same force region. The zero force rates can then be determined using one of the models described in Section 6.6.

6.7.3 *Determination of rates from probabilities*

An indirect way to measure the force-dependent transition rate from state i to j is to measure the probability $P_j(F_i, t)$ to observe a certain state j after a certain time t at a certain force F_i, if the system has been prepared in state i ($P_j(F_i, t=0) = 0$). For a first order reaction with one force-dependent rate $k_{i\to j}(F_i)$, the transition rate is related to the transition probability:

$$P_j(F_i, t) = 1 - e^{-k_{i\to j}(F_i)t} \tag{6.23}$$

If the reverse reaction $k_{i\to j}$ is also possible, solving the coupled differential rate equations yields the probability of finding the system in state j after it had been prepared in state i:

$$P_j(F_i, F_j, t) = \frac{k_{j\to i}(F_j)}{k_{j\to i}(F_j) + k_{i\to j}(F_i)} \left(1 - e^{-\left(k_{j\to i}(F_j)+k_{i\to j}(F_i)\right)t}\right) \quad (6.24)$$

Fitting these equations to unfolding or refolding probabilities obtained e.g. by jump experiments allows the determination of the rates. This is very useful for probing transitions at almost zero force, where no signal can be obtained directly in optical tweezers experiments.

Part III

RESULTS, DISCUSSION AND OUTLOOK

In this main part I present and discuss the results obtained using different experimental methods. The first two chapters deal with the structure and the folding behavior of Hsp90 and the following two concentrate on the dynamic properties Hsp90. At the end I give a short outlook to guide the future direction of single molecule Hsp90 measurements.

7

THE MECHANICAL FINGERPRINT OF HSP90

In this chapter the complex unfolding properties of yeast Hsp90 are introduced. The mechanical response of different Hsp90 monomer and isolated domain constructs are measured, which, taken together, provide a detailed structural interpretation of Hsp90. This is the basis for more elaborate work described in the following chapters.

7.1 UNFOLDING THE HSP90 MONOMER

To determine the mechanical fingerprint of Hsp90, i.e. the mechanical *unfolding pattern* of the natively-folded molecule, the Hsp90 monomer is expressed as a fusion protein with two ubiquitins, one at each terminus, which serve as spacers between the protein and the DNA handles and allow force exertion directly on the N- and C-termini [139, 125]. An illustration of the Hsp90 construct in the optical tweezers assay is shown in Figure 7.1 A. By slowly moving the beads apart with constant velocity Hsp90 is driven from the completely native state to the fully unfolded state and typical *Hsp90 force-extension traces* are obtained (see Figure 7.1 B).

For details on protein design and sample preparation see Sections A.1 and 5.2.

The trace shows *three consecutive unfolding events*, where the cooperative unfolding of the structural units leads to sudden drops in force followed by stretching of the unfolded polypeptide chain. The size of the folded elements, which are obtained by WLC fits, correspond very well to those expected for unfolding of the three Hsp90 domains (see Section 7.4). The first and smallest event is the unfolding of the structured part of the C domain, followed by the N domain at higher forces. Finally the M domain unfolds, showing the longest contour length gain and a comparable unfolding force to the N domain. After these unfolding events Hsp90 is completely unfolded and further stretching doesn't result in further unfolding events since the still folded ubiquitins usually don't unfold at the forces applied.

In the displayed example trace the C domain unfolds twice, while refolding once in between.

At low forces, where all Hsp90 domains are still intact, rapid transitions due to the Charged Linker, which is discussed in detail in Chapter 9, are observed. The stretching behavior of the DNA and the unfolded protein are well described by WLC models (see Section 6.1), if unstructured and unstable regions are considered. To allow for the fact that parts of the C domains are unfolded and that the Charged Linker is unstable an additional contribution of 48 nm contour length for WLC fitting is assumed (see Section 6.1.1).

20 nm for unstructured parts in the C domain plus 28 nm for the Charged Linker.

To gather statistics constant velocity experiments are performed at a speed of 500 nm/s. Plotting the unfolding forces against their asso-

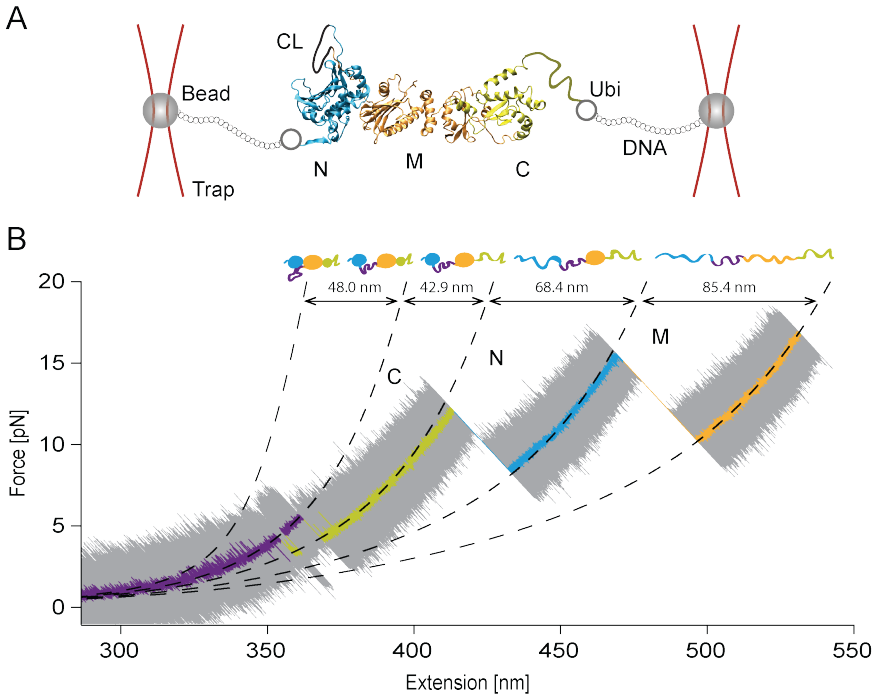

Figure 7.1: (A) Schematics of the construct Hsp90-diUbi, clamped between two beads held by optical tweezers. (B) The unfolding pattern of Hsp90-diUbi obtained by moving the beads apart at a speed of 10 nm/s. The force-extension trace shows three peaks that mark the unfolding of the C (purple to green), the N (green to blue) and the M (blue to orange) domain. The main unfolding sequence is sketched above the trace. After each unfolding event the length of the unfolded protein chain increases. The stretching behavior of the currently unfolded peptide is described well by WLC fits shown as dashed lines. Their contour length gains are shown and the first one accounts for stretching of unstructured and unstable regions within Hsp90. The unfiltered force-extension curve is shown in gray, while the colored trace is smoothed with a sliding average.

ciated contour length gains of multiple native unfolding traces yields a *scatterplot* as shown in Figure 7.2. As expected three distinct populations corresponding to Hsp90's domains are found. Their mean values are presented in Table 7.1.

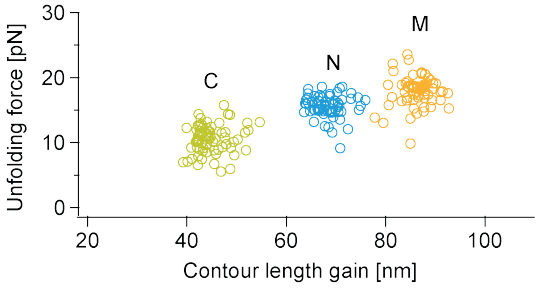

Figure 7.2: Scatterplot of the unfolding forces versus the contour length gains of Hsp90-diUbi at a pulling speed of 500 nm/s. The three distinct populations mark the unfolding of the three Hsp90 domains (N, M and C). Data are from 11 different molecules, a totality of 76 traces.

DOMAIN	CONTOUR LENGTH	UNFOLDING FORCE
N	67.6 nm ± 3.1 nm	14.9 pN ± 1.7 pN
M	85.3 nm ± 2.1 nm	17.7 pN ± 1.2 pN
C	42.3 nm ± 2.8 nm	9.9 pN ± 2.3 pN

Table 7.1: Average contour length gains and unfolding forces for Hsp90-diUbi domains calculated from 76 unfolding traces of 11 different molecules. Mean values and SD are determined by Gaussian fits. Note that unfolding force histograms are only Gaussians in the first approximation (see Section 6.7.2).

7.2 UNFOLDING BY INTERNAL FORCE EXERTION

To examine the Hsp90 monomer without ubiquitins and without the unstructured portion of the C domains, the residue at position 61 located in the N domain and the residue at position 560 at the beginning of the C domain are chosen as force-exertion points by mutating them to cysteines (see Section A.1). Figure 7.3 shows an example unfolding trace at a slow pulling speed of this variant, as well as a scatterplot of the unfolding events of multiple traces at a high pulling speed, similarly to the previous section.

These positions were also altered in fluorescence experiments [87, 115].

As expected the contour length gain measured for the N domain is shorter than in the N- and C-terminal-pulled construct (compare to Section 7.1). In addition an N domain intermediate is significantly more populated. The observed pattern for the M domain is identical to that observed in Hsp90-diUbi, since the *pulling geometry* is the

A similar intermediate is also present in Hsp90-diUbi, but shorter-lived.

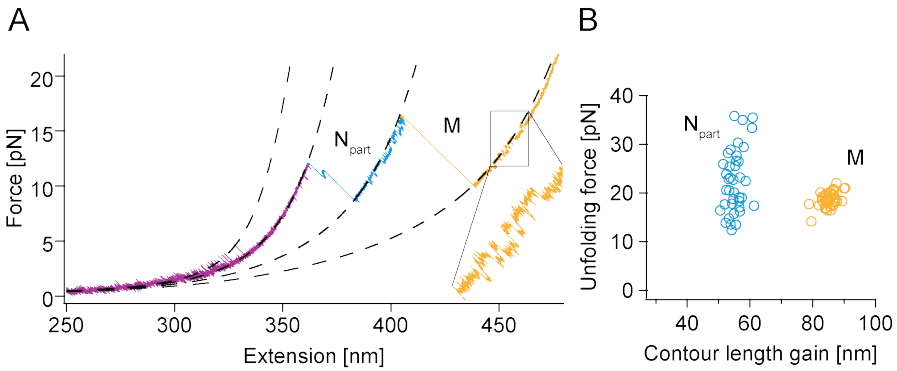

Figure 7.3: (A) Unfolding force-extension trace of Hsp90_61C_560C, an Hsp90 monomer pulled at aminoacids 61 and 560 with a velocity of 10 nm/s. Due to the different pulling geometry the unfolding event for the N domain (N_{part}) is shorter compared to the full length variant (see Figure 7.1) and shows a long-lived intermediate, whereas the M domain remains identical. The flipping in all slopes at forces between 10 and 15 pN is unfolding and refolding of the C domain (see inset). The contour length of this transition is significantly lower since only a few aminoacids are under tension. WLC models (dashed lines) are fitted to the major unfolding events.
(B) A scatterplot of the unfolding forces versus the contour length gains of the shortened N domain and the M domain at a pulling speed of 500 nm/s is displayed. Data are from 46 traces of 13 different experiments.

same. Although the N and M domains have unfolded, transitions with a short contour length change of around 4 nm at forces between 10 and 15 pN are still observed. These are due to the unfolding and refolding of the C domain, which has dramatically changed unfolding characteristics in this pulling geometry. The unfolded part of the C domain is not between the force exertion points, therefore a contour length of only 28 nm accounts for the unstable Charged Linker during WLC fitting.

In principle the transitions could also arise from rebinding of some structure to the not completely unfolded N domain, but this scenario is excluded by other measurements.

Table 7.2 shows the average contour length gains and unfolding forces of this construct obtained from the data shown in Figure 7.3 B. Comparing these values to Table 7.1 shows that the N domain is around 15 nm shorter, due to the different pulling geometry. In addition, higher unfolding forces and a broader force distribution are measured, indicating a lower force sensitivity.

DOMAIN	CONTOUR LENGTH	UNFOLDING FORCE
N_{part}	$53.7\,nm \pm 1.5\,nm$	$20.3\,pN \pm 4.8\,pN$
M	$85.6\,nm \pm 1.1\,nm$	$17.6\,pN \pm 1.1\,pN$
C_{part}	$\sim 4\,nm$	$10 - 15\,pN$

Table 7.2: Average contour length gains and unfolding forces for Hsp90_61C_560C calculated from 46 unfolding traces of 13 different molecules.

7.3 UNFOLDING OF THE INDIVIDUAL DOMAINS

Apart from the major unfolding transitions, Hsp90 monomers show many smaller, but observable transitions. For many of them it is not obvious, to which domain they belong. Therefore isolated single domain constructs of the N domain, the M domain and the C domain flanked by ubiquitins are designed (see Section A.1). Figure 7.4 displays force-extension unfolding traces of the three constructs, one example curve each for pulling speeds of 10 nm/s and 500 nm/s.

For the N domains (see Figures 7.4 A and B) the DNA fit deviates from the traces at high forces before the N domain unfolds completely. Some structure can *unfold and refold rapidly (flipping)* before the major unfolding transition sets in. This effect is clearer in the faster pulled traces. This preflipping involves more than two states and the contour length between the shortest and the longest state is around 9 nm. Furthermore the major unfolding transition passes through multiple short-lived (around 1 ms) states (not shown explicitly).

The M domains (see Figures 7.4 C and D) also show preflipping. Unlike that seen for the N domain, this preflipping seems to be just between two states, which are separated by 7.5 nm in contour length. In the major unfolding transition, the M domain passes through a

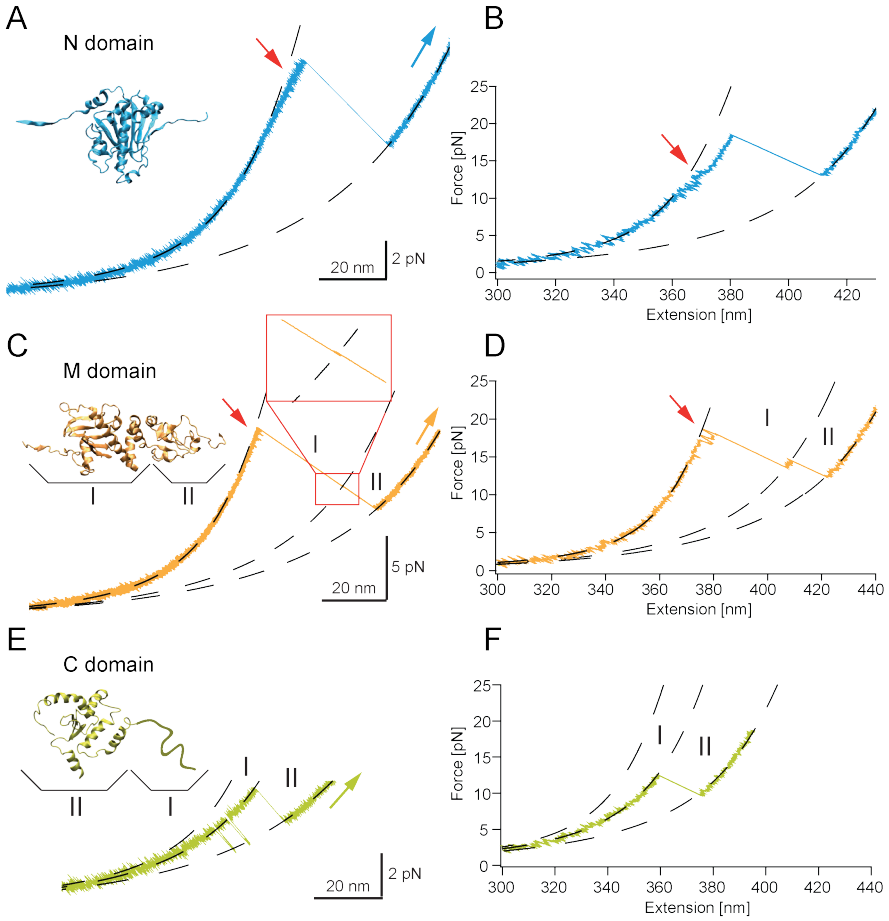

Figure 7.4: Example force-extension traces of N-diUbi at a pulling speed of 10 nm/s (A) and 500 nm/s (B) show rapid transitions (red arrows) before the major part of the domain opens.
Example force-extension traces of M-diUbi at a pulling speed of 10 nm/s (C) and 500 nm/s (D) show two-state preflipping (red arrows) before the domains unfold. Additionally a stable unfolding intermediate I_M due to the subdomain structure (region II) is observed in the fast-pulled traces as well as in the slow pulled traces (see zoom).
Example force-extension traces of C-diUbi at a pulling speed of 10 nm/s (E) and 500 nm/s (F) show no additional intermediates, only the unstructured regions (region I) have to be considered for the WLC fits.

long-lived intermediate (I_M), which seems to be obligatory (on-pathway) since it is observed in all unfolding traces. The contour length gain of the stable intermediate is shown in Table 7.3.

This intermediate is further investigated in Section 8.6.2.

The C domains (see Figures 7.4 E and F) don't show any unfolding intermediates, but unlike the N and M domains it can refold in the range of the unfolding forces on the timescale of the experiment, as already observed in the monomeric constructs.

The refolding behavior of the C domain is addressed in more detail in Section 8.7.

The total contour length gains of all domains as well as the contour length gain of the prominent M domain intermediate (I_M) are evaluated for multiple molecules at a pulling speed of 500 nm/s. Averages and SDs are shown in Table 7.3. For identical pulling speeds the unfolding forces (not explicitly shown) are similar to those observed for the monomer (see Table 7.1).

DOMAIN	CONTOUR LENGTH
N	70.35 nm ± 2.06 nm
M	85.92 nm ± 1.68 nm
I_M	28.42 nm ± 1.09 nm
C	39.26 nm ± 1.47 nm

Table 7.3: Average contour length gains of the individual domain constructs. The displayed values are means and SDs of 158 traces (11 molecules) for the N domain, 176 traces (8 molecules) for the M domain and 174 traces (7 molecules) for the C domain.

7.4 STRUCTURAL INTERPRETATION

This section gives a detailed structural interpretation by taking into consideration the unfolding events observed for the different Hsp90 variants described in the last three sections.

The main unfolding peaks observed for the monomer construct (Hsp90-diUbi) clearly match the unfolding events observed for the individual domain constructs (N-diUbi, M-diUbi and C-diUbi) and therefore validates the domain assignment. Table 7.4 compares the observed contour length gains of the individual domain constructs with the expected ones based on the number of aminoacids and the distances in the Hsp90 crystal structure [2]. For details see Section 6.1.2. The numbers for the N and M domains agree very well, only the C domain is shorter than expected (see interpretation below).

DOMAIN	DISTANCE IN CS	EXPECTED	MEASURED
N (2-208)	6.52 nm	69.04 nm	70.35 nm ± 2.06 nm
M (273-527)	6.54 nm	86.53 nm	85.92 nm ± 1.68 nm
C (538-671)	3.87 nm	45.04 nm	39.26 nm ± 1.47 nm

Table 7.4: The theoretically expected contour length gains of the different domains calculated from the number of aminoacids (regions in brackets) and the initial end-to-end distances in the folded state of the crystal structure (pdb id: 2cg9) are compared to the contour length gains obtained from WLC fits.

On a closer inspection the domains themselves show multiple states such as preflipping or unfolding intermediates, as described in Section 7.3. A structural interpretation of these events is given in Figure 7.5.

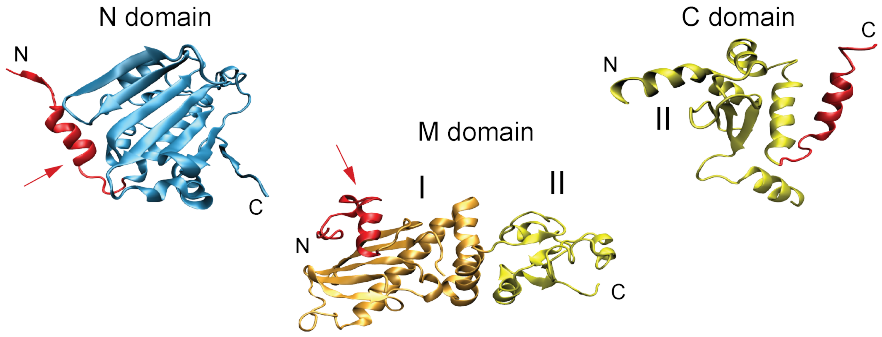

Figure 7.5: The crystal structures of the N, M and C domains are shown (pdb id: 2cg9). Different colors code for different features in the force-extension traces shown in Figure 7.4. The regions in red give rise to the preflipping observed for the N and M domains (red arrow). The yellow part of the M domain is the structure of the unfolding intermediate I_M. The last alpha helix of the C domain seems to be unfolded if not dimerized (red).

The N domain preflipping is observed for the full length construct (Hsp90-diUbi) and the individual domain construct (N-diUbi), but it is absent in the construct pulled at aminoacid 61 (Hsp90_61C_560C). It follows that the preflipping arises from aminoacids at the N-terminus. The contour length is around 9 nm, which is slightly higher than expected, if these transitions result from aminoacids 1-27. That aminoacids beyond residue 27 contribute to the preflipping is unlikely since they are buried within the domain. Therefore minor contributions from the C-terminal part of the domain cannot be fully excluded, but the contour length gain measured for the N_{part} domain (Hsp90_61C_560C) is 53.7 nm ± 1.5 nm, which is very close to the value of 52.4 nm expected for a fully folded C terminus.

N-terminal deletion mutants of Hsp90 (up to aminoacid 24), are also stable in biochemical experiments [121].

The unfolding intermediate (I_M) observed for the M domain is explained by the smaller alpha-beta-alpha subdomain at the C terminus of the M domain. The measured contour length gain is 28.4 nm ± 1.1 nm, which matches the expected one of 27.9 nm. This suggests that the preflipping observed for the M domain is due to some structure at the N-terminal end. Indeed the N-terminal alpha helix fits the contour length gain perfectly.

Changing the C terminus of the M domain influences the folding properties of this intermediate (see Section C.3.4).

The length estimate for the folded part of the C domain in the crystal structure is around 6 nm longer than the measured value (see Table 7.4). Folding and unfolding of the C domain is observed in the Hsp90_61C_560C construct with a contour length of 4 nm. Since force is applied at aminoacid 560 this corresponds approximately to

residues 538 and 560. Therefore the missing 6 nm likely result from the C terminus. The missing contour length can be explained by the last alpha helix which is not folded if the Hsp90 is in the monomeric form. This assumption is verified in Section 10.2.

7.5 DISCUSSION

7.5.1 *Structure of the Hsp90 monomer*

The single molecule tweezers assay allows the investigation of the *mechanical denaturation* process of the *large multi-domain protein* Hsp90 with high structural and temporal resolution.

The contour length gains obtained by force-extension traces match the expected ones of the N, the M and the C domains well (see Section 7.4). However, the sum of all domain contour length gains is shorter than the contour length gain expected if all 709 aminoacids of Hsp90 are fully structured. This is resolved by multiple regions of Hsp90 that are found to be unstructured or unstable: First, the Charged Linker located between the N and M domains, which appears as rapid transitions at low forces in slowly-pulled force-extension traces, doesn't contribute to the domains. Second, the large unstructured part of the C domain is stretched without any force signal. Finally, the linker between the M and C domain with a length of around 10 aminoacids appears to be unstructured as well. These unstructured or unstable regions have to be considered to extract reliable contour length gains for the domains as discussed in Section 6.1.1.

First experiments with a loop insertion mutant (see Section A.1) indicate no significant stabilization between the M and C domains.

Unlike the C domain, the N and the M domains show multiple intermediates, meaning that mechanically stable states during the unfolding process exist. Measuring different Hsp90 mutants and using different pulling directions permits the characterization of these intermediate states in high detail and allows the structure of the intermediates to be resolved (see Section 7.4).

Yeast Hsp90 is also characterized by a completely different experimental method. In AFM experiments the contour length gains of the individual domain constructs could be measured precisely as shown in Section C.2. The AFM data are in very good agreement with the optical trap data. However, measuring the force response of the full length Hsp90 monomer in the AFM is tedious, mainly because the assay used relies on non-specific adhesion to the cantilever. The experiments get more and more challenging with increasing protein size. Another limitation is that the force range interesting for Hsp90 is at the resolution limit of the AFM experiments. With the optical trap and the dumbbell assay these limitations can be overcome, offering a perfect way to study Hsp90 in mechanical experiments.

By comparing the unfolding pattern of Hsp90 from different organisms, it is possible to draw conclusions how the protein's structure

has evolved. This is done in Chapter D showing unfolding experiments with *E.coli* Hsp90.

7.5.2 *Mechanical stability of the domains*

These are moderate unfolding forces compared to other proteins.

The unfolding stabilities of Hsp90s domains are in the range of several to tens of piconewtons. As detailed in Section 6.7 the unfolding forces depend on the force-loading rate and therefore on the pulling speed. Increasing unfolding forces with increasing force-loading rates are observed for all investigated constructs.

Also changing the force-exertion points and therefore the pulling direction can modulate the unfolding forces [26], which is the case for the N and C domain in the Hsp90_61C_560C construct. For the C domain this behavior is clearly evident. If it is pulled slowly at its termini, it unfolds (and refolds) at forces of around 4 pN (see Figure 7.1). If it is pulled at the N-terminus and residue 560, only 20 aminoacids are under force and unfolding (and refolding) forces are around 12 pN (see Figure 7.3). This increase of unfolding force with reduced length suggests that the energies are similar and that by just pulling on one helix the whole C domain is unfolded.

Comparing the unfolding forces of the domains in the monomeric construct and the forces of the isolated domain constructs does not show any major difference. This suggests that at higher forces *no significant domain-domain interactions* occur [44]. This is supported by two observations: First, the sequence of unfolding events may change, meaning that the unfolding transitions of all domains in the monomer are independent of the other domains. Second, the structural model suggests unstructured aminoacids between the domains, therefore the domains are connected by flexible and not rigid linkers.

Due to the different unfolding rates there is still a most likely unfolding sequence.

Although studied extensively, no measurable changes could be induced by adding an excess of nucleotides or cochaperones to the monomer. Nucleotides ATP, ADP and AMP-PNP as well as co-chaperones p23/Sba1 and Aha1 are tested with Hsp90_61C_560C. This is somewhat surprising because K_Ds for nucleotide binding are low [85, 115, 110] and some stabilization of the N domain can be expected. The reason why no stabilization is observed might be that the reaction coordinate probed, is not sensitive to nucleotide binding. However, if Hsp90 monomers are present in solution and dimers are formed, nucleotide-sensitive changes are observed (see Section 10.5).

The Hsp90_61C_560C construct has wild type ATPase activity (see Section 9.5).

8

FOLDING BEHAVIOR OF HSP90

Large proteins consist of long aminoacid chains, often comprising many hundreds of residues and multiple domains. While the folding process of small single-domain proteins is well understood [52, 20, 132, 139, 49], studying such large systems remains challenging. Bulk experiments can only provide limited information about the *folding process of large multi-domain proteins*, because these systems typically populate numerous states including on- and off-pathway intermediates [34, 162, 15]. Therefore kinetics on different time scales and parallel processes are expected. In addition, the structural interpretation of the states and intermediates observed in bulk experiments is not trivial. With optical tweezers experiments these limitations can be overcome by actively manipulating one molecule at a time and measuring with high temporal and spatial resolution.

In this chapter the folding behavior of the yeast Hsp90 monomer is investigated *in vitro*. The influence of multiple domains on protein folding is elucidated by comparing the monomer to the isolated domain constructs. Already the N and M domains show unexpected folding behavior, and the effect of tethering the domains changes the folding process significantly. Finally, the implications of the results on *in vivo* expression and folding are discussed [53].

The first Hsp90 refolding experiments with chemical denaturant showed that folding is reversible and an unspecified intermediate occurs [55].

8.1 FOLDING OF THE HSP90 MONOMER

The Hsp90 monomer is forced into the completely unfolded state by slowly moving the trapped beads apart. Then the refolding process of the polypeptide chain is observed in real time by measuring the force while moving the beads together. Such experimental traces are shown in Figure 8.1 for three consecutive unfolding and refolding force-extension cycles of the same molecule.

Hsp90's unfolding properties are described in Section 7.1.

First refolding events of the completely unfolded chain are measured at forces around 5 pN. With decreasing force many *transiently populated states* are observed, especially before larger refolding transitions take place. The traces of the major refolding events usually overlap with the unfolding traces, suggesting that the domains have fully refolded and the preferential domain refolding sequence is reverse to the unfolding sequence.

This is true for the 2^{nd} and the 3^{rd} refolding trace shown.

The curves show that Hsp90 populates a network of many different states, some of them close to equilibrium, before it reaches its native state, where all domains are natively-folded. That Hsp90 indeed found its native state is supported by the fact that consecutive

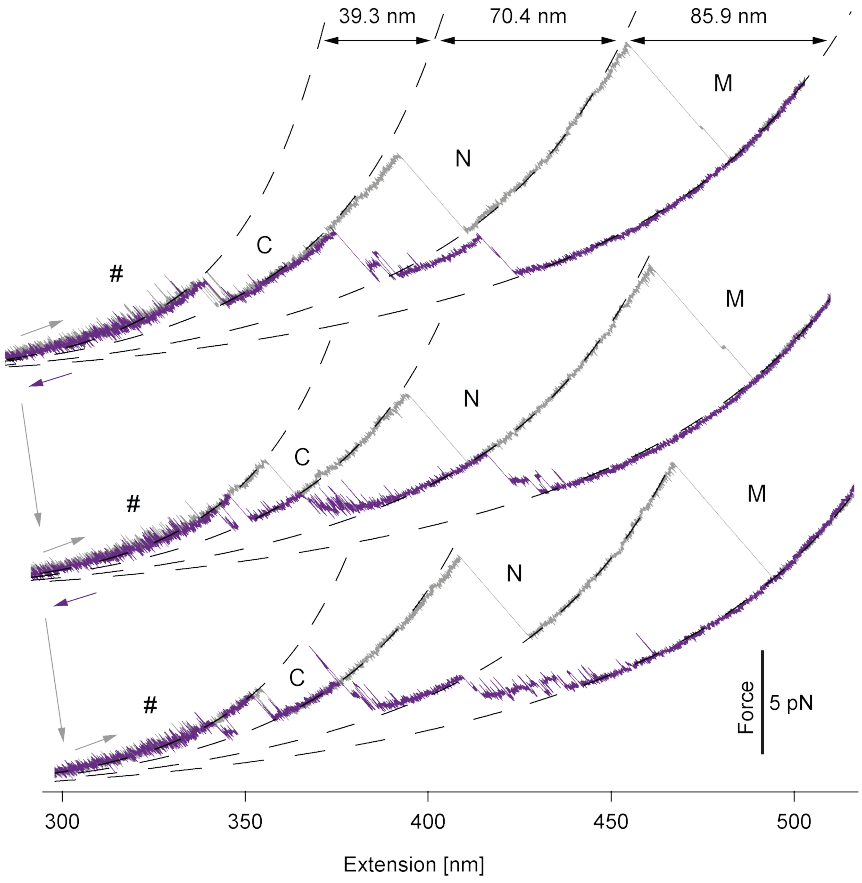

Figure 8.1: Unfolding (gray) and refolding (purple) force-extension traces of Hsp90-diUbi at a pulling speed of 10 nm/s. WLC fits to the unfolding traces mark the three domains (N, M and C). Refolding sets in at forces around 5 pN and after some large refolding transitions the refolding trace has the same contour length as the unfolding trace indicating that a complete domain has refolded. But also many rapid transitions, especially with shorter contour length gains are observed. At extensions smaller than 340 nm the unfolding and refolding trace overlap and the Charged Linker fluctuations (#) are observed. Unfolding the monomer again shows the completely native pattern, which proves that Hsp90 has fully refolded. The contour length gains of the domains shown above the uppermost trace are taken from Table 7.3.

unfolding traces always have the completely native pattern and that the Charged Linker fluctuations at an extension of around 340 nm are observed (see Chapter 9).

At slow pulling speeds of 10 nm/s the Hsp90 monomer always refolds, but for increased speeds, the protein is given less time to fold and the native pattern is not always recovered. Figure C.3 shows example traces at a pulling velocity of 500 nm/s. There only 36% ± 7% (238 traces of 5 molecules) of the unfolding force-extension traces show the fully native pattern. The time spent at forces that allow refolding is around 1 s. Using Equation 6.23, which assumes two-state behavior and neglecting the force dependence, gives a first estimate of 2.5 s^{-1} for the folding rate from the unfolded to the native state.

The distance traveled in this region is around 500 nm.

8.2 FOLDING OF THE N DOMAIN

While the monomer refolds, it populates many different states. To delineate their origins the refolding behavior of isolated domain constructs (see also Section 7.3) is investigated in the following, beginning with the N domain.

8.2.1 *The N domain shows folding intermediates*

Figure 8.2 A shows example force-extension traces at a slow pulling speed of 10 nm/s for the unfolding and the refolding of the isolated N domain. Before the N domain reaches its native state, where the refolding trace lies on top of the folded state in the unfolding trace, numerous short-lived (transient) intermediates are populated. They seem to be heterogeneous involving different contour lengths.

For details on N domain unfolding see Section 7.3.

For slow pulling and relaxing, molecules always end up in the native fold, but for higher speeds this behavior changes, as shown in Figure 8.2 B. Two possible patterns for the unfolding traces are found: Either the protein has completely folded back into its native state or it is trapped in some non-native intermediate state. Surprisingly, the completely unfolded state is never observed, suggesting it is rapidly depopulated at low forces.

That slowly-relaxed N domains always end up in the native state is expected because Hsp90 monomers show the same behavior (see Section 8.1).

The non-native traces show different contour length gains and stabilities, very different to the defined unfolding properties observed for the native state. This is visualized in the scatterplot shown in Figure 8.2 C. The *heterogeneity of the non-native states* suggests that the intermediates are an ensemble of different states. It stands to reason that they are identical or very similar to the states observed during slow refolding in Figure 8.2 A. To investigate the nature of these intermediates and their effect on protein folding the next two sections deal with possible effects of intermediates and a new measurement protocol to investigate force-dependent protein folding.

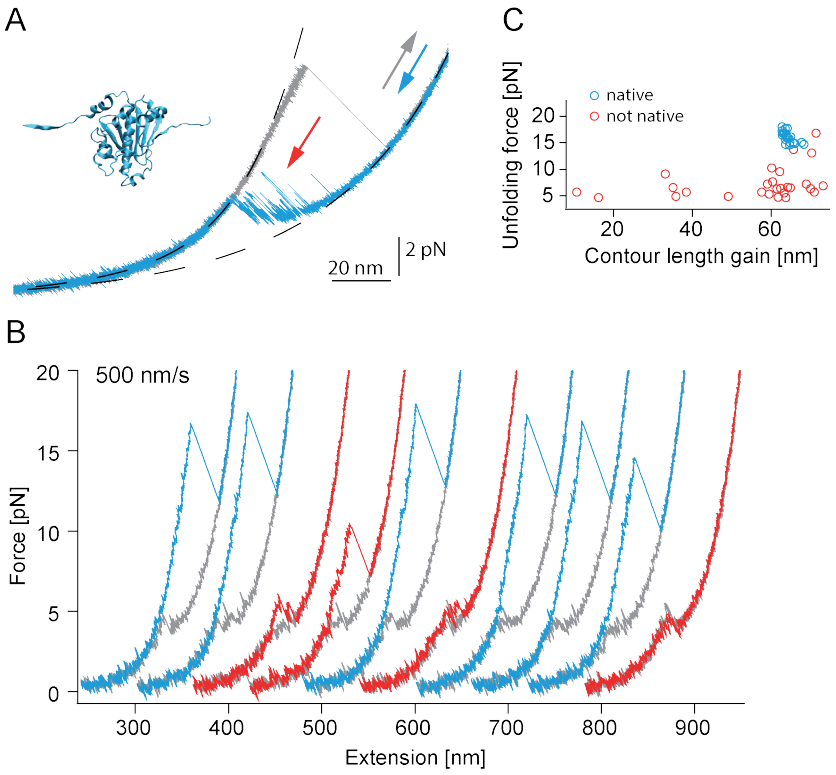

Figure 8.2: (A) Unfolding (gray) and refolding (blue) of N-diUbi (Inset from pdb id: 2cg9) at a pulling speed of 10 nm/s. The folded and the unfolded states are marked by WLC fits. Before the N domain refolds, it populates many transient intermediates, indicated by the red arrow. More example traces are shown in Figure C.4.
(B) Ten consecutive, offsetted cycles at a pulling speed of 500 nm/s. Natively-folded unfolding traces are colored in blue, non-native unfolding traces in red and refolding traces in gray.
(C) A scatterplot of the unfolding events obtained from 73 traces of an individual molecule. Native unfolding events show a well defined cluster (blue), while non-native events show very diverse contour length gains and unfolding forces (red).

8.3 INFLUENCE OF INTERMEDIATES ON PROTEIN FOLDING

Intermediates are kinetically stable states, which can occur during the folding process. On-pathway intermediates are along the folding pathway and have native intramolecular contacts, while off-pathway intermediates are not along the folding pathway and typically contain non-native contacts. Generally, off-pathway intermediates decrease the effective folding rate. How on-pathway intermediates influence the folding speed is less clear and depends on the system [151]. However, with increasing number of intermediates, the number of possible transition rates rises, complicating the characterization and interpretation of such systems.

Off-pathway intermediates are often called misfolds.

Kinetic protein folding studies are typically performed by perturbing the system with chemical denaturants and measuring *relaxation rates* in a stopped-flow apparatus. The measurement signal is tryptophan fluorescence or circular dichroism amongst others. The numbers of intermediate states and their nature are deduced from the denaturant-dependent rates [161, 10]. In these experiments often *burst-phase intermediates*, which are structures that form within the dead time of the experiment, are observed [113]. They can lead to so-called *roll-overs*, which are deviations from the linearity in rate chevron plots at low denaturant concentrations [8, 60, 10].

Like chemical denaturant, force can be used to elucidate the nature of intermediates. The effect of force on a rate between two states is described in Section 6.6. If additional on-pathway intermediates are present, force may depopulate them, leading to a further *decrease* of the overall refolding rate. The opposite is true for off-pathway or misfolded intermediates that can be depopulated by applying force, thus *accelerating* folding kinetics. In contrast to chemical denaturant experiments, single molecule optical tweezers also provide detailed structural information about the intermediates simplifying the interpretation of the experiment. However, the exact force dependence of the effective folding rate depends on the force sensitivities of all states involved.

8.4 DOUBLE-JUMP RAMP EXPERIMENTS

In constant velocity experiments the force that is applied to the protein changes with time, therefore these experiments are not well-suited to study the force dependence of the refolding process. To overcome this limitation and measure refolding probabilities after defined times at defined forces, a *special mechanical double-jump protocol*, similar to the one in [130], is established (see also Section 5.3).

This assay is illustrated in Figure 8.3 A. Starting from the completely unfolded peptide, the chain is relaxed to a certain low force value by decreasing the trap distance very rapidly (jump down). For

These jumps take less than 10 ms *in the setup used.*

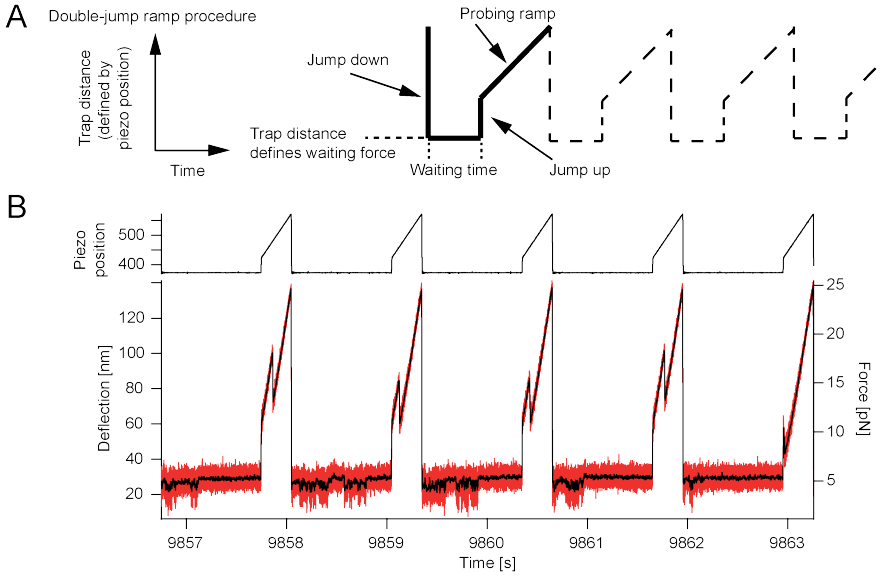

Figure 8.3: (A) Illustration of the repetitive double-jump ramp pattern, which allows the measurement of refolding probabilities with defined refolding parameters. (B) A part of such a double-jump ramp experiment with the isolated N domain shows the measured trap distance (top) and the force response of the N domain (bottom).

this trap distance the "waiting force" is defined, as the force that acts on the unfolded chain. After the protein is given a certain time to refold, the "waiting time", the force is increased rapidly to 5 to 12 pN (jump up). This quenches the refolding process, but does not yet unfold native structure or stable intermediates. Directly after jumping up the protein state is probed by an constant velocity unfolding ramp. An example for a typical force response of the protein is shown in Figure 8.3 B. This assay is especially useful for complex systems that have multiple structural units and/or multiple states, such as the Hsp90 constructs, because the unfolding force-extension traces allow the detailed characterization of the different protein states.

Waiting time and waiting force are also called refolding time and refolding force in the text.

To determine refolding probabilities for a wide range of waiting times and waiting forces, large datasets are generated. To ease evaluation a *deterministic classifier* is programmed, which is described in detail in Section B.2.2. In short, unfolding force-extension traces are transformed into contour length space and histogrammed. Using a peak finding routine allows the determination of the contour lengths of the states in the system. Related unfolding forces are found by searching for the local maximum in the force-extension trace. If the found contour length gains and unfolding forces match those expected for the native states within some tolerance, the events are classified as native.

In addition to the unfolding force-extension traces, the force signals during the waiting time, where the traps are at constant distance (see

also Section 5.3) are recorded. If the forces are high enough to obtain a signal, it is possible to follow the refolding process directly, as shown in Figure 8.3 B. In this limited force range, force-dependent refolding lifetimes can also be read out without determining the refolding probabilities.

8.5 FOLDING OF THE N DOMAIN - CONTINUATION

8.5.1 *Force-dependent folding experiments*

Employing double-jump ramp experiments, as described in the previous section, the *refolding probabilities* for different refolding times and refolding forces of the isolated N domain are determined. Figure 8.4 A shows the measured refolding probabilities in a 3D representation as black dots. The same data are presented in simplified plots with averaged probabilities in Figures 8.4 B and C. There the probabilities are plotted against waiting time and the different waiting forces are color-coded and the other way around.

The orange markers in Figure 8.4 B show the refolding probability against time at almost zero force. After around two seconds 70% of the molecules have folded, meaning that the effective folding rate is around $0.5\,s^{-1}$. Interestingly, applying small forces (up to 3.5 pN) during refolding, increases the refolding probability and therefore the refolding rate (from orange, over red, purple to dark blue markers). For higher waiting forces (light blue and teal markers) the refolding probability decreases as expected. This unusual force dependence becomes even more apparent in Figure 8.4 C, where the probabilities peak around 3.5 pN.

This atypical increase of the refolding rate with increasing force can only be rationalized by *the depopulation of the off-pathway intermediates* as reasoned in Section 8.3. Therefore the ensemble of intermediates observed for the N domain in constant velocity experiments (see Figure 8.2) constitutes mainly misfolded intermediates. The following decrease of the refolding rate at higher forces resembles the normal behavior, where force disfavors folding into the native state, as described in Section 6.6.

8.5.2 *N domain model*

A minimal kinetic model with only three states is enough to quantitatively describe the observed refolding probabilities of the N domain. The unfolded ensemble of folding-competent states [U] is in equilibrium with an ensemble of misfolded, folding-incompetent states [M]. Folding to the native state [N] is just allowed from the unfolded state ensemble [U] and is irreversible. This means that the folding process can be seen as a competition between the population of the misfolded

It is unlikely that only the completely unfolded state is folding competent, therefore all folding-competent states are referred to as the unfolded ensemble.

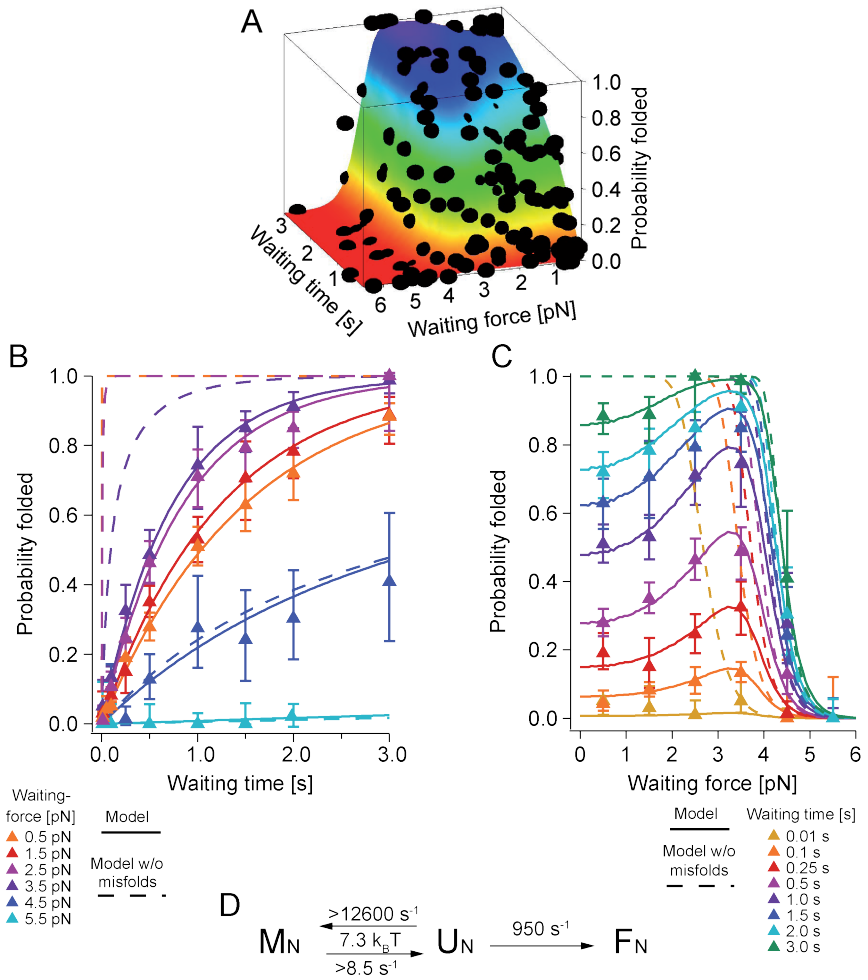

Figure 8.4: (A) Refolding probabilities dependent on time and force of N-diUbi obtained from double-jump ramp experiments are shown as black dots. Probabilities are determined from 11,570 traces of 45 molecules. The colored surface shows the best fit of the model described in Fig D.
(B, C) The same data are shown in a different representation with averaged probabilities. Fig. B shows the probabilities depending on the waiting time with the waiting force color-coded. Fig. C shows the probabilities depending on the waiting force with the waiting time color-coded. For uncertainties see Section B.2.3. At low forces an unexpected increase of the refolding probability with increasing waiting forces is observed, which identifies the previously observed intermediates as misfolded states. For solid and dashed lines see Fig. D.
(D) A three-state model quantitatively describes the observations. In this model unfolded states [U] are in equilibrium with misfolded states [M]. The native state [F] can only be accessed from the unfolded states. Numbers displayed are from the fit of Equation 8.1 to the data and from estimates. The fit is shown as a colored surface in Fig. A and as solid lines in Figs. B and C. If the misfoled states would be absent, the refolding probability would behave like the dashed lines in Figs. B and C.

states and the native state. Since the unfolded and misfolded ensembles are assumed to be in rapid equilibrium and the unfolding of the native state at the forces investigated is almost impossible, the system can be described by only four parameters.

- The zero-force folding rate from the unfolded states to the native state $k_{U\to F}$.
- Its force dependence x_{UT}, which is the contour length between the unfolded states and the transition state (T) on the way to the native state.
- The average equilibrium free energy between the misfolded and the unfolded states E_{UM}.
- Its force dependence x_{UM}, which is the mean contour length gain between the unfolded states and the misfolded states.

The probability of observing the native state P_F depending on the waiting time t and the waiting force F_U according to the model is (see also Section 6.7.3):

$$P_F(t, F_U) = 1 - e^{-P_U(F_U)k_{U\to F}(F_U)t} \tag{8.1}$$

$k_{U\to F}(F_U)$ is the force-dependent rate, which is described by Equation 6.19 and contains the first two fit parameters. $P_U(F_U)$ is the probability being in the unfolded, folding-competent states. It contains the other two fit parameters and can be written, similarly to Equation 6.11, as:

$$P_U(F_U) = \frac{1}{1 + e^{\frac{-E_{UM}(F_U, F_M)}{k_B T}}} \tag{8.2}$$

The model described by Equation 8.1 fits the measured folding probabilities very well. The fit is shown in Figures 8.4 A–C. An average free energy of $7.3(\pm 0.1)\,k_BT$ between the unfolded and the misfolded ensemble and a fast folding rate from the unfolded ensemble to the native state at zero load of $954(\pm 65)\,s^{-1}$ are found. These fit results are summarized in Figure 8.4 D. The average contour length of the misfolded states x_M and the position of the transition state x_T are fixed to constrain the fit. Varying these parameters show that $x_M = 33\,nm$ and $x_T = 39\,nm$ give the lowest value of χ^2.

The large distance between the folded state and the transition state indicates intermediates on the way to the native state.

To apply the model, average DNA and protein parameters are needed to determine the energy of the complete dumbbell system (see Section 6.3). They are obtained by WLC fits to constant velocity traces prior to the double-jump ramp experiments. The parameters used in the fit are $p_{DNA} = 18.3\,nm$, $l_{DNA} = 359.5\,nm$, $K = 400\,pN$, $p_{prot} = 0.7\,nm$, the contour length of the folded protein $x_F = 0\,nm$,

the contour length of the unfolded protein $x_U = 70.6\,\text{nm}$ and $k_{eff} = 0.176\,\text{pN/nm}$.

If misfolded states are excluded from the model and the same folding parameters are assumed, the force-dependent refolding probabilities would follow the dashed lines shown in Figures 8.4 B and C. Comparing them to the measured ones, highlights the *tremendous effect of misfolding at low forces*.

8.5.3 *Validation of the model*

In constant velocity refolding traces with a slow pulling speed (see Figures 8.2 A and C.4 A) the misfolded states are observed as transient intermediates. They appear at forces around 4 pN along a distance of roughly 20 nm. Thus their energy is approximately $E = 4\,\text{pN} \cdot 20\,\text{nm}/2 = 40\,\text{pNnm} = 9.8\,k_BT$, which is in good agreement with the value obtained from the fit, that solely depends on the refolding probabilities to the native state.

The model assumes that the misfolded states and the unfolded states are at equilibrium. This is reasonable, because very fast rates into the misfolded states are observed. This can be seen in the refolding force-extension traces at fast and slow speeds (see Figure 8.2), but also in double-jump ramp experiments after the jump down to low forces (see Figure C.5). Furthermore, if the average rate into the misfolded states would be on the order of the refolding rate, a biphasic force dependence would be expected. This is clearly not the case and all probabilities plotted against time (see Figure 8.4 B) follow single exponentials as the model demands.

To estimate the fast misfolding rate directly, constant distance traces similar to the one shown in Figure C.5 C are evaluated. The typical average lifetime of the unfolded state is 5 ms at 3.8 pN. Equation 6.19 with average DNA and protein parameters and a transition state position right in the middle between the unfolded and misfolded states, yields a zero-force misfolding rate estimate of $k_{U\to M} \geqslant 12600\,s^{-1}$. From this rate the reverse rate is easily determined as $k_{M\to U} \geqslant 8.5\,s^{-1}$ using Equation 6.16.

Since the transition state is typically closer to the folded states, this is a lower estimate.

Having determined or estimated all rates of the model (see also Figure 8.4 D) simulation of force-extension refolding traces is possible. They reproduce the measured traces well, as shown in Figure C.4 B.

Finally, the heterogeneity of the misfolded states (see Figure 8.2) is also observed in the constant distance part of the double-jump ramp experiments, as shown in Figure C.5. Analyzing such a trace with HMM suggests that conversion between different misfolded states is possible (see Figure C.6).

8.6 FOLDING OF THE M DOMAIN

Based on the experiments of the N domain in Sections 8.2 and 8.5, the isolated M domain is investigated.

8.6.1 *The M domain shows different folding intermediates*

Figure 8.5 A shows an example constant velocity cycle at 10 nm/s of the isolated M domain. Like the N domain (see Figure 8.2), the M domain populates transient intermediate states during refolding. Aside from that, a long-lived refolding intermediate is observed. It is reasonable for now to presume that it is the same intermediate observed in unfolding curves in Section 7.3.

Slowly-relaxed M domains always reach the native state.

If the pulling speed is increased, the M domain has less time to fold and the unfolding traces do not always exhibit the completely native pattern. As Figure 8.5 B shows, either native or non-native, but never completely unfolded traces are observed. Additionally, a significant number of partially folded traces are populated, where the aforementioned intermediate has successfully refolded but the remaining part hasn't. Alike the N domain, the intermediates show heterogeneous contour length gains and unfolding kinetics, which is best seen in a scatterplot of fast pulled unfolding traces (see Figure 8.5 C).

8.6.2 *The stable intermediate is on-pathway*

Different findings indicate that the stable M domain intermediate observed in unfolding and refolding traces is an on-pathway unfolding and refolding intermediate: first, the intermediate is present in all unfolding traces (see Section 7.3). Second, on the resolution of the experiment, an intermediate with very similar contour length, is always populated in refolding traces (see Figures 8.5 A and C.7 A). Third, the intermediate can be observed individually in rapidly-pulled constant velocity experiments with slightly lower unfolding stability compared to the completely native M domain (see Figures 8.5 B and C). Finally, a truncation mutant, which lacks 10 aminoacids at the C-terminal end, decreases the stability of the intermediate, which leads to the reduction of the overall folding rate (see Section C.3.4). The intermediate destabilization is in good agreement with the structural interpretation given in Section 7.3, proving that the intermediate's structure corresponds to the C-terminal subdomain of the M domain (residues 444–527). The influence of this on-pathway intermediate on the overall folding properties is further discussed in Section C.3.5.

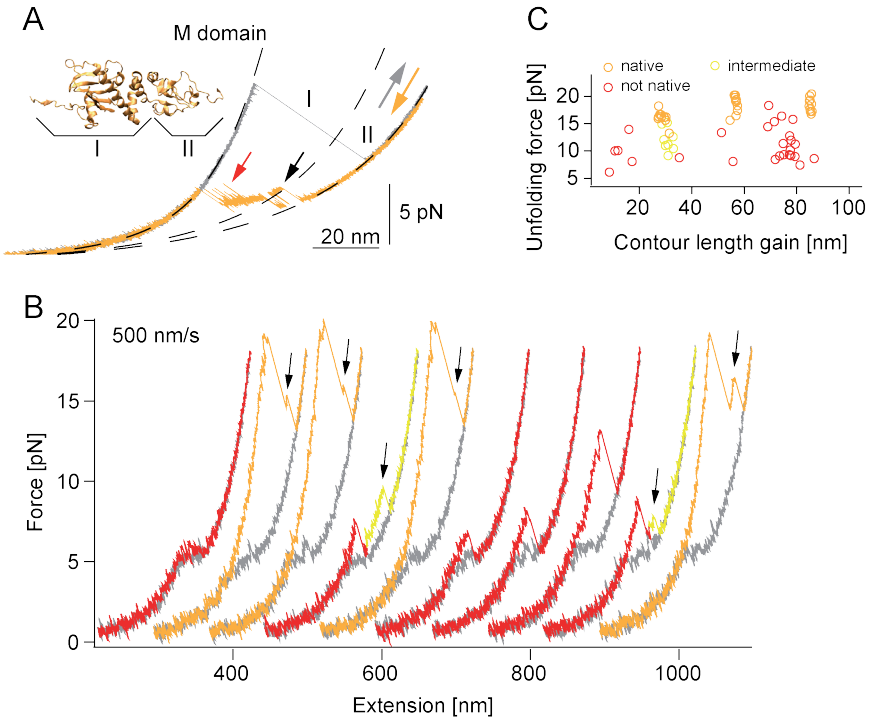

Figure 8.5: (A) Unfolding (gray) and refolding (orange) of M-diUbi (Inset from pdb id: 2cg9) at a pulling speed of 10 nm/s. The refolding protein populates many transient intermediates (red arrow) and a stable refolding intermediate (black arrow) that corresponds to the C-terminal subdomain (region II in inset). WLC fits to the unfolding trace are shown as dashed lines and additional traces are shown in Figure C.7.
(B) Consecutive, offsetted cycles at a pulling speed of 500 nm/s. Natively-folded unfolding traces are shown in orange and non-native ones in red. If just the stable M domain intermediate is folded (black arrows), it is colored in yellow. Refolding traces are in gray.
(C) A scatterplot of the unfolding events obtained from 53 unfolding traces of an individual molecule at a speed of 500 nm/s. Native unfolding events are colored in orange. Depending on the lifetime, the classification algorithm (see Section B.2.2) either picks up the stable intermediate or not. Therefore native events with the full contour length or native events consisting of two unfolding events with shorter contour length gains are observed. If only the intermediate is folded, markers are colored yellow. Apart from these events, many non-native events (red) with different lengths and stabilities are present.

8.6.3 *Force-dependent folding experiments*

To determine whether the unstable intermediates of the M domain effect folding similarly as the intermediates of the N domain, the force-dependence of the refolding process is evaluated using double-jump ramp experiments, as described in Section 8.4. Figure 8.6 shows a 3D plot of the *force-dependent refolding probabilities* as well as simplified projections where the probabilities are plotted versus refolding time or refolding force.

For raw double-jump ramp data of the M domain see Figure C.8.

The refolding probabilities at almost zero force, displayed as orange markers in Figure 8.6 B, show that after 3 seconds almost 70% of the molecules have refolded. Therefore the effective folding rate is around $0.34\,s^{-1}$. The refolding rates increase with increasing force up to 3.5 pN, identifying the transient intermediates as misfolded states. This behavior is identical to that observed for the N domain, only the effect by small forces on the refolding probabilities of the M domain is even higher (compare Figure 8.6 C to Figure 8.4 C), suggesting more stable off-pathway intermediates.

8.6.4 *M domain model*

The observations made for the M domain in constant velocity traces and force-dependent refolding probabilities show that folding is drastically slowed down by an *ensemble of misfolding intermediates* as it is the case for the N domain. Therefore the three-state model with the unfolded, misfolded and native states presented in Section 8.5.2 is also applied to quantify the M domain refolding probability data.

The best fit displayed in Figures 8.6 A–C shows the applicability of the model for the M domain. The transition rate between the unfolded states and the folded state $k_{U\to F}$ is $7651(\pm714)\,s^{-1}$. The free energy difference between the unfolded and misfolded states E_{UM} is $9.9(\pm0.1)\,k_BT$. The model and determined numbers are summarized in Figure 8.6 D. Lowest χ^2 is obtained by fixing the contour length of the misfolded states x_M to 37 nm and the position of the transition state between the unfolded and the folded states x_T to 46 nm.

By fitting WLCs to multiple molecules, average DNA and protein parameters are determined to calculate the energy contributions of the system (see Section 6.3). For the model following parameters are used: $p_{DNA} = 17.0\,nm$, $L_{DNA} = 371.1\,nm$, $K = 656\,pN$, $p_{prot} = 0.7\,nm$, contour length of the folded state $x_F = 0\,nm$, contour length of the unfolded state $x_U = 86.5\,nm$ and $k_{eff} = 0.174\,pN/nm$.

The misfolding free energy determined by the model is even higher for the M domain than for the N domain. If these misfolds are excluded from the model, the force-dependent refolding probabilities would follow the dashed lines shown in Figures 8.6 B and C.

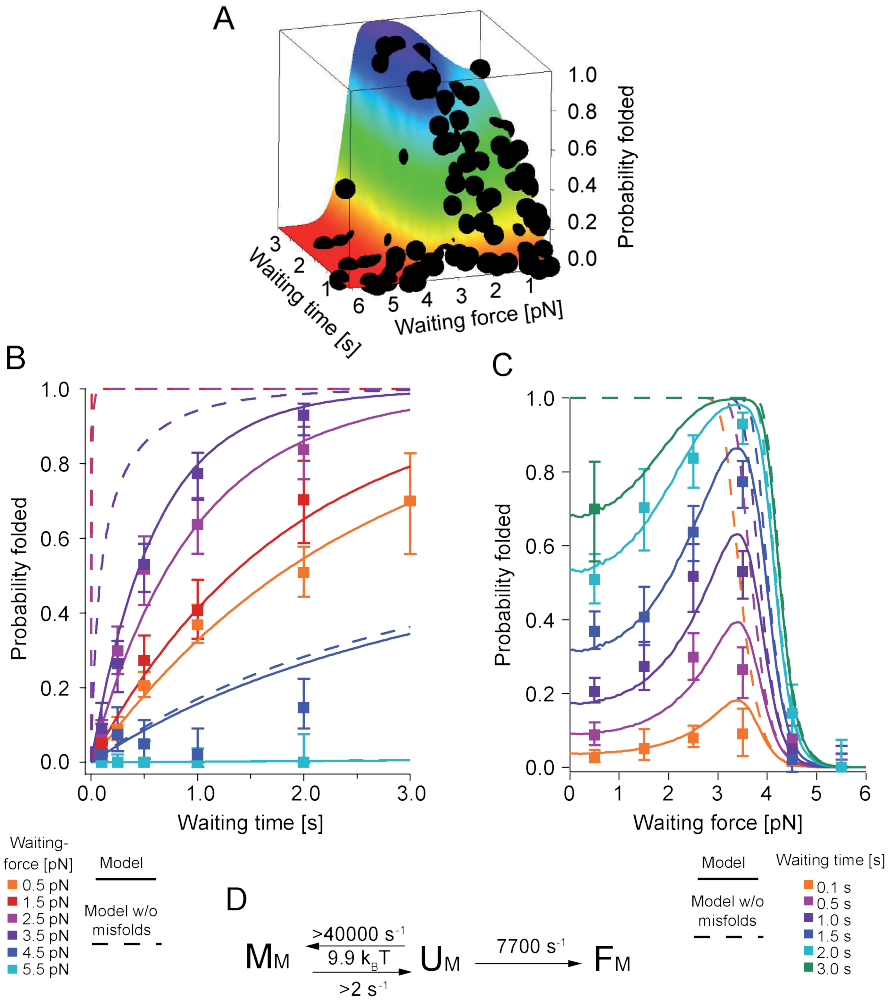

Figure 8.6: (A) Refolding probabilities dependent on the waiting time and the waiting force of M-diUbi obtained from double-jump ramp experiments are shown as black dots. They are determined from 7,927 traces of 14 molecules. The colored surface shows the best fit of the model described in Fig. D.
(B, C) The same data are shown in different representations with averaged probabilities. Fig. B shows the probabilities depending on the waiting time with the waiting force color-coded. Fig. C shows the probabilities depending on the waiting force with the waiting time color-coded. For uncertainties see Section B.2.3. The increase of the refolding probability with increasing waiting force marks the transient intermediates as misfolds. For solid and dashed lines see Fig. D.
(D) A three-state model quantitatively describes the observations. In this model unfolded states [U] are in equilibrium with misfolded states [M]. The native state [F] can only be accessed from the unfolded states [U]. Numbers displayed are derived from the fit of Equation 8.1 to the data and estimates made in the text. The fit is shown as a colored surface in Fig. A and as solid lines in Figs. B and C. If the misfolded states would be absent, the refolding probability would follow the dashed lines in Figs. B and C, showing the severe influence of the misfolded states at low forces.

The misfolding free energy can also be estimated from the slowly-relaxed refolding traces (see Figures 8.5 A and C.7 A). Since the transitions are very similar to those found for the N domain (see Section 8.5.3), an energy of around 10 k_BT is expected, which is in good agreement with the value obtained from the model.

The heterogeneity of the intermediate states is also observed in constant distance experiments during the waiting time (see Figure C.8). However it is more difficult to estimate the misfolding rate from these experiments because the on-pathway and the off-pathway intermediates mix. Hence a different approach is used: The measured refolding probabilities shown in Figure 8.6 B, don't significantly deviate from single exponential fits. Therefore the unfolded and misfolded states are already in equilibrium for waiting times greater than 0.5 s. This implies that both rates of the process are faster than $2\,s^{-1}$. Since the average free energy of the misfolded states is 9.9 k_BT, the misfolding rate $k_{U\to M}$ can be estimated to $\geqslant 40,000\,s^{-1}$ using Equation 6.16. With these rates the full system is simulated and refolding traces can be reproduced (see Figure C.7).

Using the same approach for the N domain yields a misfolding rate of $\geqslant 2960\,s^{-1}$, which is in good agreement with the estimate found in Section 8.5.3.

8.7 FOLDING OF THE C DOMAIN

In this section the refolding properties of the isolated C domain are elucidated. With 172 aminoacids it is the smallest domain of Hsp90. Around one third of the domain is unstructured, as detailed in Section 7.4. However, the folded part is still larger than the typical average domain size of around 100 aminoacids [158].

8.7.1 *The C domain shows two-state folding behavior*

Figure 8.7 A shows slowly-pulled force-extension traces of the isolated C domain. At forces around 4 pN the protein can unfold and refold multiple times during stretching and relaxation. At higher pulling speeds (see Figure 8.7 B), the C domain always folds into its native state, indicating a fast refolding rate. Unlike the N and M domains, the C domain doesn't show any stable refolding intermediates (see Figure 8.7 C). Only the folded or unfolded states are populated at the forces investigated, which is characteristic for a *two-state folder*.

8.7.2 *Force-dependent folding experiments*

The C domain's refolding behavior is very different to those of the N and M domains. To compare them quantitatively the double-jump ramp experiments (see Section 8.4) are repeated for the C domain. Figure 8.8 A shows the refolding probabilities depending on waiting time and waiting force as black dots. For Figures 8.8 B and C proba-

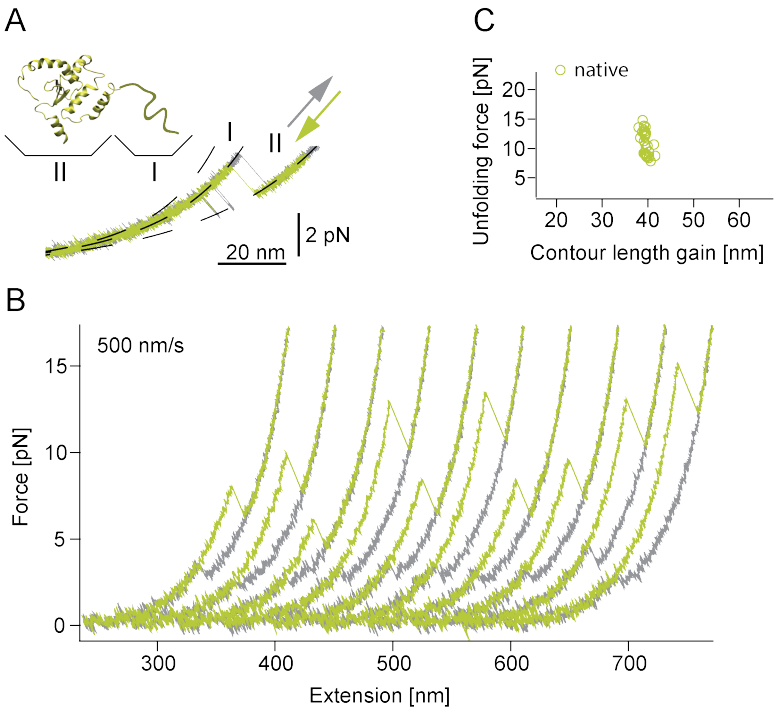

Figure 8.7: (A) Unfolding (gray) and refolding (green) of C-diUbi (Inset of pdb id: 2cg9) at a pulling speed of 10 nm/s. WLC fits are drawn as dashed lines and consider the unfolded portion of the C domain (region I). More force-extension traces are displayed in Figure C.11.
(B) Consecutive, offsetted unfolding traces (green) at a pulling speed of 500 nm/s always show the native pattern. The transition from the unfolded state into the native state is also resolved in the refolding traces (gray).
(C) A scatterplot of the unfolding events obtained from 35 traces of an individual molecule shows a well defined native population without any intermediates.

bilities are pooled and plotted versus waiting time and waiting force, respectively.

In agreement with the constant velocity experiments, fast folding rates are found. Even for very short waiting times the refolding probability is close to one at low waiting forces. For longer waiting times ($> 0.1\,s$) the unfolding rate becomes more and more dominant and the refolding probability saturates at its time-independent equilibrium value (see Figure 8.8 B). In contrast to the N and M domains, the C domain shows decreased refolding rates for increasing waiting forces in all regimes, as expected for a system that is not influenced by misfolds.

8.7.3 *C domain model*

The experiments indicate that the C domain is a two-state folder that unfolds and refolds at forces of interest. Therefore the folding behavior is investigated with a model, comprising an unfolded state [U] and a folded state [F] with the respective force-dependent transition rates. The refolding probabilities expected for such a system are described by Equation 6.24 in Section 6.7.3. The force-dependent transition rates are described by the model in Section 6.6.3, which depends on the zero-force rate and the distance to the transition state. Since the transition state position should be the same for unfolding and refolding (see Section 6.5.1), the model contains three parameters.

The fit yields zero-force folding and unfolding rates of $218(\pm16)\,s^{-1}$ and $0.154(\pm0.029)\,s^{-1}$, respectively. Fixing the transition state position x_T to 44 nm gives the lowest value of χ^2. The fit is displayed in Figures 8.8 A–C and the results are summarized in Figure 8.8 D.

The average DNA and protein parameters to calculate the energy contributions of the dumbbell system used in the fit are: $p_{DNA} = 17.9\,nm$, $l_{DNA} = 367.4\,nm$, $K = 711\,pN$, $p_{prot} = 0.7\,nm$, contour length of the folded state $x_F = 22.0\,nm$, contour length of the unfolded state $x_U = 62.7\,nm$ and $k_{eff} = 0.177\,pN/nm$.

8.7.4 *Equilibrium experiments*

In constant velocity experiments the C domain can unfold and refold on the timescale of the experiment (see Figure 8.7 A). Therefore the system is perfectly suited for constant distance experiments (see Section 5.3), where the C domain fluctuates between the unfolded and the folded states at a given pre-tension.

Figure 8.9 A shows example traces of the isolated C domain for different trap distances. Equilibrium transitions between the folded state at higher forces and the unfolded state at lower forces are observed. Increasing the trap distance and therefore the pre-tension shifts the equilibrium from the native to the unfolded state. All data points are

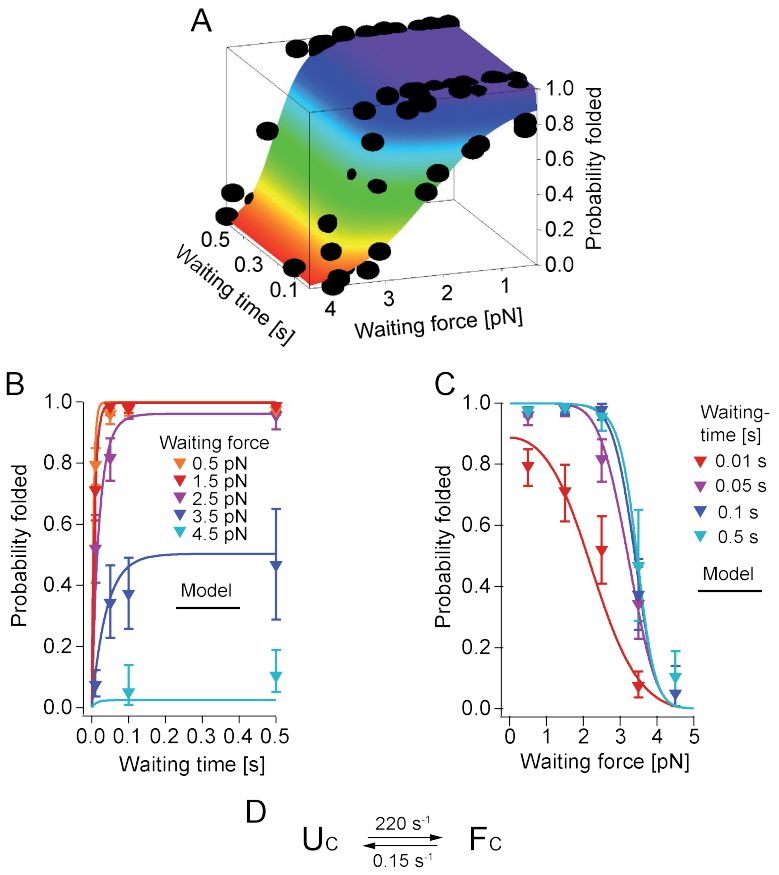

Figure 8.8: (A) Time- and force-dependent refolding probabilities of C-diUbi obtained by double-jump ramp experiments are shown as black dots. They are determined from 6,157 traces of 9 molecules. The colored surface shows the best fit using the model described in Fig. D.
(B, C) The same data are shown in different representations with averaged probabilities. Fig. B shows the probabilities against the waiting time with the waiting force color-coded. Fig. C shows the probabilities against the waiting force with the waiting time color-coded. For uncertainties see Section B.2.3. For solid lines see Fig. D.
(D) A two-state model with force-dependent folding and unfolding transition rates, is used to describe the data. Numbers presented are the result of fitting Equation 6.24 to the refolding probabilities. The best fit is shown as a colored surface in Fig. A and as solid lines in Figs. B and C.

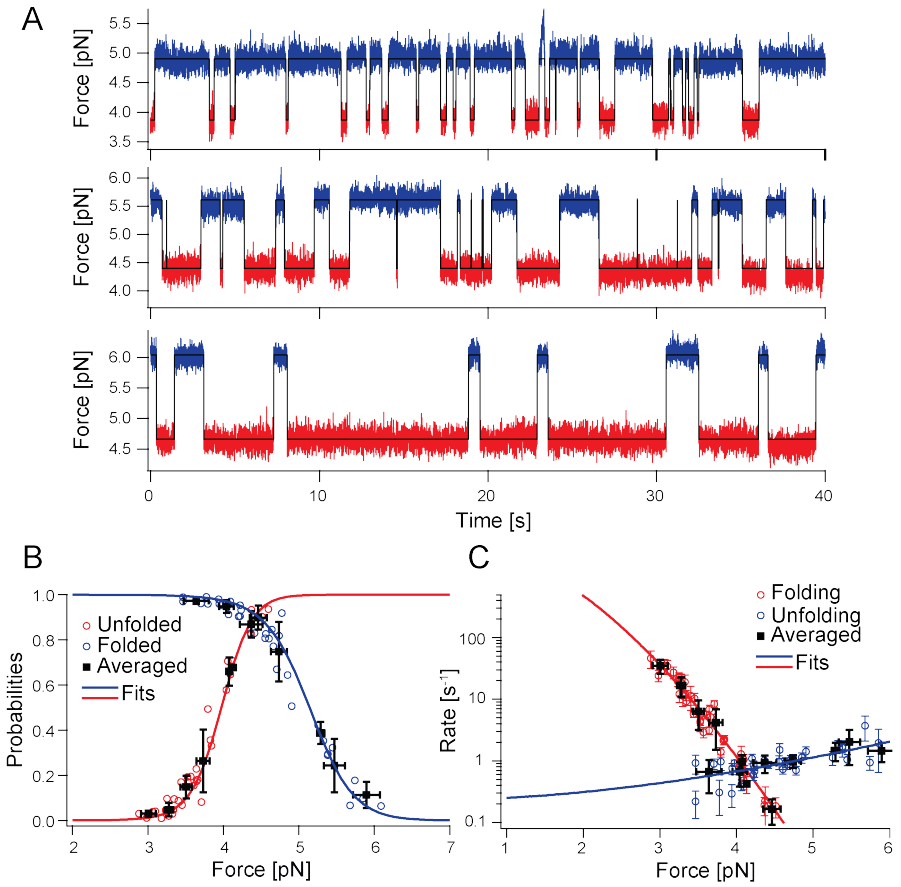

Figure 8.9: (A) Example constant distance traces of C-diUbi allow the observation of the folding kinetics in real time. If the trap distance is increased (top to bottom), the average pre-tension rises and the populations are shifted from the folded state (blue) to the unfolded state (red).
(B) Individual (red and blue markers) and averaged (black markers) force-dependent population probabilities of both states derived from constant distance traces of 15 different molecules. The occupancies are well described by a global fit (dashed lines) with a folding energy of around $10\,k_BT$.
(C) Folding and unfolding rates are determined from the lifetimes of multiple experiments. Fitting the model (solid lines) described in Section 6.6.3 to the averaged data (black markers) permits extrapolation to zero force. The folding and unfolding rates are $3700\,s^{-1}$ and $0.23\,s^{-1}$, respectively.

assigned to one of the two states by HMM, as described in Section 6.2.

From such *constant distance trajectories*, the occupancies and the lifetimes of both states at different forces can be read out directly (see Chapter 6). They are shown in Figures 8.9 B and C. Because the transitions are relatively slow, it is difficult to collect meaningful lifetime histograms for one single molecule for the accessible force range in one experiment. Therefore force-dependent probabilities and rates of 15 different molecules are pooled.

The population probabilities of both states are fitted globally using Equation 6.11 and the folding free energy is determined to $10.4(\pm0.1)$ k_BT. Both rates are modeled using Equation 6.19 and zero-force folding and unfolding rates of $3700(\pm1600)\,s^{-1}$ and $0.23(\pm0.16)\,s^{-1}$, respectively, are found. Additionally, the fit yields the contour length differences from the initial states to the transition states, which are $35.9(\pm3.8)$ nm for folding and $5.9(\pm3.0)$ nm for unfolding. Their sum matches the contour length of the folded C domain (see Table 7.3), which supports the validity of the rate fits. However, the folding rate is around an order of magnitude higher than the one derived in double-jump ramp experiments (see previous Section 8.7.3). This discrepancy is revisited in the discussion.

To fit the probabilities and the rates average DNA and protein parameters of all investigated proteins are calculated: $p_{DNA} = 12\,nm$, $l_{DNA} = 375\,nm$, $K = 650\,pN$, $p_{prot} = 0.7\,nm$, protein contour length of the folded state $x_F = 18.0\,nm$, protein contour length of the unfolded state $x_U = 62.4\,nm$ and $k_{eff} = 0.182\,pN/nm$.

Furthermore, very similar rates and energies for the C domain are obtained from an independent experiment using the full length Hsp90 construct (Hsp90-diUbi), which can also be held at constant distance at forces where the C domain unfolds and refolds, while the N and M domain remain folded (data not shown).

8.8 CROSS-DOMAIN MISFOLDING

As found in Section 8.1, slowly relaxed monomer chains always reach the completely native state, but rapidly relaxed ones don't. After the folding properties of the isolated domains are well characterized and understood, the question arises, whether the refolding properties of the Hsp90 monomer are described by its individual domains or if the *presence of neighboring domains* influences the folding process.

To address this question, the time- and force-dependent refolding probabilities of the Hsp90 monomer are investigated, employing the double-jump ramp protocol (see Section 8.4). Figure 8.10 compares scatterplots of the unfolding events observed at two different refolding forces after a waiting time of 1 s. Examples of individual force-extension traces are shown in Figure C.13.

For the monomer this protocol is especially useful, since the unfolding ramp allows the determination of how many and which domains have refolded.

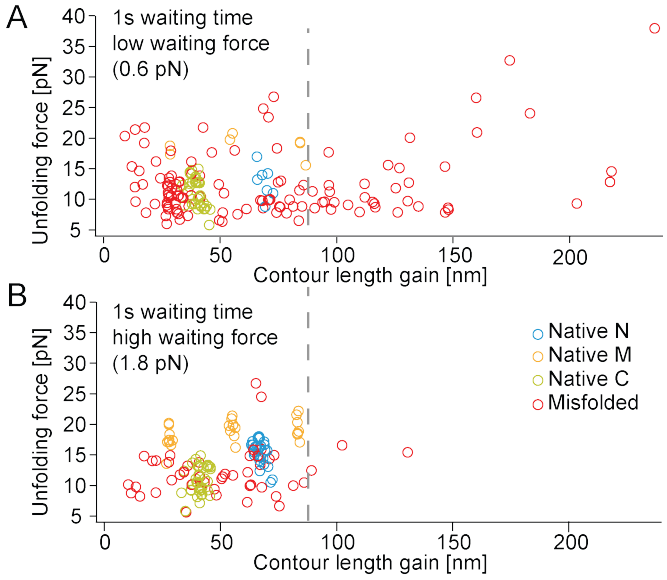

Figure 8.10: Example scatterplots of unfolding forces against contour length gains obtained in double-jump ramp experiments with the same Hsp90-diUbi molecule. The protein is given 1 s to fold at a low (A) and a high (B) waiting force. The scatterplots are determined from 60 and 45 unfolding traces, respectively. Blue, orange, and green circles mark the unfolding events of the native N, M, and C domains, while red markers show non-native events. The misfolded events vary significantly in contour lengths and stabilities and many of them are even longer than the contour length of the largest Hsp90 domain (to the right of the gray line), suggesting misfolds across domains. At higher waiting forces, as shown in Fig. B, the fraction of misfolded to native events is greatly reduced and cross-domain as well as high-stability misfolds are almost absent. The stable M domain intermediate (at around 30 nm), which is significantly populated at low waiting force is also colored red for simplicity. More scatterplots at low and high refolding forces for different waiting times are shown in Figure C.14 B.

As expected from the isolated domain experiments (see Sections 8.5.1 and 8.6.3), more natively-folded N and M domains are observed at higher refolding forces. However, a novel feature are the non-native intermediates with contour length gains longer than the largest Hsp90 domain (M domain, 68 nm). Their lengths demand that two or even all three domains of Hsp90 are involved in the formation of these intermediates. These cross-domain intermediates are preferentially populated at low waiting forces and almost absent if the refolding force is increased. This suggests that small forces can successfully prevent cross-domain misfolding by keeping distant regions of the unfolded polypeptide chain apart.

In addition, the misfolds observed for the monomer are more stable than those of the isolated N and M domains (compare Figure 8.10 A to Figures 8.2 C and 8.5 C). Some of them don't even unfold during a single unfolding ramp and multiple cycles are needed to disintegrate the intermediate (for examples see Figure C.14 A).

To investigate the time dependence of the refolding probability of the individual domains within the monomer qualitatively, scatterplots like those displayed in Figure 8.10 are measured for waiting times of 0.5 s and 3 s for both low and high refolding forces (see Figure C.14 B). Therefrom refolding probabilities of the individual domains for different waiting times at low and high force are calculated, which are shown in Figure C.14 C. The N and M domains' refolding probability doesn't change significantly over this time range at low forces, but at higher forces a steep increase is observed. Also the C domain seems to be involved in forming cross-domain misfolds, because its refolding probability is significantly reduced (down to 50%), although almost 100% refolding efficiency is expected from the isolated domain.

8.8.1 *Comparison of the monomer with the isolated domains*

To quantify the impact of the cross-domain misfolds on folding speed of the Hsp90 monomer, the probabilities of observing the completely native Hsp90 in the unfolding ramp after the double-jump are evaluated for different refolding times and forces. Figure 8.11 shows averaged refolding probabilities plotted versus waiting time for a low ($0.3 - 0.8\,\mathrm{pN}$) and a high ($1.8 - 2.2\,\mathrm{pN}$) waiting force range. Non-averaged data are shown in Figure C.14 D.

Refolding into the completely native state is very slow for the low force range, but already small forces can speed up folding significantly. The overall rates, characterized by single exponential fits to the time-dependent probabilities, are $0.024 \pm 0.013\,\mathrm{s}^{-1}$ for low forces and $0.54 \pm 0.07\,\mathrm{s}^{-1}$ for high forces. The small additional force speeds up the overall folding rate by a factor of 25, although folding is more difficult at higher forces. This factor is much larger than the one ob-

For the N and M domain it is only a factor of around 1.4 and 2.1, respectively.

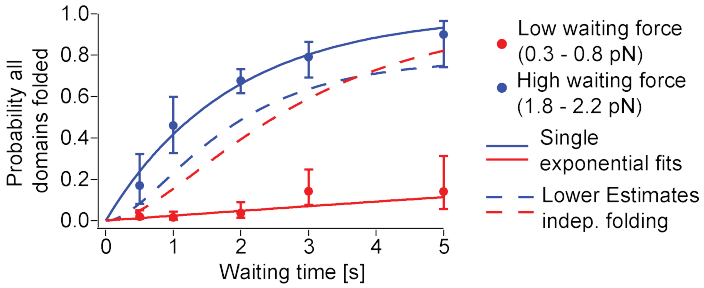

Figure 8.11: Time-dependent probabilities of observing the completely refolded Hsp90 monomer (Hsp90-diUbi) in double-jump ramp experiments. Red and blue dots mark the average refolding probabilities for a low and a high waiting force range, respectively. Data are from 1,067 unfolding traces of 10 different molecules, for uncertainties see Section B.2.3. At low forces even long refolding times yield poor refolding behavior, while at slightly higher forces folding is greatly accelerated. Single exponential fits are shown as solid lines. To compare the measured refolding probabilities to those expected from the isolated domain constructs, lower limit estimates of the refolding probabilities are shown as dashed lines. For high refolding forces the estimate lies below the measured probabilities, indicating independent folding and only little influence of cross-domain misfolds. In contrast, at low forces, the estimate lies well above the data, revealing severe influence of cross-domain misfolds.

served for the refolding of the isolated domains (see Figures 8.4 B and 8.6 B), highlighting the tremendous assistance of force in *avoiding cross-domain misfolds*.

To show this more rigorously, the expected minimal refolding probabilities assuming independent folding of the individual domains are estimated for both force ranges. Since the waiting force refers to the force of the completely unfolded peptide, and folding of the domains increases the applied force, while the protein is held at constant trap distance the force ranges accessible to the proteins are determined from the average protein and DNA parameters. For the lower force range ($0.3 - 0.8\,\mathrm{pN}$) the accessible force range is $0.3 - 0.99\,\mathrm{pN}$. For the higher force range ($1.8 - 2.2\,\mathrm{pN}$) the accessible force range is $1.8 - 3.13\,\mathrm{pN}$. For a *lower limit estimate* the smallest refolding probabilities of all isolated domains are taken in the respective force range. Therefore the minimal refolding probability $P_{NMC}(t)$, assuming independent folding of all domains (N, M, C), at low forces is

$$P_{NMC}(t) = P_N(t, F = 0.3\,\mathrm{pN}) P_M(t, F = 0.3\,\mathrm{pN}) P_C(t, F = 2.2\,\mathrm{pN}) \quad (8.3)$$

and similarly at high forces:

$$P_{NMC}(t) = P_N(t, F = 1.8\,\mathrm{pN}) P_M(t, F = 1.8\,\mathrm{pN}) P_C(t, F = 3.13\,\mathrm{pN}) \quad (8.4)$$

The estimates are directly calculated from the isolated domain models (see Sections 8.5.2, 8.6.4 and 8.7.3) and are shown as dashed lines in Figure 8.11. For high refolding forces the lower estimate lies below the measured probabilities, indicating independent refolding and no or little effect of the cross-domain misfolds, while at low forces the lower estimate is well above the data, showing severe disturbance of folding by the stable cross-domain misfolds.

8.9 DISCUSSION

Interesting insights into protein folding are obtained by studying the *multifaceted folding properties* of the large three-domain protein Hsp90 with optical tweezers. The highlights are the reduction of the folding speed by misfolding intermediates of the large N and M domains, the on-pathway intermediate of the M domain, which speeds up folding, the effect of the cross-domain misfolds on multi-domain protein folding and the fact that applied force can speed up protein folding. These points are discussed here and related to the existing literature. Ultimately, the impact on *in vivo* protein folding is elucidated.

8.9.1 *Folding of the individual domains and intradomain misfolds*

Protein folding is typically studied for proteins with only one domain [23, 66, 169, 125]. Some of these proteins, especially the smaller ones, show two-state folding behavior. Others populate kinetically stable intermediates, which can be on-pathway or off-pathway and happen during an early or a late stage of the folding process. Whether an observed intermediate promotes or hinders protein folding and its structural bases are difficult to investigate and hence remain mostly unclear [151]. However, the investigation of exactly such intermediates is essential to further study protein folding and to understand protein aggregation, which is relevant for many diseases [15].

Hsp90 has three *non-homologous domains*, which are all studied independently. The smallest Hsp90 domain, the C domain with slightly more than 100 structured residues, folds in a two-state manner within milliseconds without populating any detectable intermediates. The folding behavior of the larger N (211 residues) and M domains (266 residues) is dominated by a *broad range of intermediates*. Measuring the force-dependent folding rates revealed that these intermediates are off-pathway. Therefore the entire refolding behavior is quantified by a three-state model comprising unfolded, misfolded and native states. The determined average free energies of the misfolded states are considerable, $7.3\,k_BT$ for the N and $9.9\,k_BT$ for the M domain. The population of the misfolded states reduces the folding speed by a factor of 1000 to 10000, from $950\,s^{-1}$ to $0.5\,s^{-1}$ for the N domain and

from 7700 s^{-1} to 0.34 s^{-1} for the M domain. All folding rates determined are summarized in Table 8.1.

DOMAIN	EFFECTIVE RATE	DIRECT RATE
N	0.5 s^{-1}	950 s^{-1}
M	0.34 s^{-1}	7700 s^{-1}
C (double-jump)	n.a.	220 s^{-1}
C (const. dist.)	n.a.	3750 s^{-1}

Table 8.1: Refolding rates of the isolated Hsp90 domains. Effective rates for the N and M domains include the effect of misfolds, while the direct rates show the rate from the unfolded states to the native states, which is the rate if misfolds are absent. For the C domain the measured rates for double-jump ramp experiments and constant distance (CD) experiments are displayed. Possible reasons for the slight deviation are discussed in Section 8.9.1.2.

The misfolded states of the N and M domains are very dynamic (especially under force) and get populated rapidly. Estimates suggest misfolding rates greater than 12,600 s^{-1} for the N and 40,000 s^{-1} for the M domain, which is in agreement with the concept that only the native or misfolded states are observed, but never the completely unfolded peptide. A similar competition between folded and misfolded states has already been observed in optical tweezers experiments with calcium binding proteins NCS-1 [47] and calmodulin [139]. The off-pathway intermediates observed for these molecules are very few (1-2) and have characteristics of defined states. This is clearly different for the N and M domains of Hsp90 that show large heterogeneity of the misfolded conformations with different contour length and kinetics. Because of this heterogeneity a structural interpretation is difficult and it cannot be excluded that the misfolded states contain many native contacts. Under force these misfolded states are very dynamic and interconversion between different misfolded states is possible.

These small proteins consist of 2 domains with 2 EF-hands each.

For the three isolated Hsp90 domains the effect of misfolding intermediates increases with domain size. It stands to reason that the number of misfolds observed is to some extent correlated to the length of the polypeptide chain. The misfolds observed for the N and M domain are likely the same that other groups report as heterogeneous burst-phase intermediates in refolding bulk experiments of larger proteins like MBP and Tim barrel protein [153, 165]. These burst-phase intermediates can manifest in roll-overs, which are a common and little-understood feature in the chevron plots of folding experiments with chemical denaturants [8].

Of great interest is also the effect of the stable M domain on-pathway folding intermediate. Destabilization of the intermediate leads to impaired folding of the complete M domain. However, the slightly reduced folding rate of this intermediate is likely not responsible for

the impaired folding of the complete domain. Rather the propensity to form misfolds with the remaining part of the M domain is favored, which can lead to a much stronger decrease of refolding speed. This suggests a possible role for on-pathway intermediates, that form to avoid off-pathway intermediates.

8.9.1.1 *Native structure and folding speed*

The folding behavior of the isolated domains is examined in great detail and it is possible to calculate the direct rates from the unfolded states to the native states without misfolds. Possible determinants of the direct rates like the native structure are discussed here.

The *structural complexity* of a protein can be described by its contact order (CO), a measure of the average sequence distance between native contacts. The contact order is typically used to predict the folding speed of two-state folders [106]. Table 8.2 compares the predicted folding rates calculated from the individual domains of the crystal structure (pdb id: 2cg9) to the measured ones without misfolds.

Contact order and predicted rates are calculated using a script freely available on the Baker lab website.

DOMAIN	CO	RELATIVE CO	PRED. RATE	MEAS. RATE
N	21.6	0.103	$6548\,s^{-1}$	$950\,s^{-1}$
M	12.5	0.0488	$325136\,s^{-1}$	$7700\,s^{-1}$
C	10.5	0.0744	$51328\,s^{-1}$	$220\,s^{-1}/3750\,s^{-1}$

Table 8.2: Contact order, relative contact order and predicted rates compared with measured rates. For the N and M domains the direct rates without misfolds and for the C domain the refolding rates determined by double-jump ramp and constant distance experiments are shown.

The N domain shows a high relative contact order reflecting the complex fold, which is described in Section 2.1. The M domain consists of multiple small subdomains, which are arranged linearly (see Section 2.1). Therefore the M domain has a much lower relative contact order than the N domain. The complexity of the C domain lies in between these extremes.

The rates predicted generally overestimate the rates observed. For the N and M domains this is not surprising because the rate prediction is based on proteins smaller than 100 residues and the CO is normalized by the numbers of residues (relative CO). In addition the N and M domains likely populate short-lived on-pathway intermediates during folding, as the transition state positions are not very close to the folded states (see Sections 8.5.2 and 8.6.4). Qualitatively, the predicted rate is faster for the M domain which has a much simpler architecture and is divided into two independently folding subunits, which might explain the fast direct folding rate for the large protein

On pathway intermediates are also suggested by the refolding transitions observed at constant trap distance (see Figures C.5 and C.8).

domain. Apart from that the direct folding rate of the N domain is not far from the rate predicted.

Since the C domain folds in a two-state manner, the contact order prediction should be valid, although slightly different rates between the double-jump ramp and the constant distance experiments are measured. Nevertheless the CO prediction overestimates the refolding rate by 1 to 2 orders of magnitude.

A deviation from the CO prediction on this order is not uncommon [106].

The comparison shows that contact order may capture some determinants of the folding speed but cannot fully describe it. Also chain length and free energy of the protein should play an important role [9]. For similar off-rates the productive folding rate should increase with free energy. But higher free energies should also increase the propensity of misfolding. This is exactly what is observed for the limited dataset of three protein domains.

8.9.1.2 *Comparison of equilibrium and double-jump rates*

The refolding rate of the C domain determined by double-jump ramp experiments is $220\,s^{-1}$ (see Section 8.7.3), while in constant distance experiments a refolding rate of $3750\,s^{-1}$ is found (see Section 8.7.4). To investigate the *discrepancy* of one order of magnitude the force-dependent rates of both experiments are plotted on top of each other in Figure 8.12. Therefore the measured refolding probabilities of the double-jump experiments with a waiting time of 10 ms are converted into rates using Equation 6.23.

For the 10 ms waiting times the unfolding rate is negligible.

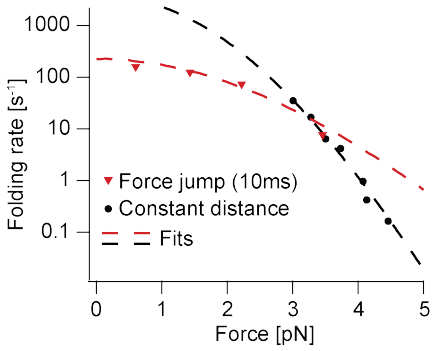

Figure 8.12: Red and black markers show the measured force-dependent folding rates obtained by double-jump ramp experiments and constant distance experiments. The respective fits are shown as dashed lines.

Double-jump ramp experiments explore the force dependence close to zero force, while constant distance experiments yield rates at higher forces. The measured data points overlap at 3.5 pN suggesting that both experiments give correct rates. However the curvature is too strong to fit the combined data with the model of Section 6.6.3. This is also seen in the difference of the transition state position. In constant distance experiment the transition state is much closer to the

native state (6 nm) than in double-jump ramp experiments (22 nm). Since the transition state position is typically close to the native state and the rate predicted by contact order is very fast, it is likely that additional processes slow down the folding rate at low forces. Scenarios, which can explain the observed effect, are listed in the following:

- A likely reason is the population of misfolded states at very low forces, similar to the N and M domain, but far less pronounced. These misfolds would have to decay fast ($\leqslant$ ms) under force, since no misfolded states are observed in the unfolding ramps. This theory is supported by the fact that Figure 8.8 C shows slightly lower refolding probabilities at 0.5 pN than 1.5 pN for all waiting times greater than 10 ms.
- The shift in the transition state position can also be explained by a pathway change, meaning that the protein folds along different pathways with different transition states depending on the force applied.
- Also non-equilibrium effects can't be fully excluded, for example new states introduced by force-jumps, as shown in [50].

8.9.2 *Cross-domain misfolds*

Most proteins, especially eukaryotic proteins, consist of multiple domains [44]. The organization of the protein chain into multiple domains is a key property for protein folding [159]. Domains can stabilize each other, speeding up folding [44], but cross-domain misfolds between identical or homologous domains can also greatly slow down protein folding. It is easy to imagine, that swapping of secondary structure for very similar domains can lead to stable misfolds, which has been reported for the long immunoglobulin domain chains of the muscle protein titin [12, 96], poly-fibronectin [103] and poly-ubiquitin [166]. How non-homologous unfolded proteins influence each other is unclear. To address these questions the refolding properties of the full-length, three domain protein Hsp90 are examined.

If the completely unfolded, 709-amino-acids polypeptide chain is relaxed rapidly to almost zero force, more misfolded conformations are formed. They have large heterogeneity in contour length and many of them exceed the length of the largest Hsp90 domain, indicating that misfolds involve multiple domains. These cross-domain misfolds are more stable than those observed for the isolated domains. Comparing the refolding probabilities of the monomer and the isolated domain shows the drastic effect on folding of these intermediates. The mere fact that the three domains are tethered together slows down the refolding process from seconds to minutes.

This is in agreement with the prediction made in the last section that with *growing chain length*, the number of misfolded states will

increase [9]. For small single domain proteins, nature may found a way to optimize the sequence for a smooth energy landscape without stable misfolds, but with ever growing length chain these misfolds will become inevitable. This suggests that for large proteins, such as Hsp90, *cross-domain misfolds are a general feature*, which limit their folding rates. To what extent intradomain misfolding triggers interdomain misfolding can not be answered at this point, but if one of the domains reaches its native state it doesn't contribute to misfold formation and can speed up folding indirectly. A *simplified energy landscape* [98] summarizing the folding properties of the Hsp90 monomer is shown in Figure 8.13.

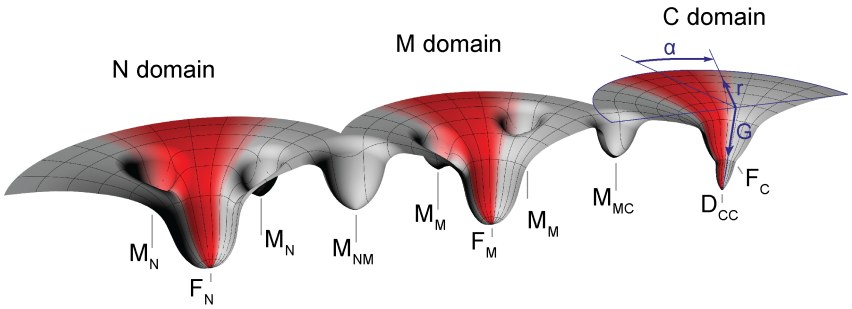

Figure 8.13: Simplified energy landscapes for folding of Hsp90. Since the Hsp90 domains don't stabilize each other directly, their folding process can be described by three individual energy landscapes with their global minima representing the native states [F_N, F_M and F_C]. The N and M domain populate intra-domain, off-pathway intermediates [M_N, M_M] which decrease refolding speeds from milliseconds to seconds. If the full-length monomer is unfolded, stable cross-domain misfolds [M_{NM}, M_{MC}] are formed, which further slow down folding to the minutes timescale. Small forces greatly change the energy landscape and favor folding. To explain the effect, the cylindrical coordinate system shown for the C domain applies for all domains. G refers to the free energy. The inverse of the radial coordinate describes the nativeness of the system. The inverse of the absolute value of the angular coordinate is interpreted as the average distance between residues in misfolded conformations. This distance depends on force and small forces can restrict the conformational search to the red shaded areas, helping to avoid misfolded intermediates. This effect is especially strong for the cross-domain misfolds. After folding of the C domains, Hsp90 can assemble into a functional chaperone [D_{CC}], as discussed in Chapter 10.

8.9.3 *Chaperoning by force*

Forces applied to proteins tilt the protein folding energy landscape, increasing the height of the transition state barrier and therefore decrease protein folding rates (see Section 6.6). In some experiments however, the at first glance impossible behavior of increasing rates with increasing forces is found. This observation can be reconciled by assuming off-pathway states, which trap the protein in folding-incompetent states hindering folding. Small forces can effectively depopulate these misfolds. Although the protein still has to fold against higher forces, the avoidance of misfolded states dominates the folding process. For the isolated N and M domains this effect speeds up folding by a factor of around two. It is more drastic for the full-length Hsp90, where small forces prevent the formation of cross-domain misfolds, which involve distant regions of the protein chain. An average force difference of only 2 pN can increase the refolding speed of the Hsp90 monomer by a factor of 25. Therefore force can "chaperone" the protein by *restricting the energy landscape* during the folding process. This is illustrated by the red shaded areas in Figure 8.13.

Notably, to observe this behavior the transition state position for native folding has to be closer to the unfolded state than the contour length of the misfolded conformations.

Generally, measuring the force dependence, is a powerful approach to get insights into complex systems. It can reveal processes and states that can not be identified otherwise. Correct modeling of the data allows the determination of rates that are otherwise inaccessible.

8.9.4 *Relevance in vivo*

The ribosome synthesizes the protein chain sequentially by appending one aminoacid after the other. In yeast, the synthesis rate is a few aminoacids per second [59]. Since all Hsp90 domains are able to fold within seconds, the individual domains can fold, in the absence of the not yet synthesized peptide chains of the other domains. This co-translational folding process [36] is a very effective way to avoid cross-domain misfolds. The translational speed can be modulated by the DNA sequence itself by rare codons that effectively stall the ribosome [148]. This could give domains that fold slower than Hsp90 additional time for folding. Interestingly, there is also evidence that folding of protein structures generates forces on the nascent chain, which can control the speed of the translation process [40]. In optical tweezers experiments co-translational folding can be mimicked by the application of small forces to the protein, keeping distant regions of the protein apart.

After Hsp90 is fully folded, the domains may transiently unfold, but the condition where the whole chain is unfolded and may misfold is very unlikely to happen again. Therefore Hsp90 is a very robust protein, which is somewhat expected for a protein that has to work under cellular stress conditions.

Chaperones are essential for the folding of many client proteins. Some of them, GroEl/GroES [134], ClpX [77] or Hsp70[24], can exert force by different means to unravel misfolded clients and favor their native folding, similarly as in optical tweezers experiments. Another mechanism for avoiding cross-domain misfolds is the stabilization of partially folded intermediates by chaperone binding, as shown for trigger factor [80].

9 HSP90'S CHARGED LINKER

Hsp90's domains are connected by linkers, which *facilitate the global rearrangements of the Hsp90 dimer* during its chaperone cycle. A unique, extended linker connects the N and M domains. It consists mainly of charged residues and is therefore named the Charged Linker (CL). This elongated linker is *exclusive to eukaryotic Hsp90s* and spans aminoacids 211-272 in yeast. While large parts are dispensable [75] and only lead to smaller deficiencies in client chaperoning, complete deletion is lethal *in vivo* [43]. Interestingly, a small artificial linker can revert lethality [43]. Although studied intensively *in vitro* and *in vivo*, the biological role of the Charged Linker is still far from understood [149, 157, 150, 173, 142].

The crystal structure of the yeast Hsp90 dimer (see Figure 2.1) could only be obtained by deleting a large part of the Charged Linker [2]. This and the fact that no structure is suggested by any biochemical assay led to the common assumption that the Charged Linker is unstructured. However, in mechanical experiments with Hsp90 a weak force signal associated with the Charged Linker is picked up. Using this as a starting point, the structural and dynamic properties of the Charged Linker are investigated in optical tweezers experiments. The effect of the CL on the large N-terminal conformational rearrangements is measured by single molecule FRET experiments. Together with other biochemical assays the Charged Linker and its influence on the Hsp90 chaperone is elucidated [54].

9.1 IDENTIFICATION OF THE CHARGED LINKER

The unfolding force-extension traces of the Hsp90 monomer shown in Chapter 7, show rapid transitions with significant contour length at very low forces prior to domain unfolding. Since all domains match the expected contour lengths (see Table 7.4), the Charged Linker is a likely candidate causing these transitions. To show this directly Hsp90-diUbi, which is described in detail in Figure 7.1, is compared to a Charged Linker deletion mutant (Hsp90ΔCL-diUbi; see Section A.1), as shown in Figure 9.1.

This is a similar deletion mutant as in [149].

The CL-dependent transitions shown in Figure 9.1 C seem to populate only two states. The one with shorter contour length at higher forces, which is observed as short-lived events, is a state where the CL's end-to-end distance is very short and the N and M domain are in close proximity. This state is called the *docked state* for reasons that will become clear in the next section. In the other state with longer

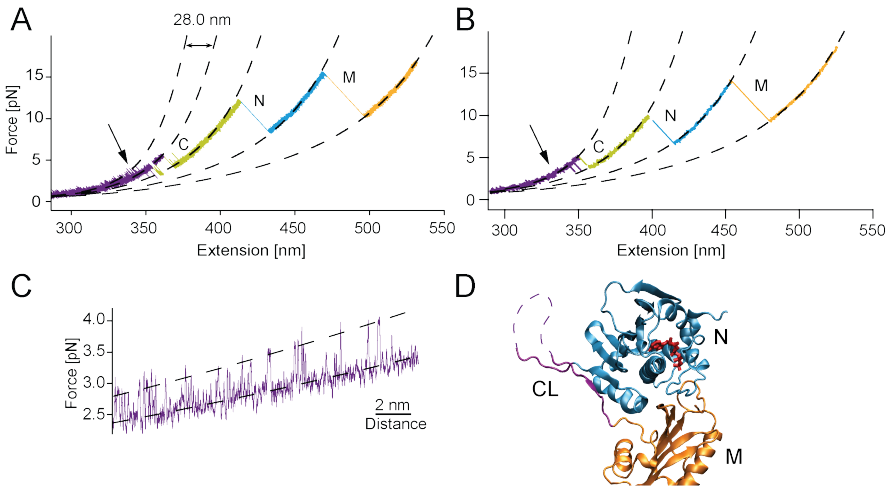

Figure 9.1: (A,B) Unfolding force-extension traces at a pulling speed of 10 nm/s of Hsp90-diUbi in Fig. A and Hsp90ΔCL-diUbi in Fig. B show identical domains (N, M and C) fitted with WLCs (dashed lines). However, the low force regions before the domains unfold marked by arrows show only transitions for the wild type. (C) An expansion of the low force region of Fig. A. suggest that two states get populated. (D) The Charged Linker region in the crystal structure (pdb id: 2cg9). The resolved aminoacids of the CL between the N domain (blue) and the M domain (orange) are colored in purple.

contour length, the Charged Linker is extended and behaves like an unstructured protein chain.

Fitting a WLC directly to the short-lived events of the docked state is difficult, hence the contour length is estimated. Assuming a contour length of 28 nm, which is approximately the length expected for the 62-aminoacids Charged Linker, fits well (see Figure 9.1 C).

The expected contour length of the CL is slightly lower, which can be explained by the reorientation of the domains and the changes in persistence length due to the charged nature (see Section 6.1).

9.2 THE CHARGED LINKER FORMS STRUCTURE

The measurements above show that a stable state exists, where the CL is not unstructured and the N and M domains are in spatial proximity. However, it is not obvious which parts of Hsp90 interact to stabilize this conformation. To get a structural interpretation of this state different CL substitution and deletion mutants are tested.

9.2.1 *Substitution mutants*

The crystal structure of the Hsp90 dimer in the twisted, closed state [2] suggests direct contacts between the N and the M domains (see Figure 9.1 D). Also the C-terminal part of the Charged Linker (aminoacids 264-272), which is still resolved, seems to associate with the N domain via a small beta-strand that proceeds the large beta-sheet

of the N domain. To test whether these interactions are eligible to explain the observed transitions, two Charged Linker substitution mutants, where aminoacids 211-263 (Sub211-263) or aminoacids 211-272 (Sub211-272) are replaced by glycine-glycine-serine (GGS) repeats of identical length, are prepared. These mutants have force-exertion points at aminoacid positions 61 and 560 like the construct Hsp90-_61C_560C shown in Section 7.2. For details on the constructs see Section A.1.

GGS repeats are very soluble and form random coils without any structure.

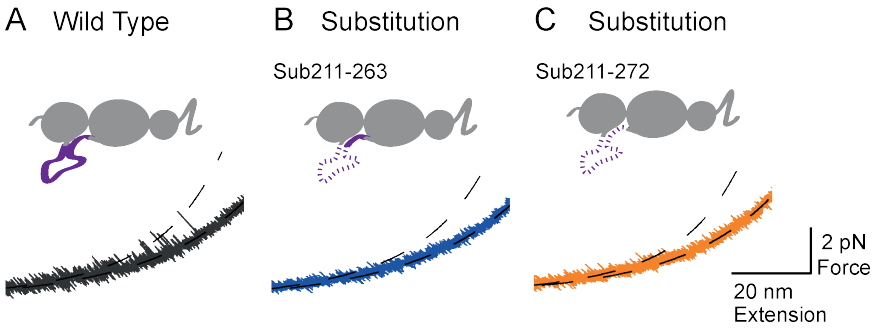

Figure 9.2: Close-up views to the low force regime of unfolding force-extension traces of Hsp90_61C560C (A) and the substitution mutants Sub211-263 (B) and Sub211-272 (C). While the wild type protein shows Charged Linker transitions, both supplementation mutants do not. Therefore neither the interactions between the N and M domains nor the interactions between the N domain and the C-terminal part of the Charged Linker facilitate the docked state. Full length traces as well as constant distance traces of the low force regions are shown in Figures 7.3 and C.15.

Surprisingly, *no stabilization between the N and M domains* is observed if the Charged Linker is substituted partly (Sub211-263) or completely (Sub211-272) by unstructured aminoacids (see Figure 9.2). Only the stretched, undocked conformation, where the tethered N and M domains can move freely apart is populated. In other words, neither the interactions between the domain surfaces, nor the interactions of the N domain with the C-terminal part of the Charged Linker are strong enough to explain the docked state. This implies that the sequence or at least residues between aminoacids 211 and 263, which are commonly considered as unstructured, are crucial for the observation of the docked state.

9.2.2 *Deletion Mutants*

The docked state is not facilitated by domain-domain interactions. Therefore two reasonable possibilities remain: Either the Charged Linker forms a stable domain itself, or it binds to the N or the M domain, or to both of them simultaneously. This question is resolved by two different Hsp90 domain-deletion mutants. The first comprises

only the Charged Linker and the M and C domains (ΔN), the second only the Charged Linker and the N domain (ΔMC). For details see Section A.1.

Figure 9.3 shows, that the construct, where only the N domain and the Charged Linker are present, shows similar transitions at low forces as the wild type. The possibility that the CL forms the docked state on its own, is ruled out by the other deletion mutant. Consequentially the docked state is the Charged Linker forming a *stable structural element together with the N domain*. Docking depends on some or all aminoacids ranging from 211 to 263, as follows from the previous section.

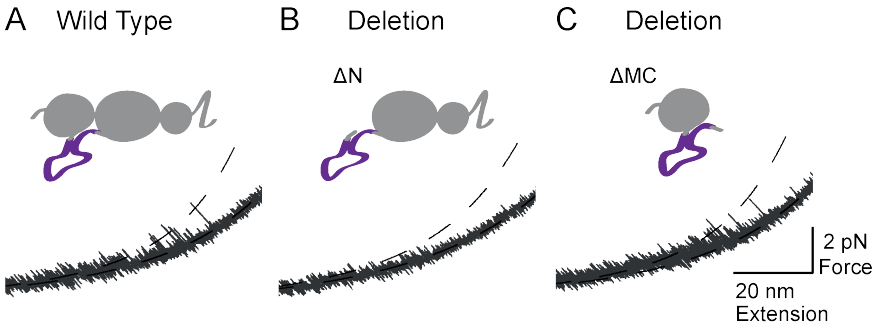

Figure 9.3: Zooms into the low force regimes of unfolding force-extension traces of the wild type construct Hsp90_61C_560C (A) and the domain deletion mutants ΔN (B) and ΔMC (C). If the N domain is absent, the docked state is not populated. But already the N domain with the Charged Linker alone shows the docked state. Hence the docked state results from the Charged Linker binding to the N domain. Full length traces as well as constant distance traces of the low force regions are shown in Figures 7.3 and C.16.

9.3 ENERGETICS AND KINETICS OF THE DOCKING TRANSITION

Already small forces strongly depopulate the docked state. To quantify the free energy of docking as well as the rates involved, constant distance experiments (see Section 5.3) are employed. Figure 9.4 A shows the force response of the Charged Linker in a typical experiment. Stepping up the trap distance, increases the average force applied to the molecule and shifts the equilibrium towards the undocked state. This is best seen in small sections of the trajectory at two different trap distances shown in Figure 9.4 B. Although the thermal noise is very large at these low forces, it is still possible to use HMM (see Section 6.2) to assign the datapoints to the two states observed. With this it is possible to directly read out the state occupancies at different pre-tensions. They are shown for a single example molecule in Figure 9.4 C. The energy difference between the docked and the undocked state is determined by globally fitting Equation 6.11 to

both probabilities. The energy contributions of the dumbbell system are calculated from the DNA and protein parameters of the respective molecule (see Section 6.3). Analysis of multiple molecules yields a *very low stabilization free energy* of $1.1\,k_BT \pm 0.4\,k_BT$ (34 molecules; 4315 seconds of analyzed constant distance traces; Error is SD between experiments). This energy is so small that even at zero force the molecule is able to populate the undocked state 25% of the time, while it stays docked 75% of the time.

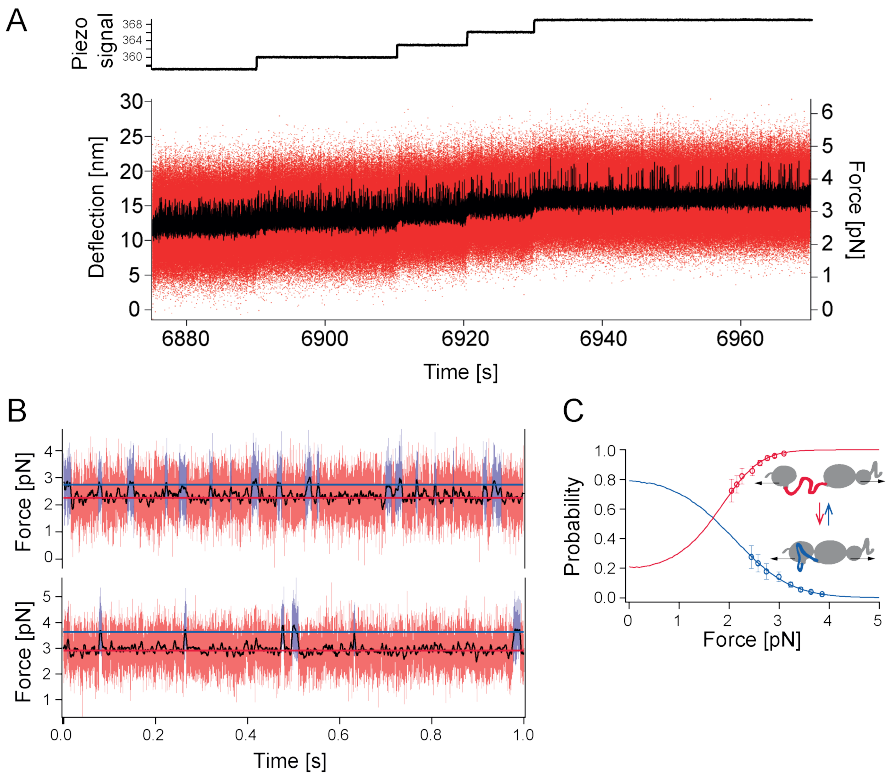

Figure 9.4: (A) A typical constant distance experiment at low forces with the Hsp90 monomer (Hsp90_61C_560C) to determine the energy and the rates of the docking and undocking transitions. Measuring at different trap distances (piezo signal) enables the exploration of the transitions at different pre-tensions. The force trace is displayed unfiltered (red) and filtered (black). (B) One second sections of the constant distance trajectories at a low (top) and a high (bottom) pre-tension. Higher forces shift the equilibrium from the docked (blue) to the undocked (red) states. (C) Force-dependent probabilities to occupy the docked and the undocked state of a single molecule (markers) are fitted to determine the equilibrium free energy between the states. The average of 34 molecules yields a free energy of $1.1\,k_BT$, hence the docked state of the unperturbed system is populated 75% of the time. Probability errors are calculated as in [139].

From the same data the transition rates between the docked and the undocked states are extracted. Figure 9.5 A shows the integrated lifetime histograms of both states at different forces for the example experiment. The lifetimes have *single exponential distributions*, indicating that docking and undocking is a two-state process. By fitting the lifetime histograms the force-dependent rates are determined, which are plotted in Figure 9.5 B. As expected the docking rates decrease with force, while the undocking rates increase.

The docked state is short-lived. Cutoffs are introduced to account for missed events, which are considered in the analysis [138].

A force-dependent rate model described by Equation 6.19 fits the measured docking and undocking rates well and determines the rates at zero force. Averaging over 34 experiments, a zero-force docking rate of $173\,s^{-1} \pm 102\,s^{-1}$ and a undocking rate of $75\,s^{-1} \pm 41\,s^{-1}$ are found. The contour length changes to the transition state are 20.8 nm from the undocked state and 8.5 nm from the docked state. The energy contributions of the dumbbell components, necessary for the fit, are precalculated from the DNA and protein parameters of the respective molecule.

As expected they add up to the complete Charged Linker length.

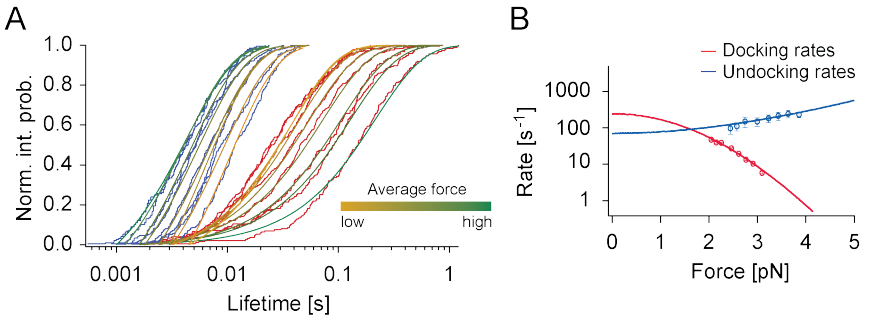

Figure 9.5: (A) Integrated lifetimes histograms for the docked (blue) and the undocked (red) states at different forces derived from a single molecule. Single exponential fits (orange to green) describe the lifetimes well, indicating that the Charged Linker dynamics are a two-state process. (B) The force-dependent docking and undocking rates derived from Fig. A. A force-dependent model yields zero force rates of around $100\,s^{-1}$ for both docking and undocking.

The thermodynamic and kinetic analysis shows that the Charged Linker mediates *very dynamic conformational changes* within the Hsp90 monomer, which are also relevant if the system is not under load. The values measured are summarized in a simple energy landscape shown in Figure C.17. Notably, the transitions are not directly influenced by the nucleotides ATP, ADP, ATPγS or AMP-PNP (not explicitly shown).

9.4 CL SUBSTITUTION MODULATES N-TERMINAL DYNAMICS

9.4.1 *Single molecule FRET experiments*

The Hsp90 dimer undergoes large conformational changes between the *N-terminally open and closed states*, which are important for the function of Hsp90 *in vivo* (see Chapter 2). Whether the intra-monomer dynamics between the docked and the undocked state influence the N-terminal inter-monomer dynamics is investigated by single molecule FRET experiments. In these experiments the distance between two dyes, termed donor and acceptor, can be determined via their FRET efficiency (for a short introduction see Section B.3.1). The FRET efficiency is very sensitive to small distance changes, therefore the open and closed states of Hsp90 can be discriminated easily. The conformational changes are evaluated in a confocal setup, where snapshots of the conformations of different single Hsp90 dimers are obtained, as well as in a surface-based TIRF assay, where trajectories of individual molecules are measured.

An established donor-acceptor pair for the Hsp90 dimer, which is described in [87], is used. The donor is attached to position 61 in the N domain of one monomer and the acceptor at position 385 in the M domain of the other monomer. To avoid dissociation of the Hsp90 dimer at the picomolar concentrations needed for single molecule experiments, a C-terminal motif which forms a stable coiled-coil with the opposing monomer is included in the constructs [87]. The Charged Linkers' influence on N-terminal dynamics is measured by comparing wild type homodimers of Hsp90 to heterodimers, where one wild type monomer is replaced by a CL substitution mutant, which cannot populate the docked state. The substitution mutants used, Sub211-263zip_61C and Sub211-272zip_61C, have the same glycine-glycine-serine repeats as those used to characterize the Charged Linker in the optical tweezers assays (see Section 9.2). For more details on the constructs and labeling see Sections A.2 and B.3.2.

Figure 9.6 A shows a schematic representation of the fluid chamber in the confocal smFRET experiments. If a labeled molecule diffuses through the confocal volume, its fluorophores are excited by pulsed lasers. The efficiency of the distance-dependent FRET process is then determined by comparing the fluorescence intensities of both dyes. For details see Section B.3.3. A large number of molecules is measured for each construct in different nucleotide conditions and collected in FRET histograms (see Figures 9.6 B–D).

Replacing one wild type monomer with one, where the Charged Linker is substituted by GGS repeats, results in a shift from the open states to the closed states as shown in Figure 9.6 B. Interestingly, the shift seems to be larger for the complete substitution mutant (Sub211-272). Preincubating the sample with 2 mM ATP further shifts

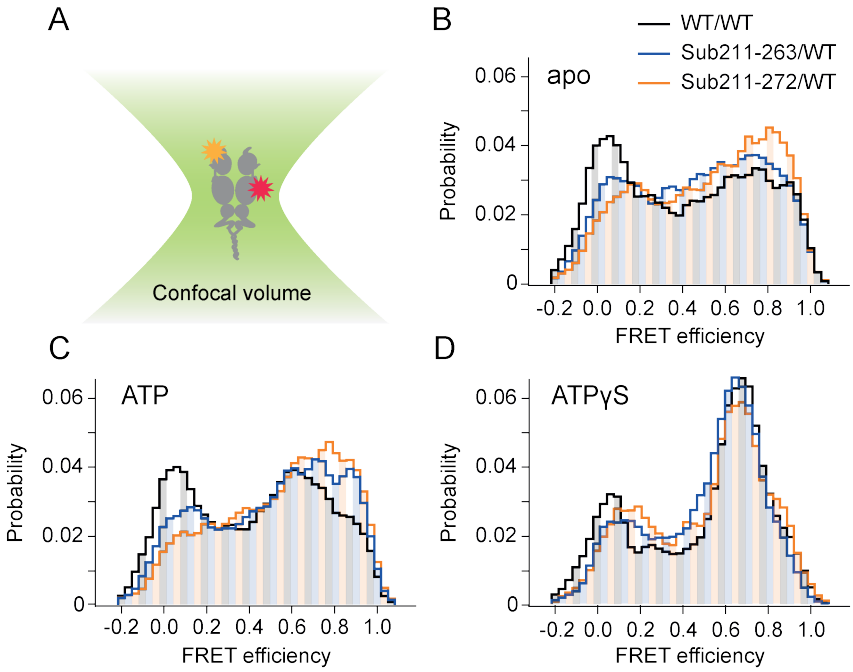

Figure 9.6: (A) Hsp90 molecules diffusing through the confocal spot result in fluorescence bursts. The ratio between donor and acceptor fluorescence determines the conformation of the molecule. (B-D) All FRET efficiency histograms of the wild type homodimer (WT/WT in black) and the GGS substitution heterodimers (Sub211-263/WT in blue and Sub211-272/WT in orange) show two populations. Low FRET values correspond to N-terminally open states, high FRET values correspond to closed states of Hsp90. Each histogram contains a minimum of 923 (up to 5,565) FRET events. Data are measured and evaluated by Björn Hellenkamp. (B) FRET histograms show that the substituion mutants are shifted towards the closed states in the absence of nucleotides (apo). (C) Addition of ATP equally shifts all dimer constructs dimers further into the closed state. (D) The ATP analog ATPγS further stabilizes a closed state. All constructs can reach this state and it seems to be identical.

all constructs similarly towards the closed states (see Figure 9.6 C). This effect has already been reported for wild type protein [87]. The unnatural ATP analog ATPγS partially traps Hsp90 in some stable closed state [48]. This is shown in Figure 9.6 D, where the FRET histograms for samples preincubated with 2 mM ATPγS, have high peaks at high FRET efficiency. In addition, the shapes of the peaks differ from those in experiments without nucleotide or ATP, suggesting a different closed state.

To confirm the findings that the undocked state shifts the equilibrium to the N-terminally closed states the same constructs as above are measured in single molecule TIR-FRET experiments in the absence of nucleotide. For this assay individual Hsp90 dimers are addi-

tionally biotinylated and immobilized on a PEG surface (see Figure 9.7 A). The donor is excited in the evanescent field and donor and acceptor fluorescence are recorded over time until the dyes bleach. For details see Section B.3.3.

Figure 9.7 E shows FRET histograms derived from FRET trajectories, which are calculated from donor and acceptor intensities (see example traces in Figures 9.7 B–D). In the FRET histograms, the heterodimers carrying the substitution mutant are shifted to the closed states, verifying the confocal experiments. In addition all constructs show the characteristic *opening and closing dynamics on the second timescale* that were observed previously for the wild type variant [87].

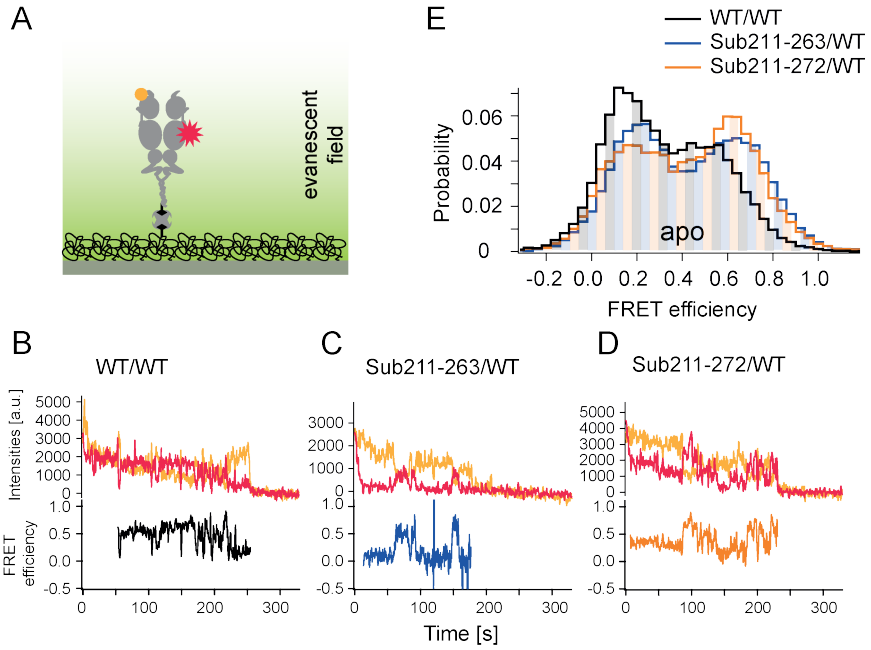

Figure 9.7: (A) Hsp90 is immobilized in the evanescent field of the excitation laser. By measuring the donor and acceptor intensities over time, the transitions between the open and the closed states are observed. (B-D) Example traces of the measured donor (orange) and acceptor (red) fluorescence intensities and the calculated FRET efficiencies (black) for the WT dimer (B), the Sub211-263/WT dimer (C) and the Sub211-272/WT dimer (D). They all show the typical N-terminal opening and closing behavior observed previously. (E) Collecting multiple FRET efficiency traces in histograms show that the substitution mutants are shifted to the closed states, confirming the previous experiments. Each histogram contains a minimum of 34 molecules.

9.4.2 *Ensemble fluorescence experiments*

The transition from the open states to the closed states of Hsp90 can be measured in ensemble FRET experiments, using a similar FRET system as in the single molecule experiments above [48]. Surprisingly, the shift caused by the natural substrate ATP is not enough to give a measurable signal in these assays. Only the addition of ATPγS or AMP-PNP, which efficiently trap Hsp90 in a closed state makes it possible to observe the closing dynamics.

To investigate the influence of the CL with this assay the Hsp90 constructs prepared for single molecule experiments are used (see Section 9.4.1). However, controls show that the fluorescence of the donor-only constructs as well as the acceptor-only constructs increases drastically upon the addition of ATPγS or AMP-PNP. This effect can only be explained, if the dyes have different quantum yields for different Hsp90 conformations. In other words, if Hsp90 isn't in the state trapped by ATPγS or AMP-PNP the protein environment quenches the dye. If Hsp90 is in the compact, trapped state, the dye has a *different protein environment* and is de-quenched. This finding can be used as a direct and easy readout for the conformation of Hsp90, which is shown in the following.

This explains the FRET signals obtained in [48], for dye positions that shouldn't give any FRET signal.

The *de-quenching experiments* are carried out with zipped homodimers of Hsp90 WT, Sub211-263 and Sub211-272. All mutants are labeled at aminoacid position 61 (see Section B.3.2). For experimental details see Section B.3.4. The fluorescence de-quenching experiments for ATPγS are shown in Figure 9.8 A. All constructs get trapped in the ATPγS closed state as observed for the heterodimers in smFRET experiments (see Figure 9.6 D). De-quenching has similar kinetics for all constructs but the substitution mutants show a greater increase in fluorescence. If the experiments are repeated with AMP-PNP (see Figure 9.8 B), WT and Sub211-263 show slower but identical behavior to ATPγS. However, the complete CL substitution mutant (Sub211-272) shows no kinetics and doesn't reach the AMP-PNP state. However, if one monomer of the complete CL substitution mutant is replaced by an unlabeled wild type monomer, the heterodimer can reach the AMP-PNP state although less efficiently (see Figure 9.8 C). These findings suggest that the C-terminal end of the Charged Linker (aminoacids 264-272) is important to reach the AMP-PNP closed state, which is different from the ATPγS closed state.

This highlights the difference of AMP-PNP and ATPγS, which are often thought to have the same effect on Hsp90.

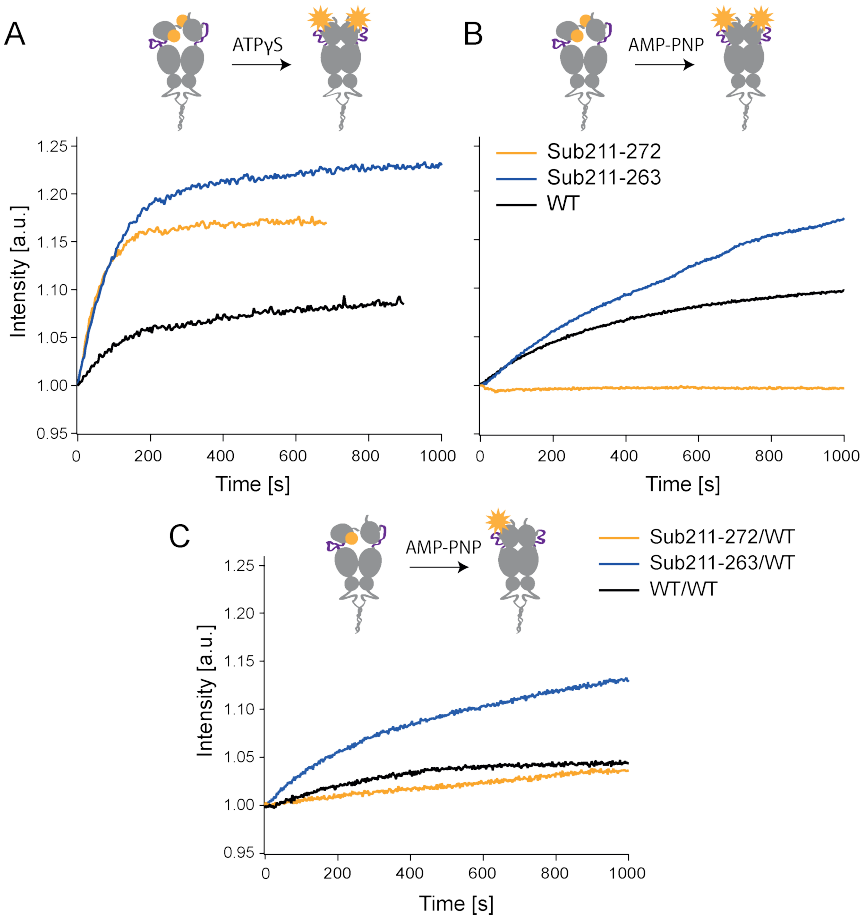

Figure 9.8: (A) Fluorescence signal after the addition of 2mM ATPγS to N-terminally labeled, zipped homodimers of WT (black), Sub211-263 (blue) and Sub211-272 (yellow). The nucleotide-dependent shift into the closed conformation increases the dyes fluorescence (de-quenching). The effect is larger for the substitution mutants indicating that the orientational freedom of the N domains leads to increased quenching.
(B) Same experiments as in Fig. A, using the nucleotide AMP-PNP. While the WT and Sub211-263 dimers show similar behavior as for ATPγS, the substitution mutant, where the CL is replaced completely (Sub211-272) doesn't show any kinetics. Hence aminoacids 264-272 are necessary to reach the closed AMP-PNP state.
(C) If one monomer of the substitution mutants Sub211-263 and Sub211-272 is replaced by WT, the Sub211-272/WT heterodimer can be partially rescued and AMP-PNP can shift it into the closed state. All experiments have been repeated multiple times.

9.4.3 *Cross-linking experiments*

Finally, the flexibility of the N domains is tested by cross-linking experiments. Cysteines in both N domains at positions 61 of wild type, Sub211-263 and Sub211-272 are allowed to form cross-links mediated by the short, cysteine-specific cross-linker $BM(PEG)_3$. For experimental details see Section B.4.

The degree of cross-linking in the presence of different nucleotides is analyzed by SDS-PAGE comparing the fraction of dimers and monomers, as shown in Figure 9.9. Although the cysteines in the N domains point away from each other in the closed crystal structure cross-linking is still possible for the wild type protein, where the Charged Linkers are mainly docked. For both substitution mutants, which just populate the undocked state, the cross-linking propensity is increased significantly. Since cross-linking should be impossible in the open states, because the N domains are well separated, the data suggest the existence of closed states, where the aminoacids at position 61 point at least transiently towards each other.

On a closer inspection the effects of the added nucleotides are also seen. Preincubation with ATPγS decreases cross-linking for all proteins, presumably by trapping them in the closed, compact state, where cross-linking is impossible. The same is observed for AMP-PNP for Sub211-263, but not for Sub211-272. The cross-linking results are in excellent agreement with the optical tweezers experiments, which suggest rotational freedom of the N domains, especially if the CL is permanently undocked, and with the fluorescence dequenching experiments, which suggest no influence of AMP-PNP on the complete CL substitution variant.

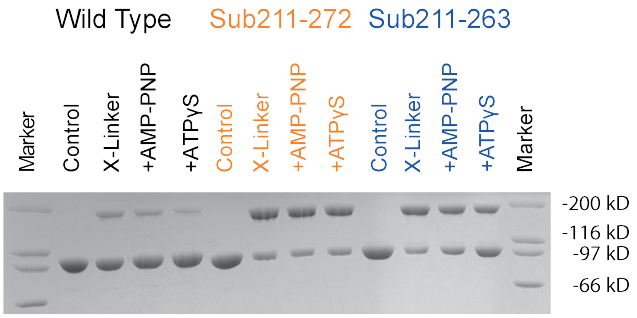

Figure 9.9: N-terminal cross-linking of Hsp90zip_61C, Sub211-272zip_61C and Sub211-263zip_61C without nucleotide (X-linker), ATPγS or AMP-PNP. In the control lanes no cross-linker is added. If the N domains cross-link, dimers (upper band) are formed, if not, monomers are observed (lower band). Cross-linking is possible for wild type, but significantly increased for both substitution mutants. Preincubating the sample with ATP analogs traps the proteins in closed conformations, avoiding some cross-links.

9.5 THE CHARGED LINKER AFFECTS HSP90'S ATPASE CYCLE

How *Hsp90's ATPase activity* couples to Hsp90's conformations has been studied intensively in bulk assays [120, 119, 48, 71]. There the hydrolysis rates of different Hsp90 mutations and the effect of co-chaperones are evaluated. One proposed ATPase cycle is presented in the introduction (see Figure 2.2). To characterize how the Charged Linker kinetics and the impossibility of reaching the AMP-PNP induced, closed state affects the ATPase cycle, WT, Sub211-263 and Sub211-272 (see Section A.1) are compared in ATPase and analytical ultracentrifugation assays.

Hsp90_61C_560C shows identical experimental results as WT, therefore it is called WT in the text.

The ATPase activity measured with a coupled enzymatic assay (see Section B.5) shows a mild increase for Sub211-263 and a mild decrease for Sub211-272 compared to the wild type (see Figure 9.10 A).

The co-chaperone Aha1 can stimulate the ATPase activity of Hsp90 significantly, presumably by facilitating the closing of Hsp90 [101, 119]. This effect is significantly reduced for the substitution mutants, by a factor of three for the partial CL replacement (Sub211-263) and by a factor of ten for the complete CL replacement (Sub211-272), as shown in Figure 9.10 B. To exclude the possibility that the reduced activation is due to impaired binding of Aha1, analytical ultracentrifugation experiments (for details see Section B.6), where labeled Aha1 can form complexes with the Hsp90 variants, are performed. Figure 9.10 C shows that all mutants bind Aha1 at the concentrations used. The reason why Aha1 can't deliver its full potential may be due to important residues of the CL necessary for the correct interaction of Hsp90 and Aha1.

Since all Aha1 proteins are bound, it can't be fully excluded that they bind with different affinities.

The co-chaperone Sba1/p23 can inhibit ATPase activity by binding to and stabilizing the AMP-PNP, closed state of Hsp90, observed in the crystal structure [135, 123, 2]. ATPase inhibition is observed for wild type and Sub211-263, but not for Sub211-272 (data not shown). The reason therefore is that Sba1/p23 can't bind to Sub211-272 as shown in ultracentrifugation experiments in Figure 9.10 D. This suggests that the complete CL substitution variant cannot reach the state necessary for Sba1/p23 binding upon addition of AMP-PNP, which is also the case in bulk fluorescence experiments (see Section 9.4.2).

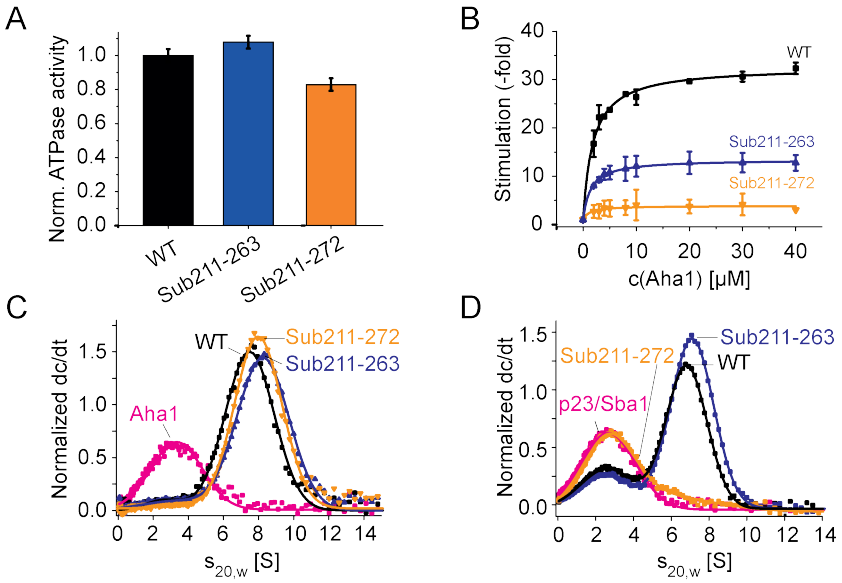

Figure 9.10: (A) ATPase activity of Hsp90 wild type, Sub211-263 and Sub211-272. The values are normalized to wild type, whose hydrolysis rate is $0.36(\pm 0.01)\,\mathrm{min}^{-1}$.
(B) The concentration-dependent activation of the ATPase by Aha1 is reduced for both GGS substitution mutants. The stimulation is normalized to the ATPase activities shown in Fig. A and the fits follow Michaelis-Menten equation.
(C) Ultracentrifugation experiments show that all Hsp90 constructs can bind labeled Aha1, which is rationalized by the shift between Aha1-only and the Aha1-Hsp90 complexes in dc/dt plots.
(D) dc/dt plots show that labeled Sba1/p23 can form complexes with wild type and Sub211-263, but not with Sub211-272. ATPase assays were performed together with Alexandra Rehn and ultracentrifugation runs were done by Klaus Richter.

9.6 DISCUSSION

The integrated approach, combining single molecule force and fluorescence experiments with biochemical assays, gives great insight into the previously unresolved structure, dynamics and function of the Charged Linker.

9.6.1 *The Charged Linker regulates the flexibility of the N domain*

Optical tweezers experiments revealed that the *Charged Linker is not as unstructured* as previously assumed. The Charged Linker can dock to the N domain forming a stable state, where the N and M domains are in close proximity. However, this conformation is stabilized by a free energy of only 1.1 k_BT, meaning that *thermal fluctuations* can depopulate this state easily. In the dissociated (undocked) state, which is populated around 25% of the time, the Charged Linker acts like a random coil, giving additional *translational and rotational freedom* to the N domain. This is also confirmed by cross-linking experiments. The intradomain transitions between the docked and the undocked states are on the *millisecond timescale*, much faster than all previously observed kinetics of Hsp90. Such unstable and rapid conformational changes of native structures usually remain undetected in bulk studies, although they are very important for the function of proteins [164].

Aminoacids within residue region 211-263 of the Charged Linker are essential for this interaction.

9.6.2 *The Charged Linker modulates N-terminal dynamics*

The N-terminal conformational changes between the open and closed states are influenced by the Charged Linker. In single molecule FRET experiments, mutants that only populate the undocked CL state show a higher probability to be in the closed states. This suggests that the increased flexibility of the N domains facilitates reorientations necessary to reach the closed states. That rotation of the N domains is necessary for closure has already been suggested for *E. Coli* Hsp90 [141]. On the contrary, the docked state seems to prevent formation of contacts between the N domains, suggesting that the Charged Linker acts as some kind of *return spring*. This is consistent with hydrogen exchange experiments, where similar substitution mutants show decreased protection of the N-domain and increased protection of the M-domain compared to WT [150].

Cross-linking between the N-terminal residues, which are pointing away from each other in the closed crystal structure, is possible for the wild type Hsp90 and even more likely for the mutants that can only populate the undocked state. This suggests that the undocking and docking transitions are also possible in the N-terminally closed states.

The shift to the closed states is also seen in the sedimentation profiles of the ultracentrifugation experiments (see Figure 9.10).

9.6.3 *The C-terminal part of the Charged Linker is essential for reaching the AMP-PNP closed state*

During Hsp90's ATPase cycle Hsp90 seems to populate *multiple closed states* (see Figure 2.2). The so-called Closed 1 state [72], is a semi-closed state, which is stabilized by the co-chaperone Aha1 [119] and is maybe the same that can be reached without nucleotides in single molecule FRET measurements [87]. The twisted, nucleotide-dependent closed state, that has been crystallized in the presence of AMP-PNP [2], is the binding-competent conformation of the co-chaperone Sba1/p23 and called Closed 2. Since AMP-PNP cannot be hydrolyzed by Hsp90 it is believed to be the *hydrolysis competent state*. ATPγS, which is hydrolyzed very slowly, is believed to stabilize the same state [48].

The experiments presented here show significant differences between the partially (Sub211-263) and complete substitution (Sub211-272) mutants that only differ in the last 9 aminoacids of the Charged Linker. If these aminoacids are also replaced by GGS repeats, ensemble fluorescence experiments show that AMP-PNP can't promote N-terminal closure, cross-linking propensity is not reduced by AMP-PNP and ultracentrifugation experiments show the impossibility of Sba1/p23 binding. Therefore this region is important for reaching the AMP-PNP Closed 2 state. It might be that the interaction between this region and the N domain, which is suggested by the crystal structure (see Figure 9.1 D), is only important in the Closed 2 state, which would mean that the N-M domain interface may be completely different in the absence of nucleotide, as it is the case for *E. Coli* Hsp90 in the presence of ADP [133].

Some groups denote this region (aminoacids 264-272) already to the M domain.

Surprisingly, the complete CL substitution mutant (Sub211-272), which cannot reach the Closed 2 state shows only a mild decrease in ATPase activity. On the other hand the ATP analogue, ATPγS, readily shifts all Hsp90 mutants into a closed state, as observed in smFRET experiments as well as in de-quenching fluorescence experiments. This implies that that ATPγS and AMP-PNP induce different Hsp90 states and that hydrolysis is not limited to the AMP-PNP state.

Although the Closed 2 state should represent the hydrolysis competent state of Hsp90.

9.6.4 *Hsp90 states mediated by the Charged Linker*

Figure 9.11 summarizes the findings of the different experiments. Optical tweezers experiments show that the Charged linker facilitates a docked and an undocked state. While Hsp90 is in its open conformation, 56% of the time both Charged Linkers will be docked (Open 1a), but almost equally likely, 38% of the time, one of both CLs is in the undocked state (Open 1b). Cross-linking experiments suggest that docking and undocking is also possible in the Closed 1 state, therefore it is renamed to Closed 1a and Closed 1b. To reach the AMP-

The 6% probability that both Charged Linkers are undocked is omitted.

PNP-induced Closed 2 state aminoacids within positions 264-272 are essential.

The new states facilitated by the Charged Linker are rapidly exchanging and could be easily stabilized by co-chaperones or substrates. This would enable *new stable global conformations of Hsp90*, which may be important in the chaperone cycle. Such small modulations may also hold the key to chaperoning different sets of substrates as is the case for Hsp90.

Optical tweezers experiments studying the N-terminal opening and closing of Hsp90 will be shown in Chapter 10.

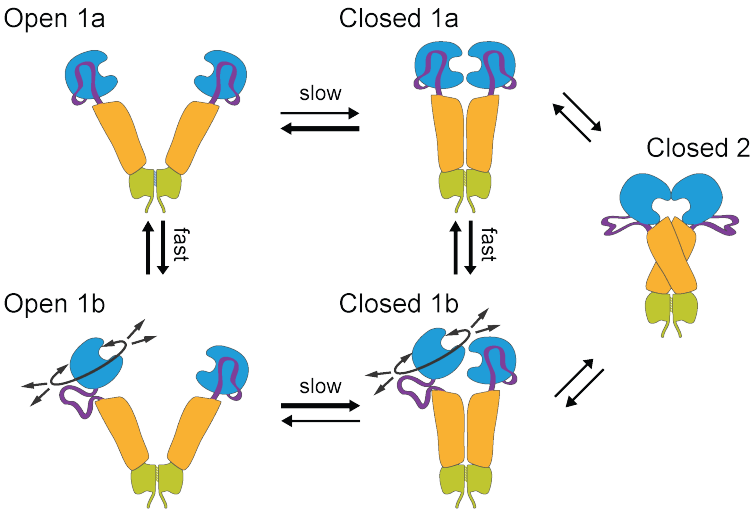

Figure 9.11: The Charged Linkers facilitate docked (Open 1a and Closed 1a) and undocked (Open 1b and Closed 1b) states in the Hsp90 dimer, which exchange rapidly. Because undocked states favor the N-terminal closed states, the CL modulates the much slower N-terminal dynamics of Hsp90 (Open and Closed). The C-terminal region of the Charged Linker is essential for reaching the nucleotide-dependent Closed 2 state.

10 DIMERIZATION AND N-TERMINAL DYNAMICS

The biologically active form of Hsp90 is the Hsp90 dimer, whose formation is mediated by the C domains. After the dimer is assembled, Hsp90 can transition between its N-terminal open and closed states, which are essential for client chaperoning (see Chapter 2). To investigate these important protein-protein interactions with optical tweezers, *dimeric Hsp90 constructs* are designed. They permit the investigation of the C-terminal dimerization kinetics in great detail and are promising for the study of the N-terminal dynamics.

10.1 DESIGN OF HSP90 DIMERS FOR OPTICAL TWEEZERS

To design dimers for optical tweezers experiments, the constructs measured in single molecule fluorescence experiments (see Sections A.2 and 9.4.1) are used as a starting point. They contain alpha helices at the C-termini, which form coiled-coils between the two opposing monomers. However, these coiled-coils are not stable enough to keep the monomers together during optical tweezers experiments, since they readily dissociate at forces of around 10 pN [14]. To increase their stability, cysteines that can covalently cross-link the coiled-coils are introduced at the N-termini of the alpha helices. The positions of the cross-linking cysteines are critical. First attempts to insert them between Hsp90 and the coiled-coil didn't result in disulphide bridges. However, mutation of the first a-position in the heptad repeats of the coiled-coil yields rapidly-formed and stable disulphide bonds.

To avoid interaction with streptavidin beads, also the Strep-tag is removed.

The second challenge is finding optimal cysteine positions for DNA attachment within the proteins. The dimers contain four cysteine residues, two for cross-linking, which have to react, before the other two, which are used for oligonucleotide attachment (see also Section 5.2). Therefore only cysteine positions can be chosen that are well-separated in the dimer and form disulphide bonds slowly. This explains the impossibility to construct dimers, where force is exerted at the N domains via cysteines at residue 9. However, two other amino-acid positions, 61 in the N domain and 452 in the M domain, give excellent results. These two constructs, which permit force exertion on both N domains or both M domains, while the monomers are linked covalently at their C termini are discussed in the following.

These constructs are sketched in Figures 10.1 and 10.2.

10.2 N-TERMINALLY PULLED DIMERS

Figure 10.1 shows unfolding and refolding force-extension traces of the yeast dimer construct slowly pulled at both N domains at amino-acid 61, which is described in the previous Section 10.1. As expected Hsp90 domains are observed twice, proving successful dimer design. Due to the pulling geometry the N domains have shorter contour length gains, identical to the monomer construct Hsp90_61C_560C discussed in Section 7.2. After both Charged Linkers are forced into the undocked states, an additional event, which unfolds at forces around 10 pN is observed. *Dissociation of the dimer* is followed by rapid unfolding of the C domains, which are not stabilized any more and are only visible as short events. If the chain is relaxed slowly and all domains have refolded, dimerization is observed, even against forces of around 3 pN. Additional cycles are shown in Figure C.18.

Notably, N domains pulled at aminoacids 61 show worse refolding properties than those pulled at the N-terminus. This suggests that the N-terminal force-free part hinders refolding, in agreement with the observations made in Chapter 8.

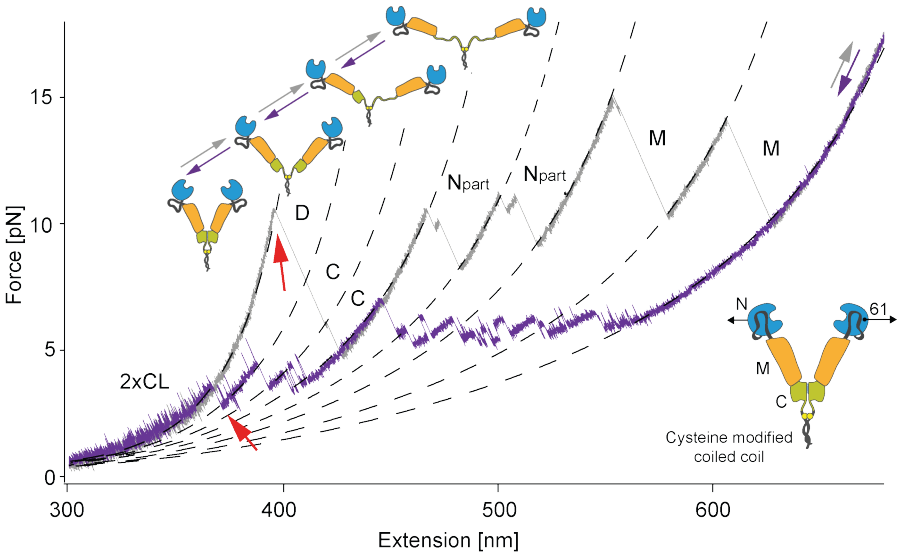

Figure 10.1: Unfolding (gray) and refolding (purple) force-extension traces of the Hsp90 dimer (Hsp90zip_61C_715C), pulled at amino-acids 61 in the N domains (see lower right-hand inset). The pulling speed is 10 nm/s. The unfolding events associated with the Hsp90 domains (Npart, CL, M and C) appear twice and are fitted with WLCs assuming some unstructured regions (see Section 6.1.1). In addition a high peak (red arrow on gray trace; D) is observed. There the dimer dissociates and the unstructured regions of both C domains are stretched, followed by rapid unfolding of the C domains. After all domains have refolded, dimerization is observed (red arrow on purple trace). The sequence of dimerization and folding of the C domains is sketched above the trace.

The contour length gain associated with the dimerization, evaluated by fitting WLCs to slowly-pulled traces of seven molecules, is 40.6 ± 1.1 nm. The unstructured regions of the C domains, which are

not under force in the dimerized state, are not sufficient to explain this contour length gain. As already speculated in Section 7.4, the last alpha helix of the folded C domain may be unstable if it doesn't form the four helix bundle involved in dimerization. Both scenarios are compared in Table 10.1. The model assuming the unstable helix (aminoacids 656-674) shows perfect agreement with the measured data. This proves that the monomers form dimers through the helix pairs in the C domains (residues 640-672) as already suggested by the crystal structure and that the last alpha helix is unstable in the monomeric state.

Moreover, contributions of the N domains to dimerization can be clearly ruled out since both Charged Linkers are already undocked.

HELIX (UNSTRUCT.)	DIST. CS	EXPECTED	MEASURED
Stable (675-715)	0.8 nm	29.1 nm	40.56 nm ± 1.06 nm
Unstable (656-715)	1.7 nm	41.4 nm	40.56 nm ± 1.06 nm

Table 10.1: To explain the contour length gain observed after the dissociation event two possibilities are considered. One where the C-terminal alpha helix in the C domain is stable and one where it is unstable if undimerized. The monomers are cross-linked 6 aminoacids after the C terminus of the wild type protein (corresponds to aminoacid 715). Calculating the expected contour length gains from the assumed unstructured regions (see Section 6.1.2), shows that the model assuming the last C-terminal helix as unstable is in perfect agreement with the data. Errors are SD.

10.3 M-TERMINALLY PULLED DIMERS

To investigate the C-terminal dimerization in more detail the other dimeric Hsp90 mutant is used, where force is applied to the M domains via aminoacid residues 452 (see also Section A.2). This construct has considerable advantages. First, under force the N and M domains are not or are only partially unfolded, therefore constant velocity experiments can be performed without getting trapped in misfolded states (see Chapter 8). Second, the Charged Linkers are not under force and the absence of these transitions simplifies the analysis of the traces. Finally, the N-terminal opening and closing of the Hsp90 dimer should result in a measurable contour length change, which will be important in Section 10.5.

Figure 10.2 shows two examples of force-extension cycles of this construct at different pulling speeds. Similarly to the N-terminally pulled constructs (see Figure 10.1), dissociation followed by unfolding of the C domains and association of the Hsp90 monomers are observed. In Figure 10.2 B also both M domains are partially unfolded. These parts are the same that give rise to the stable M domain unfolding and refolding intermediate, as described in Sections 7.3 and 8.6.2. The high unfolding forces and the refolding properties are in good

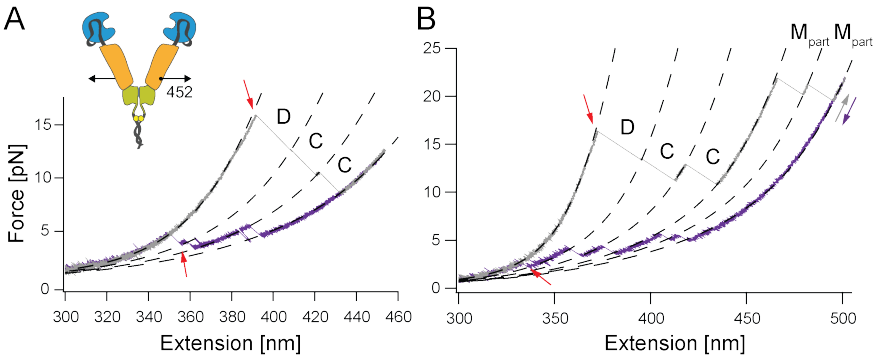

Figure 10.2: Example force-extension traces of the dimer construct pulled at the M domains (Hsp90zip_452C_715C; see inset) stretched (gray) and relaxed (purple) with a velocity of 10 nm/s (A) and 500 nm/s (B). Dissociation and association of the C-terminal dimer interface are observed (D; red arrows). After dissociation the C domains readily unfold. All states are marked by WLCs shown as dashed lines. In Fig. B also partial unfolding of both M domains is observed (M_{part}), which is very stable and refolds rapidly.

agreement with the observations in Section C.3.5, that suggests two-state folding of this intermediate and stabilization by the remaining, still-folded subdomain.

10.4 KINETICS OF THE C-TERMINAL DIMERIZATION

The most accurate way to evaluate the kinetics of the C-terminal dimerization is using constant distance experiments (see Section 5.3). Figure 10.3 shows an example trajectory of such an experiment, where the dimerization kinetics are observed at equilibrium. However at pre-tensions that allow the observation of these transitions, rates are slow and the force range is restricted due to the unfolding of the C domains. Hence gathering statistics of the force-dependent lifetimes necessary for rate calculation is possible but tedious. To circumvent this problem, a different approach is applied.

Dimerization energy determined by constant distance experiments is roughly $10\,k_BT$ *(data not shown).*

Since the dissociation forces are related to the dissociation rate and the association forces to the dimerization rate, force-dependent rates can be determined from constant velocity experiments. Therefore a method described by Oberbarnscheidt *et al.* (see [95] and Section 6.7.2) is applied to an example experiment comprising 94 force-extension cycles at a speed of 200 nm/s. Figure 10.4 shows the number of dissociation and association events and their respective lifetimes pooled in force-bin histograms as well as the resulting force-dependent rates.

To calculate the zero-force rates the force-dependent rates are fitted with Equation 6.19. A *low dissociation rate* of $0.038(\pm 0.029)\,s^{-1}$, with a low force dependence of $2.0(\pm 0.3)$ nm (contour length difference

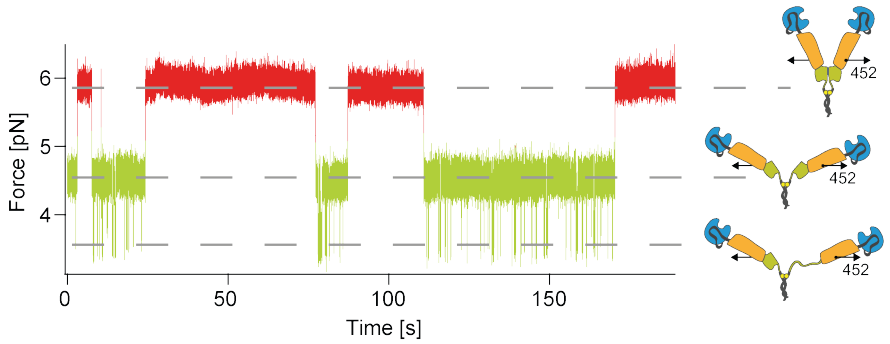

Figure 10.3: The C-terminally dimerized state (red) and the dissociated state (green) can be observed at equilibrium in constant distance experiments while pulling at both M domains (Hsp90zip_452C_715C). At the pre-tension applied in this example, one of the C domains can unfold and refold if the dimer is in its dissociated state, leading to short excursions to the lowest force level. The three states observed are sketched next to the trace. In general, the dimerization kinetics measured at equilibrium are relatively slow.

to the transition state) is found. The association rate is $137(\pm 52)\,s^{-1}$ with a distance of $27(\pm 4)$ nm to the transition state.

Seven other molecules, evaluated the same way, show similar rates, but the kinetics are more complex than presumed. A significant number of dissociation events occur at low forces and they seem to cluster in successive force-extension cycles. This suggests that a slow isomerization process within the polypeptide chain leads to two dimer states with a 'weak' mode and a 'strong' mode. Although the 'strong' mode dominates the weak mode, it is also captured in a slowly-pulled force-extension trace (see Figure C.19). To what extent this behavior is introduced by the repeatedly stretching of the chain in optical tweezers is yet unclear. A candidate residue which could explain this behavior is the proline at position 661 in the turn of the alpha helical pair that forms the dimer interface [124].

More low force events than expected from the determined rate and transition state position.

Additionally, the finite bin size and slight deviations from the WLC fits, which are used to determine the lifetimes of the states, may increase the error. Therefore the obtained rates should be seen as estimates.

These errors are especially a problem for the association rates in the lower force range.

The equilibrium dissociation constant K_D measured in bulk biochemical experiments is 60 nM [120]. In the optical tweezers experiment the association rate depends on the tethering of the monomers, while in the biochemical experiments it depends on the Hsp90 concentration. From the measured rates as well as the K_D it is possible to calculate the effective concentration of the Hsp90 monomers in the optical tweezers assay. This concentration is 220 µM, which corresponds to 1 particle in a sphere with a radius of 12 nm. This value meets the expectations, since the monomers are separated by the unstructured

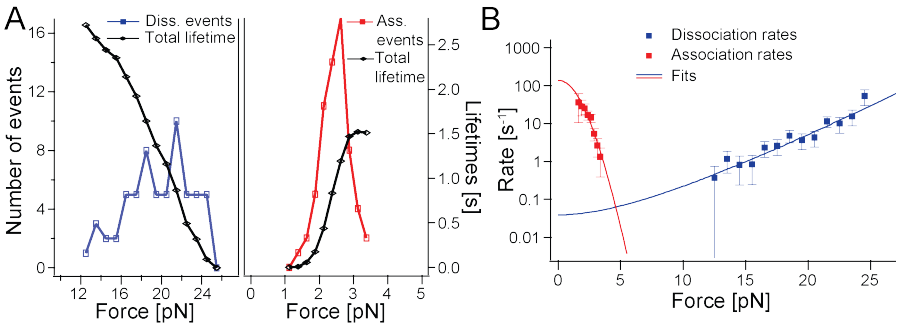

Figure 10.4: (A) Number of dissociation (blue) and association events (red) for different force bins as well as the total time spent in the respective force bin without having transitioned yet. Data are from 94 force-extension cycles of one example molecule (Hsp90zip_452C_715C) and only dissociation events higher than 12 pN are evaluated. (B) Force-dependent dissociation (blue markers) and association (red markers) rates are calculated from the data in Fig. A and extrapolated to zero-force yielding a dissociation rate of $0.038\,s^{-1}$ and an association rate of $137\,s^{-1}$.

regions of both C domains, indicating consistency between the bulk and the single molecule measurements.

10.5 EVIDENCE FOR N-TERMINAL DYNAMICS

So far only the interaction between the C domains of both monomers is observed in optical tweezers experiments. The literature suggests also interactions between both N domains, which give rise to the opening and closing dynamics of Hsp90, which are important for ATP hydrolysis and client chaperoning, as introduced in Chapter 2. Indeed an additional state can be identified in two independent optical tweezers experiments, if the nucleotide ATP is added.

The N-terminal dynamics are already measured using fluorescence experiments in Sections 9.4.1 and 9.4.2.

10.5.1 *N-terminal closing of dimer constructs*

Figure 10.5 A shows two example trajectories of constant distance experiments with the M-terminally-pulled dimer construct in the presence of ATP, similar to the one without nucleotide shown in Figure 10.3. At the trap distance measured, the dimer mainly stays in its C-terminally dimerized state, although sometimes it transitions into its undimerized state at lower forces in agreement with the kinetics described in the previous section. In addition a rare but reproducible state at higher forces is observed. This state is exclusive or at least preferentially populated if ATP is in solution, since it is not observed in nucleotide-free conditions. The contour length is around 14 nm smaller than the C-terminally dimerized state, indicating a *very compact state*. The ATP dependence and the contour length suggest that

It is impossible to tell from the experiment, if one or both Hsp90 monomers have to bind ATP to reach this state.

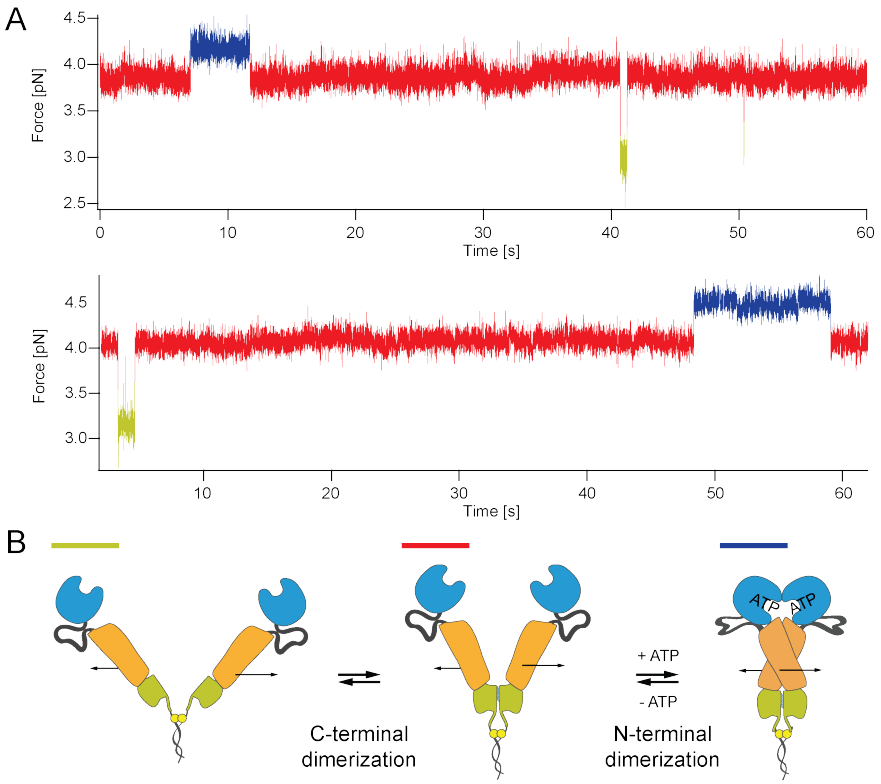

Figure 10.5: (A) Example constant distance traces of two different Hsp90zip_452C_715C molecules. The dimers transition between the dimerized (red) and the undimerized (green) conformation. In the presence of ATP a new state with shorter contour length is observed (blue).
(B) The observed states and their connectivity are described schematically. The new, ATP-dependent state is due to N-terminal closing into a compact state.

this state is an N-terminally closed state, likely very similar to the compact, entwined conformation of the crystal structure [2].

Although the N-terminal closing is measured at low forces around 4 pN, the *closing rates are very slow*, on the order of a few hundred seconds. If Hsp90 reaches the closed state, it is *stable for a few seconds*. Owing to the slow rates, the full force dependence of the opening and closing dynamics hasn't been investigated yet.

10.5.2 *N-terminal dynamics measured with the Hsp90 monomer*

A very similar state, like the N-terminally closed one presented in the previous section, is observed with an Hsp90 monomer construct. The state appears in constant distance experiments, like those described in Section 9.3 to characterize the Charged Linker, if wild type Hsp90 and ATP, both at saturating conditions, are added in solution. Example traces in Figure 10.6 show not only the well-studied docking and undocking transitions of the Charged Linker, but also a long-lived state with similar contour length as the docked state.

At the forces investigated, the lifetime of this state is a few seconds and therefore much longer than the lifetimes of the docked Charged Linker conformation. The rate into this state is very low, similar to the one observed for the dimeric construct. The related kinetics and the fact that the state is only populated if Hsp90 monomers and ATP are in solution leads to the conclusion that this state is indeed the same N-terminally closed, compact state observed for the dimer in Figure 10.5.

Some of the experiments are performed in the presence of Aha1, which seems to increase the closing rate.

As expected for the pulling geometry, dimerization and N-terminal closing don't lead to a significant change in contour length (see also Figure 10.6 B) and the docked state and the N-terminally closed state can only be separated by their different kinetics. Whether the N-terminally closed state is populated from the docked state or the undocked state, cannot be answered at this point.

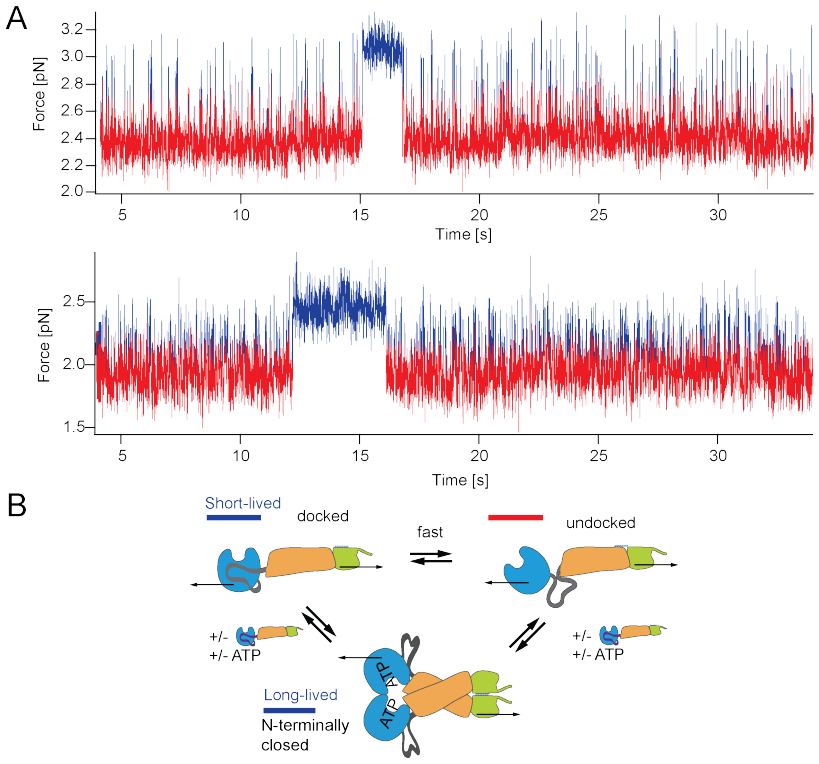

Figure 10.6: (A) Example constant distance traces of two different experiments with the Hsp_61C_560C monomer show the rapid CL docking (blue, short-lived) and undocking (red) transitions. If wild type Hsp90 and ATP are present in solution, long-lived states (blue, long-lived) with similar contour length as the docked state are observed. They can be explained by a mechanically stable closed state, which is dependent on monomer and ATP binding.
(B) Schematics of the states populated in the constant distance traces. The long-lived state depends on ATP and Hsp90 monomers in solution, but it is not possible to tell from the experiment, at what point dimerization occurs and how many ATPs are bound.

10.6 DISCUSSION

10.6.1 *Assembly of Hsp90*

At physiological conditions Hsp90 exists as a dimer [120], which is the *biologically active form* required for the function *in vivo* [156]. Therefore after Hsp90 monomers have successfully folded, they have to find each other and associate. This process as well as the dissociation process can be directly monitored and manipulated in optical tweezers experiments.

There is also some speculation about tetramers, because HtpG has been crystallized in this form [133].

Dissociation of the dimer in unfolding force-extension traces happens at high forces and is followed by the rapid unfolding of the C domains, suggesting that the C domains are the main interaction site and that they are stabilized by dimerization. This is in agreement with bulk denaturation experiments, where increasing protein concentrations show increasing stabilities [156]. The interaction between the C domains is mediated by a *four helix bundle*, one helix pair per monomer, as resolved by the crystal structures of yeast and *E. coli* Hsp90s [46, 2, 133]. The optical tweezers experiments additionally revealed that the last C-terminal helix is not folded if the monomers are not in the dimerized state, which goes well with the increased hydrogen incorporation observed in exchange experiments [41].

The C-terminal kinetics of Hsp90 are of great interest, because some studies suggest that they are important for the Hsp90 chaperone cycle [117]. First results yield a slow dissociation rate of $0.038\,s^{-1}$ and a dimerization rate of $137\,s^{-1}$, which is dependent on the effective concentration defined by the linkage of the monomers. They are in agreement with the K_D of 60 nM determined in bulk experiments [120]. On the contrary, smFRET experiments of Hsp90 encapsulated in vesicles [117], show dimer lifetimes of only 3 seconds, which are incompatible with the dominant dimerization peaks observed and the rates determined using optical tweezers.

K_D compared to yeast Hsp90 [46]; see also Figure D.3.

The results suggest that due to the slow dissociation rate and the high cellular Hsp90 concentration, the Hsp90 dimer will only *transiently populate its monomeric state*. Therefore it remains possible, but unlikely, that the monomeric form is an important feature in the chaperone cycle. This is supported by the fact that no co-chaperones or nucleotides significantly modulate Hsp90's dimerization properties. However, the importance of the bivalent stability of the dimer has to be addressed in the future.

C-terminally fused Hsp90 constructs also support yeast viability [112].

10.6.2 *N-terminal conformational changes*

The N-terminal dimerization of Hsp90 and the related ATP binding and hydrolysis are both essential for Hsp90's function [97, 100, 109]. Although extensively studied, the basis of this interaction [111, 109,

If the proline is indeed responsible for this behavior, it should be mentioned that the co-chaperones Cpr6 and Cpr7 have PPIase activity [83].

121, 2], its coupling to nucleotides [120, 87, 115], the number of closed states involved [120, 48, 87] and why closed states are necessary to chaperone clients [84, 140, 61] remain poorly understood. Studies are further complicated by the fact that the conformational cycle is different for different organisms [136, 41, 118]. The only commonality is that the non-hydrolysable nucleotide AMP-PNP traps Hsp90s in a closed compact conformation, likely the twisted one that has been crystallized [2]. While ATP induces only small population shifts into the closed states in yeast and human Hsp90 [48, 87, 41], the effect is much larger in *E. coli* Hsp90 [42, 41, 118].

In optical tweezers experiments a state is identified, which has *all characteristics of a closed state*. It is stable in two pulling geometries, along the monomer as well as across the M domains. Additionally, it shows a very short contour length suggesting a compact conformation. It is *strongly coupled to nucleotides*, since the state is only observed in the presence of ATP. Although this conformation is investigated at low forces, the on-rate is very slow and the closed state is populated rarely, which is in agreement with the small ATP-dependent shifts mentioned above. One can hypothesize that this state is similar to the closed compact state of the closed crystal structure [2] and that at zero-force the closing rate will be on the order of the ATPase rate, indicating that the rate limiting step for hydrolysis is closing into this state, as already suggested in [120, 42, 112].

There is also no evidence that other closed states exist.

Nucleotide binding studies argue against ATP association or ADP dissociation being the rate limiting step [85, 115, 110].

Single molecule FRET experiments, as shown in Section 9.4.1 and [87] suggest that, even in the absence of nucleotides, multiple open and multiple closed conformations of Hsp90 are populated equally with transition rates of around $0.25\,s^{-1}$ [87]. The fast rates and the observation of the closed states without nucleotide are incompatible with the optical tweezers measurements. Since both experiments are carried out under very similar conditions, it is likely that the *different methods themselves lead to different results*. Single molecule FRET experiments use hydrophobic dyes as reporters while optical tweezers experiments use hydrophilic DNA handles to tether the protein. Attaching different dyes to different Hsp90 residues can greatly shift (up to 40%) the populations of Hsp90 (personal communication by Philipp Wortmann), therefore it is likely that the increased attraction or the different environments of the dyes lead to the observation of additional kinetics and additional closed states. On the other hand, the DNA handles may lead to a more soluble Hsp90, hampering N-terminal dimerization.

Quantitative optical tweezers jump-experiments, performed by Katarzyna Tych, probing the conformation of Hsp90 at zero-force suggest only a fraction of around 5 % in the compact closed state.

However, the results from optical tweezers experiments are in good agreement with other experiments [120, 41]. Having established the Hsp90 dimer system for optical tweezers allows detailed investigation of the influence of co-chaperones and substrates on the closed states of Hsp90 and will shed light on the differences between optical tweezers and FRET experiments.

11 OUTLOOK

In this thesis, the *Hsp90 system is established for optical tweezers experiments*. Different properties have been characterized, opening various possible future directions. However, the most interesting and pressing questions are how the Hsp90 alters the structure of its clients and for what reason ATP is hydrolyzed. To get a real understanding of this, some features of the Hsp90 machinery have to be clarified first. Exploiting the strength of single molecule optical tweezers experiments should give valuable insights.

First, the N-terminal kinetics, which are essential for chaperoning and modulated by nucleotides have to be investigated in more detail. Therefore it is important to reconcile optical tweezers and single molecule fluorescence measurements (see Section 10.6.2) to have independent methods to investigate the system.

Several experiments suggest the existence of *multiple closed states*, but their conformations and the molecular interactions that stabilize the different closed states are unclear. The only hint is given by the crystal structure with the ATP analog AMP-PNP, which shows a very compact, entwined Hsp90 dimer. Two models try to explain the stabilization of the N-terminally closed states. The first suggests domain swapping of the N-terminal beta strands between both Hsp90 monomers. The second nucleotide-induced remodeling of the N domain and exposure of hydrophobic residues. Optical tweezers experiments are well suited to distinguishing between different closed states, which should make it possible to determine the number of distinct states and to reveal their molecular interactions. Studying different Hsp90 mutants and measurements in the presence of nucleotides and/or co-chaperones should shed light on the N-terminal kinetics and the regulation by co-solutes.

Whether the stability of this beta strand is influenced by nucleotides could be addressed by investigating the preflipping behavior of the N domain (see Figure 7.4.)

A promising and simpler candidate to study the conformations and the influence of nucleotides on Hsp90s chaperone cycle is the *E. coli* ortholog, which isn't regulated by co-chaperones. Biochemical experiments indicate that already the natural substrate ATP induces a significant population shift to the closed states. First experiments with a dimeric *E. coli* Hsp90 look promising, as shown in Figure D.3.

Possible candidates are the Δ8 Hsp90 mutation and the co-chaperone Aha1.

Besides the ambiguity of Hsp90's conformations, bulk assays have limited success in *revealing Hsp90's chaperoning mechanism in vitro* (see Chapter 2). Whether these difficulties arise from not reaching the Hsp90 conformation necessary for client binding or the clients cannot be prepared in their chaperone-dependent form is unclear. Optical tweezers provide two ways to tackle this problem. Either Hsp90 is

held by optical tweezers and the influence of the clients on the Hsp90 chaperone cycle is monitored or clients are clamped and the effect of Hsp90 in solution on the substrate is directly measured. The latter approach provides the possibility of actively modifying the client by force, perhaps inducing Hsp90-binding-competent states.

One main problem for the second approach is finding clients suitable for optical tweezers experiments. First attempts establishing the pseudo substrate of *E.Coli* Hsp90, Δ131Δ, in optical tweezers failed (see Section A.4). However, initial experiments suggest that *E.Coli* Hsp90 in solution can interact with unfolded or partially folded states of itself (see Figure D.4). Other promising clients are the glucorticoid receptor, v-src kinase and tau (see Chapter 2).

Also calmodulin, which has been reported to interact with Hsp90, doesn't show any changes in folding kinetics in the presence of E.Coli Hsp90 (data not shown).

If the substrate problems can be overcome, *technical advances of optical tweezers setups* provide new ways to gain additional insights during experiments. Combined setups with single molecule fluorescence or FRET detection facilitate measurements of an additional reaction coordinate. These experiments would render it possible to measure the conformations of the Hsp90 machine and the client simultaneously. In addition, the important asymmetry of Hsp90 complexes could be investigated.

Furthermore, using microfluidics with different channels in the optical tweezers setup permits the measurement of the same molecule in different environments. This can be used to show the nucleotide dependence on the same Hsp90 molecule. With different co-chaperones and substrates in different channels, one could follow the piece-by-piece assembly and chaperoning process of the Hsp90 machinery.

Part IV

APPENDIX

In the first appendix chapter an overview on all protein samples is given. Methods and results that were omitted in the main part to increase readability are shown in the following two chapters. The final chapter presents optical tweezers data of *E.coli* Hsp90.

A

PROTEIN SAMPLES OVERVIEW

The following sections give an overview on the protein constructs designed for this thesis. They also display some interesting constructs that are only briefly mentioned to ease protein design for future experiments.

A.1 YEAST HSP90 MUTANTS

The experiments mainly use protein constructs of the heat-inducible isoform Hsp82 of yeast [13], which is referred to as Hsp90 in the text. Numbering of aminoacids always relates to Hsp82 wild type.

Figure A.1 displays the monomeric mutants produced and investigated by optical tweezers. Hsp90-diUbi, Hsp90_61C_560C, N-diUbi, M-diUbi, MΔ10-diUbi and C-diUbi are used to characterize the unfolding patterns and folding behavior of Hsp90 (see Chapters 7 and 8). Choosing the M domain boundary at aminoacid position 527 as it is for MΔ10-diUbi, shows decreased protein stability (see Section C.3.4), therefore it is moved to residue 537 (M-diUbi).

MC-diUbi and MCextlink-diUbi are constructs to measure interactions between the M and C domains (experiments not shown).

Hsp90-diUbi, ΔCL-diUbi, Hsp90_61C_560C, ΔN, ΔMC, Sub211-263 and Sub211-272 proteins are utilized studying the Charged Linker in Chapter 9. In ΔCL-diUbi aminoacids 214-261, analogous to [149], are deleted. Sub211-263 and Sub211-272 are constructs, where the indicated aminoacids of the Charged Linker are replaced by glycine-glycine-serine repeats of the same length.

A.2 YEAST HSP90 DIMERIC MUTANTS

Figure A.2 displays the dimeric mutants used in fluorescence, crosslinking and optical tweezers experiments. To ensure stable dimerization at very low concentrations, fusion constructs of yeast Hsp90 with the coiled-coil motif of the kinesin neck region of *Drosophila melanogaster* at the C terminus are engineered (see also [87]). For fluorescence experiments the cysteines of constructs Hsp90zip_61C, Hsp90zip_385C, Hsp90ΔCLzip_61C, Sub211-263zip_61C and Sub211-272zip_61C are labeled with fluorescent dyes and monomers are exchanged to create FRET pairs (see Section B.3.2). In the main text these constructs are called WT, ΔCL, Sub211-263 and Sub211-272 for better comprehensibility.

The first two constructs are identical to those in [87].

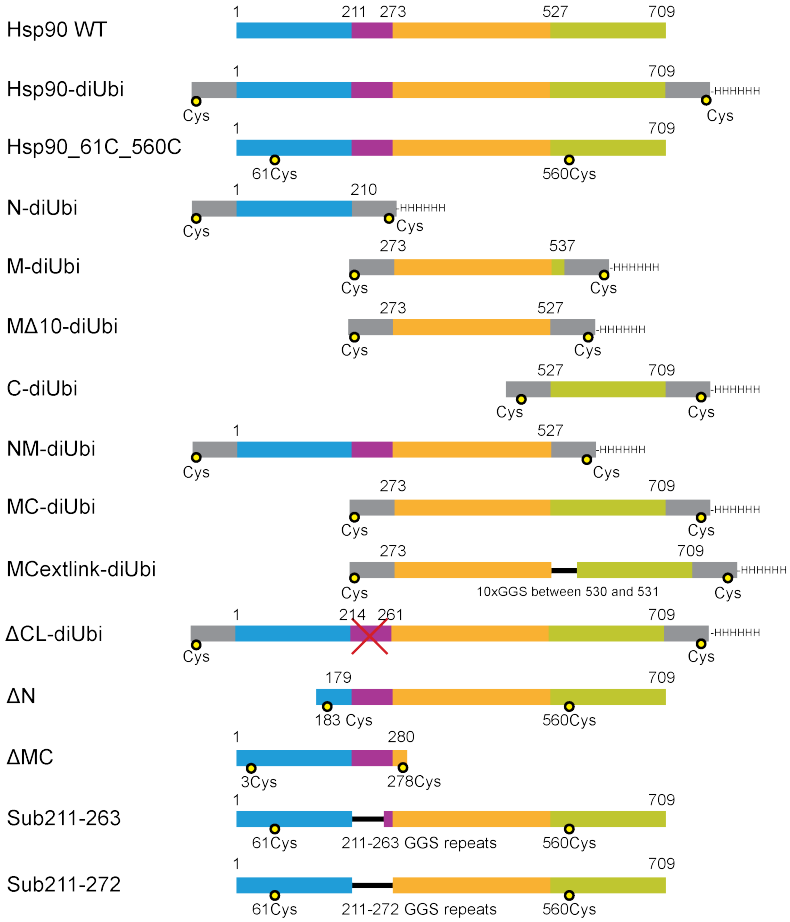

Figure A.1: Overview on monomeric Hsp90 constructs. Colored regions mark the N domain (blue), the Charged Linker (purple), the M domain (orange) and the C domain (green). Many constructs are expressed as fusion proteins with human ubiquitins shown in gray. Cysteines that are used for DNA-protein coupling are shown as yellow circles.

The stability of the coiled-coil is not sufficient to withstand force in optical tweezers experiments, because it dissociates at around 10 pN [14]. Therefore an additional cysteine is introduced to covalently cross-link both monomers. Initial attempts replacing the aminoacid (at position 712) between the C-terminal end of Hsp90 and the beginning of the coiled-coil didn't yield cross-linked monomers. However, replacing the first aminoacid (residue 715) of the coiled-coil, which is an a-position and therefore pointing directly to the opposing cysteine residue, yields covalently linked monomers. For oligonucleotide attachment on the dimeric construct the position of the second cysteine is critical (see Section 10.1), hence Hsp90zip_61C_715C and Hsp90zip_452C_715C can be used, but Hsp90zip_9C_715C not. The bottom-most construct, Hsp90zip_61C_560C, is designed to study monomer unfolding, while the opposing monomer (Hsp90zip) is hybridized.

Also the Strep-tag, which may interact with the streptavidin beads, is removed.

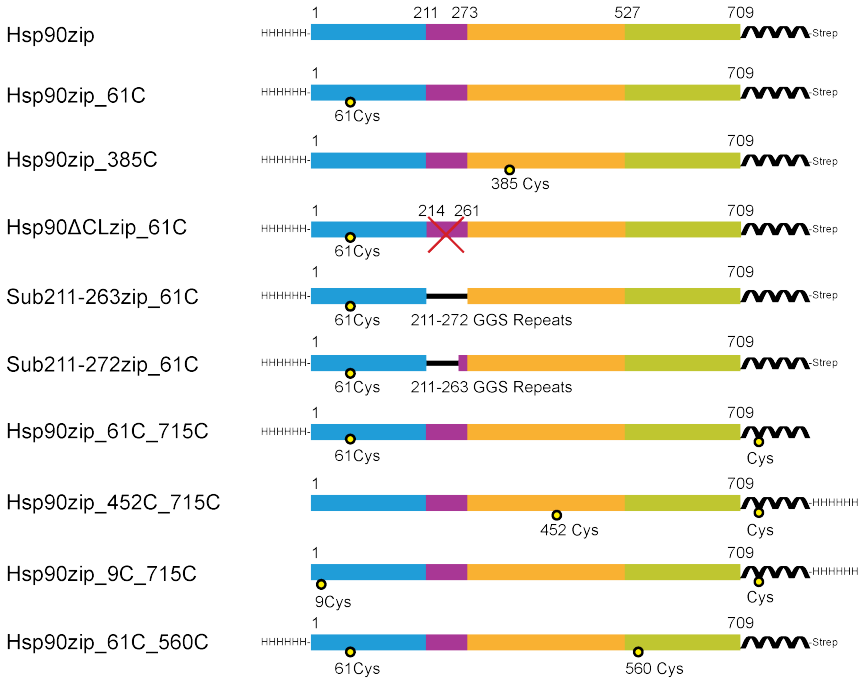

Figure A.2: Overview on dimeric Hsp90 constructs. Domains and cysteines are labeled as in Figure A.1. To stabilize the Hsp90 dimer, all constructs have C-terminal helices, which form coiled-coils upon dimerization. Hsp90zip_61C, Hsp90zip_385C, Hsp90ΔCLzip_61C, Sub211-263zip_61C and Sub211-272zip_61C are used in fluorescence and cross-linking experiments. Hsp90zip_61C_715C, Hsp90zip_452C_715C, Hsp90zip_9C_715C and Hsp90zip_61C_560C are used in optical tweezers experiments. The first three have cysteines within the helices to form covalent disulphide bonds with the opposing monomers.

A.3 E. COLI HSP90 MUTANTS

Figure A.3 shows monomeric and dimeric mutants of *E. coli* Hsp90 (HtpG). HtpG-diUbi and HtpGΔC-diUbi are used to characterize the unfolding pattern of HtpG, as shown in Section D.1. HtpG_61C_521C is a ubiquitin free construct, in analogy to Hsp90_61C_560C, but it performed badly in experiments for unknown reasons.

HtpGzip_61C_632C, HtpGzip_3C_631C and HtpGzip_3C_634C are designed to form dimers as those for yeast (see Section A.2). They vary in the linker length and composition between the HtpG and the coiled-coil motif. Since the first two constructs rarely formed dimers and mostly monomers are measured the linker length is increased. HtpGzip_3C_634C has a slightly longer linker and yields an acceptable amount of dimers (see Figure D.3). Different dimer-cross-linking strategies or cysteine cross-links within the HtpG sequence could fur-

HtpG doesn't have unstructured regions at the C-terminus, therefore the linker length may be critical.

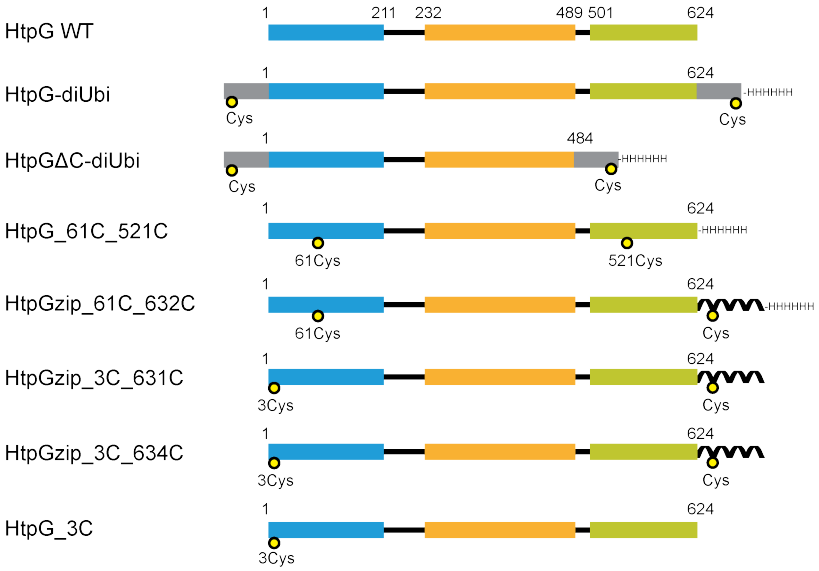

Figure A.3: Overview on HtpG constructs. Domains and cysteines are labeled as in Figure A.1.

ther improve the results. Another possibility is to rely on the dimerization properties of the protein itself, like in the construct HtpG_3C.

A.4 CO-CHAPERONES AND SUBSTRATES

Figure A.4 shows constructs related to Hsp90. The co-chaperone p23 /Sba1 and Aha1 consist of two domains and are used in some experiments. Moreover a cysteine free Aha1_nocys variant is produced, which would be useful for optical tweezers and single molecule FRET experiments. Unfortunately, this mutant is not very stable, therefore a mutant with an unnatural artificial aminoacid (uaa) is expressed, which is a sophisticated way to introduce a unique functional group into the protein. This group can be targeted specifically and used for dye labeling or oligonucleotide attachment.

These Aha1 mutants were produced as part of the bachelor's thesis of Maximilian Mühlbauer.

The unstable fragment of staphylococcal nuclease Δ131Δ [1] is a pseudo substrate for HtpG [140]. Unfortunately investigating it directly in the optical tweezers assay didn't work so far, because the positively-charged protein aggregates with the negatively-charged oligonucleotides.

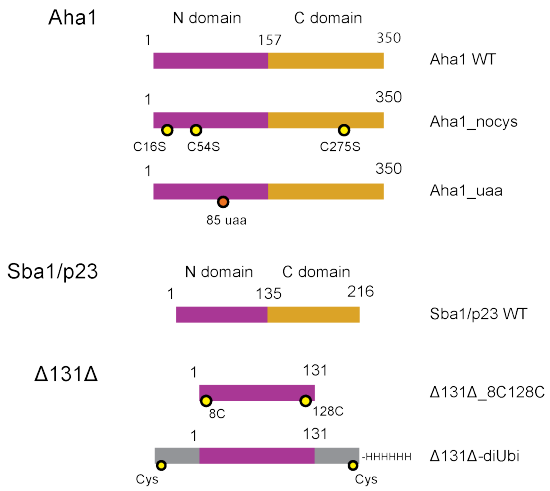

Figure A.4: Overview on constructs of the co-chaperones Aha1 and p23/Sba1 and the HtpG client Δ131Δ of staphylococcal nuclease.

B

SUPPLEMENTAL METHODS

B.1 PROTEIN PREPARATION

The basis of protein preparation of the samples shown in Chapter A are described in the following three sections (see also [54]).

B.1.1 *Molecular Cloning*

Most plasmids, coding for proteins, are modified using standard biomolecular techniques such as restriction and insertion or PCR. Some genes, synthesized by assembly PCR, are purchased commercially. For plasmid multiplication *E. Coli* strains XL1, XL10, DH5α or DH10b are used.

B.1.2 *Protein Expression*

For protein expression plasmids are transformed into *E. Coli* strain BL21DE3 cod+. Bacteria are grown in LBo with antibiotics kanamycin or carbenicillin, dependent on the plasmid's resistance. At an optical density (OD) of 0.6 protein expression is induced by the addition of ITPG or arabinose, dependent on the expression system (T7 or ParaBad). Proteins are expressed for 3 to 4 hours at 37°C or overnight at 20°C.

B.1.3 *Protein Purification*

Pelleted bacteria containing the expressed His-tagged proteins are resuspended in loading buffer (100 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8) and lysed in a French Press. They are loaded onto a Ni-NTA tag column (GE Healthcare), washed with loading buffer and eluted by a linear gradient with elution buffer (100 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8).

Some proteins have an N-terminal SUMO domain (Small ubiquitin-related modifier), which carries the His-tag and can be cleaved specifically. If the construct has a SUMO tag it is digested over night with SENP protease during dialysis to loading buffer and additionally purified on the Ni-NTA column by collecting the flow-through. This strategy using the affinity column twice enhances protein purity significantly. Strep-tagged proteins are purified on a Strep-Tactin Superflow column (IBA) with supplied buffers and protocols instead.

This is the case for all constructs shown in the figures of Chapter A that have no remaining affinity tag.

All constructs are further purified by anion exchange on a Q HP or a Mono Q column (GE Healthcare) using a linear gradient between low salt and high salt buffer (40 mM Hepes, 50 mM / 1 M NaCl, 1 mM TCEP, pH 7.5). For Aha1 and Δ131Δ a cation exchange SP HP column (GE Healthcare) and buffers with a pH of 6 are used instead. If the protein still contains impurities after the aforementioned steps, it is further purified by size exclusion on a Superdex 200 column (GE Healthcare) in 40 mM Hepes, 200 mM KCl, 1 mM TCEP, pH 7.5.

Protein purity is tested by SDS-PAGE after each purification step. Aliquots of the purified samples are frozen in liquid nitrogen and stored at -80°C.

B.2 OPTICAL TWEEZERS ADDITIONAL METHODS

B.2.1 *Assembly and measurement of optical tweezers constructs*

To bind single-stranded DNA oligonucleotides to the samples, the proteins' cysteines are reduced by adding 10 mM of TCEP for 10 to 15 minutes at 20°C. The buffer is rapidly exchanged to PBS with a pH of 6.7 and remaining TCEP is removed. The concentrations of the protein samples are adjusted to $10-200\,\mu M$ and a 2-fold molar excess per cysteine residue of 3'-maleimide-labeled, 34-basepair-long DNA oligonucleotides (Biomers) is added. After 1 to 2 hours of incubation at room temperature the sample is purified with a Superdex 200 gel filtration column (40 mM Hepes, 200 mM KCl, pH 7.5) and unreacted oligonucleotides are removed (for an example see Section C.1).

Proteins with DNA oligonucleotides are then hybridized for about 1 hour at room temperature with the single stranded overhang of 545-basepair-long DNA handles. These DNA handles are amplified from lambda phage using PCR similarly as previously described [18]. The quality of the oligonucleotide reaction as well as the concentration ratio between constructs and handles can be analyzed by agarose gels, where unbound DNA handles, construct with one DNA handle and construct with two DNA handles can be distinguished very well.

Running different handles-protein ratios can be used for optimization.

The DNA handles have two different modifications, to bind to the two types of silica beads. One half of the handles has 3 biotins to bind to streptavidin-coated 1 µm silica beads (Bangs Laboratories), the other half has 3 digoxigenin modifications to bind the respective antibodies on the in-house-functionalized, fluorescent 1 µm silica beads.

The handle-protein constructs are incubated for several minutes with one of the bead types and then diluted in measurement buffer (40 mM Hepes, 150 mM KCl, 10 mM $MgCl_2$, pH 7.4), that already contains the other bead type. To avoid photodamage by radicals a scavenger system containing glucose, glucose oxidase and glucose catalase is used. The sample is filled into a BSA-passivated custom-built

fluid chamber, mounted on the optical tweezers setup and beads are trapped by the laser beams. At the laserpowers used, the temperature at the spot of the tethered protein construct is approximately 30°C.

B.2.2 *Automatic classification of force-extension traces*

To evaluate the protein states in the force-dependent refolding experiments (see Sections 8.5.1, 8.6.3 and 8.7.2) tens of thousands force-extension traces are analyzed. Figure B.1 describes how this analysis is automatized.

After the assignment all traces of each class are plotted on top of each other and checked manually. This is necessary because sometimes sometimes subtle features (see Section 7.3) of the N or M domain are missing although the event is classified as native. Rarely, proteins don't refold for many successive cycles, although the conditions would favor refolding. Such *dead states*, which may arise from isomerization of aminoacids, are observed for all Hsp90 domains and are neglected for refolding probability evaluation.

B.2.3 *Uncertainties in optical tweezers experiments*

Measurements and analyses of optical tweezers experiments have different error sources. Major errors during measurements are from the force calibration and drifts. During analysis WLC fits may also contain some error due to incorrect fitting or incorrect zero values.

The uncertainties displayed for the force-dependent state occupancies from constant distance experiments are estimated from the error due to the finite length of the measurement as described in [139]. Errors displayed for the force-dependent population lifetimes are those from the single exponential fits to the lifetimes histograms. They are the only errors that are considered, while extrapolating rates to zero force. Since uncertainties in force calibration and WLC fits are neglected, the error might be higher. However, the data presented here are mostly measured at low forces, therefore the extrapolation range is short and the additional errors will be limited.

B.2.3.1 *Uncertainties of the refolding probabilities*

Uncertainties of the force-dependent refolding probabilities are estimated by the Clopper-Pearson confidence interval (95%). For the refolding probabilities of the isolated domains, an additional error estimating the probability change for a 0.2 pN force shift (divided by the square root of experiments at this waiting time and waiting force) is added.

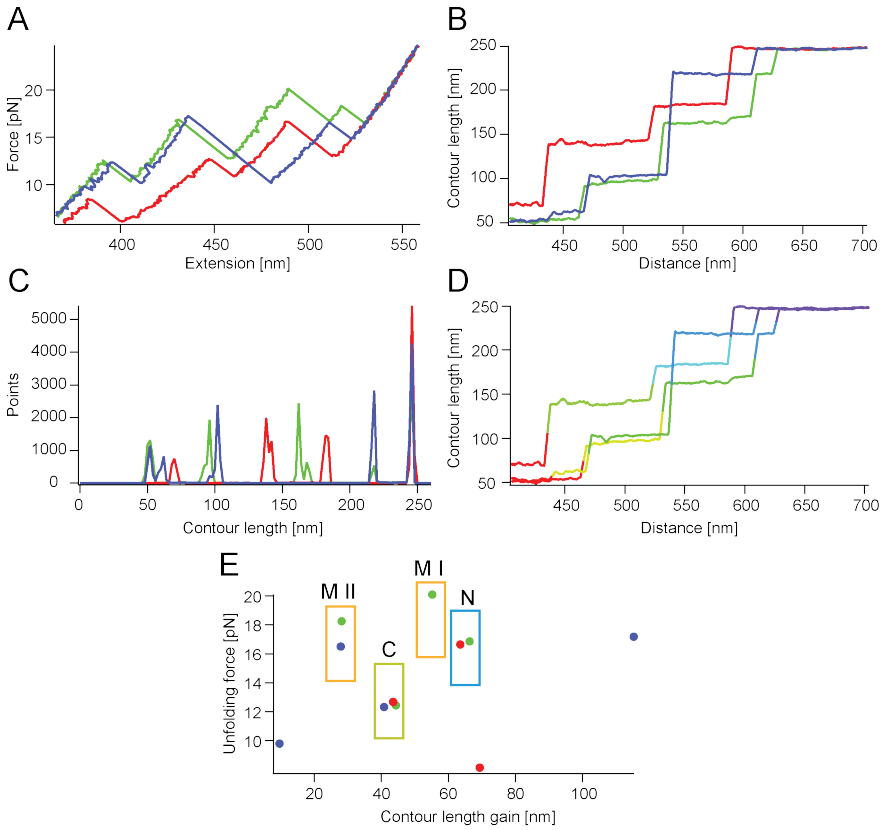

Figure B.1: The multistep classification procedure for 3 example traces is shown. (A) Three differently folded force-extension traces of Hsp90-diUbi. (B) Traces are transformed into contour length space by calculating the contour length for every point with the previously determined DNA parameters. (C) Contour length histograms of Fig. B shows the contour lengths gains of the force-extension traces. Using a peak finding routine puts numbers on them. (D) They are used to assign states (rainbow colors) to force-extension traces. Finding the local maxima of the assigned regions yield the associated unfolding forces. (E) After the contour length gains and the unfolding forces of all events are determined, they are collected in a scatterplot. Classification of the natively-folded events is done by a deterministic approach, where each event that has a certain contour length gain and unfolding force within some tolerance is assigned to the respective state. The squares show the approximate regions for the 3 Hsp90 domains.

B.3 FLUORESCENCE EXPERIMENTS THEORY AND METHODS

B.3.1 *FRET theory*

Förster resonance energy transfer (FRET) is a common tool to study biomolecular systems (see Chapter 3). If two fluorescent dyes, with overlapping emission spectrum of the first dye (donor) and absorption spectrum of the second dye (acceptor), are brought into close proximity, energy from the excited donor is transferred to the acceptor by a radiation-free dipole-dipole interaction. The distance dependent efficiency $E(r)$ of this transfer is:

$$E(r) = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \tag{B.1}$$

R_0 is the Förster radius, which is typically around 5 nm. Due to the power law behavior the FRET efficiency is very sensitive to small length changes around the Förster radius. R_0 depends on various parameters, like the overlap integral of the donor-acceptor pair, the relative orientation of the dipoles and the surrounding environment.

In single molecule FRET experiments the donor and acceptor intensities I are measured to calculate the FRET efficiency.

$$E(r) = \frac{I_{acceptor}}{I_{acceptor} + \gamma I_{donor}} \tag{B.2}$$

γ is the difference between the quantum yields of both dyes. To obtain the correct values, also the leakage from one detector channel into the other and the direct acceptor excitation have to be considered.

B.3.2 *Fluorescence labeling and monomer exchange*

Figure A.2 shows the protein constructs used for FRET experiments. The wild type and the mutants with the cysteine at position 61 are labeled with the donor dye Atto 550 (Atto-tec) and the wild type with the cysteine at position 385 is labeled with the acceptor dye Atto 647N using maleimide chemistry as described in [87].

For the single molecule FRET experiments the constructs with the donor and the acceptor dyes are mixed, yielding heterodimers with donor-acceptor pairs. The monomer exchange protocol is described in [87].

For surface-based TIR-FRET experiments the constructs with the acceptor dye are unspecifically biotinylated with NHS-PEG-Biotin for surface immobilization after dye labeling [87].

B.3.3 *Confocal and TIRF smFRET experiments*

FRET efficiencies of freely diffusing single molecules are determined in a confocal, custom-built setup similar to those described in [57, 91]. Labeled proteins diffusing through the confocal volume are excited by two pulsed lasers and the fluorescence intensities of the donor and the acceptor are recorded by avalanche photo diodes. If the total intensity of a burst exceeds a certain threshold value, the FRET efficiency and the stochiometry of this event are collected in a 2D histogram. Selecting stochiometries of around 0.5, which is the case if one donor and one acceptor are present, yields the histograms shown in Figure 9.6. For nucleotide-dependent measurements labeled Hsp90 dimers are preincubated with the respective nucleotide for 1 hour, before dilution to around 50 pM. The temperature in the fluid chamber is around 23°C.

TIR-FRET measurements are carried out in a custom-built prism-type TIRF microscope, which is described in [116]. The samples are immobilized on neutravidin-functionalized and PEG-coated quartz slides. The donor is illuminated by a green laser and fluorescence light of the donor and the acceptor are monitored on an ultra-sensitive EMCCD camera (Andor). The measurement interval is 220 ms, composed of 200 ms illumination and 20 ms camera readout. The donor and acceptor fluorescence trajectories are converted into FRET trajectories, for examples see Figure 9.7. FRET efficiencies are smoothed with a sliding average of 5 frames and multiple traces are pooled into FRET histograms (see Figure 9.7 E). Measurements are performed at 30°C.

For all smFRET experiments the same buffer as in optical tweezers experiments is used (40 mM Hepes, 150 mM KCl, 10 mM $MgCl_2$, pH 7.4).

B.3.4 *Bulk fluorescence*

The change in the fluorescence properties of a single dye attached to the N domains of Hsp90 can report on the N-terminal conformational changes. Therefore protein samples are labeled like the donor molecules described in Section B.3.2. Heterodimers of CL mutants and wild type are generated by monomer exchange (see Section B.3.2) adding a 10-fold excess of Hsp90zip (see Figure A.2).

The fluorescence intensity changes of 500 nM labeled protein after the addition of different ATP analoga are recorded in a Jasco FP-8500 spectrofluorimeter (530 nm excitation/ 580 nm emission). Results are shown in Figure 9.8. There all intensity traces are normalized to 1 upon the addition of nucleotide. Buffer is the same as in optical tweezers and smFRET experiments.

B.4 CROSS-LINKING EXPERIMENTS

Hsp90zip_61C, Sub211-263zip_61C and Sub211-272zip_61C (see Figure A.2) with a concentration of 3 µM are subject to cross-linking experiments. Before the cross-linking reaction proteins are freshly reduced and buffer is exchanged to PBS pH 6.7. Then 10 µM of cross-linking agent BM(PEG)3 (1,11-bis(maleimido)triethylene glycol) that can bind two cysteines, is added and incubated for 15 minutes at room temperature. The reaction is quenched by adding an excess of DTT and samples are analyzed on a reduction-free 10% SDS-PAGE gel. Buffer is the same as in optical tweezers and smFRET experiments.

B.5 MEASUREMENT OF ATPASE

The ATPase activity is determined by an enzyme-coupled, ATP-regenerating assay described in [145, 100]. Measurements with protein concentrations of 3 µM in 40 mM Hepes, 150 mM KCl, 5 mM $MgCl_2$, pH 7.4 are carried out at 30°C. For the Aha1 activation measurements protein concentration is reduced to 1 µM and the buffer's KCl concentration is reduced to 20 mM.

B.6 ANALYTICAL ULTRACENTRIFUGATION EXPERIMENTS

Unspecifically labeled co-chaperones 0.5 µM Aha1-FAM or 1 µM Sba1 /p23-Atto488 are mixed with 3 µM Hsp90 WT or mutants and the complex formation is analyzed by analytical ultracentrifugation with fluorescence detection (Beckman/Coulter) [72]. Buffers are the same as in the ATPase measurements and the interaction assays involving Sba1/p23 are done in the presence of 2 mM AMP-PNP.

C

SUPPLEMENTAL RESULTS

C.1 COMPARISON OF DIFFERENT ATTACHMENT CHEMISTRIES

The first attempts attaching the typically used thiol-modified oligonucleotides to the Hsp90 protein didn't give satisfying results. Therefore the conjugation method is changed to maleimide oligonucleotides, as described in Section B.2.1. Maleimide-thiol chemistry offers a more reactive and easier-to-use alternative compared to the disulphide bond formation. This is seen in gelfiltration runs displayed in Figure C.1. A complete reaction is important, because for large proteins it is impossible to separate unreacted protein or proteins with only one oligonucleotide from protein with two oligonucleotides attached.

To circumvent this problem anion exchange purification instead of gelfiltration could be used.

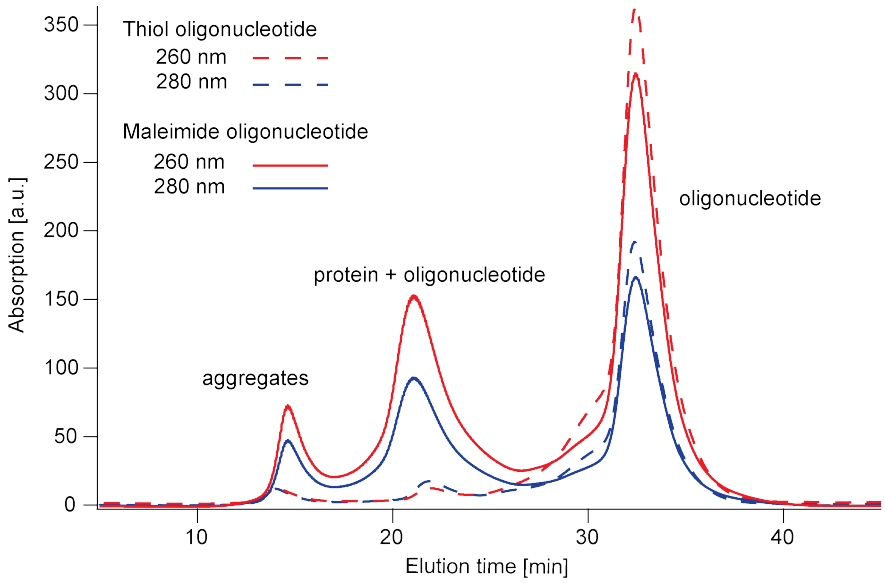

Figure C.1: The chromatograms (260 nm and 280 nm) of two gel filtration runs are overlaid. For one thiol oligonucleotides are reacted with Hsp90-diUbi, for the other one maleimide oligonucleotides are used. The latter yields significantly more labeled Hsp90 molecules, which can be seen by the high peak after 22 minutes.

C.2 UNFOLDING OF HSP90 USING AFM

In first mechanical experiments yeast Hsp90 is characterized with a custom-built AFM. Figure C.2 displays the setup and example unfolding force-extension traces for four selected protein constructs. To exert force to the termini of the proteins, Hsp90 variants are expressed as fusion proteins between the domains filamin 1-5 of *Dictiostelium diodecum*. The fourth filamin domain, which is the first domain at the C-terminal end of the investigated protein, has a characteristic unfolding intermediate. If three filamin domains are observed and one of them is the fourth domain, also the Hsp90 variant is subject to force. This fingerprint is important to discriminate between properly-pulled and incorrectly-pulled proteins. The occurrence of unusable force-extension traces is relatively high, because picking up a molecule relies on unspecific interactions between the first two filamin domains and the cantilever. Furthermore the large surfaces involved can lead to unwanted interactions.

The fusion proteins are only specifically attached to the surface at the C-termini via their His-tags.

However, the contour length gains obtained for the individual domains match the values obtained in optical tweezers experiments (compare to Table 7.3). Although the force-loading rates are typically higher than in optical tweezers experiments unfolding events take place at around 20 pN which is at the lower end of AFM resolution. Due to the limitations described, it is almost impossible to measure the large full-length Hsp90 construct.

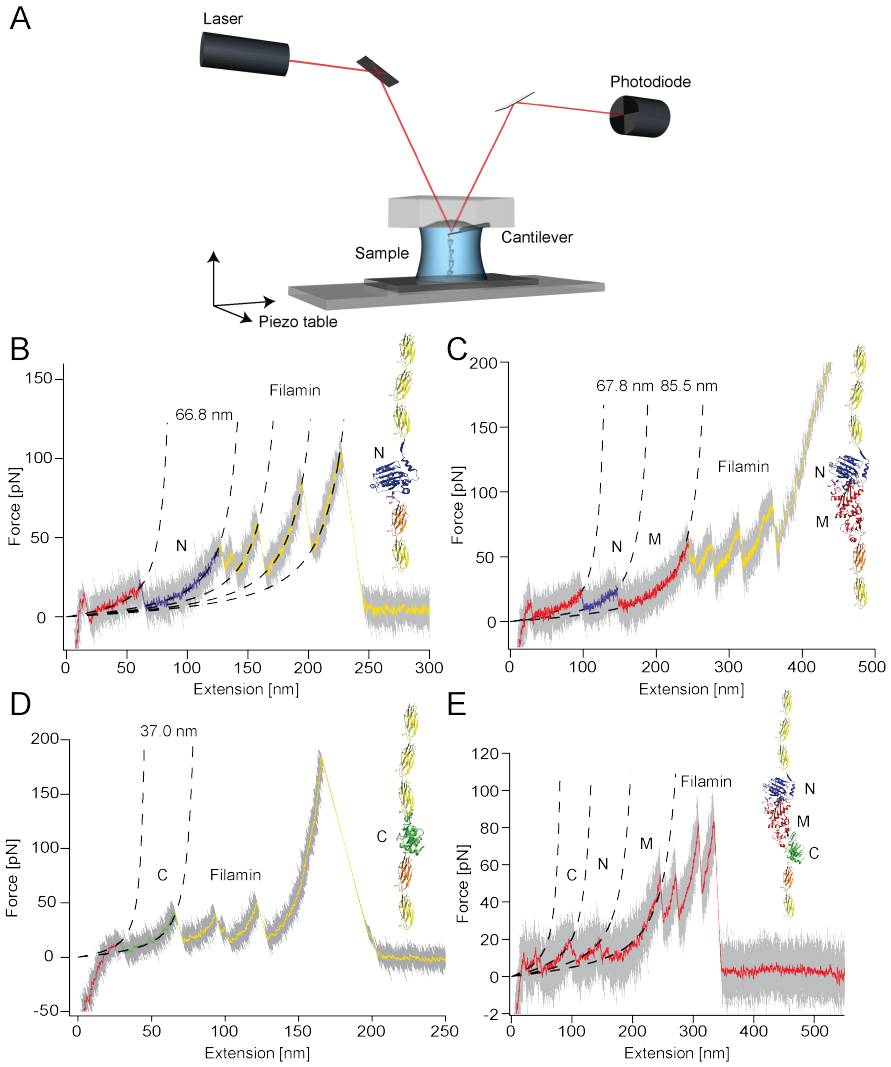

Figure C.2: (A) In AFM experiments the sample is clamped between the cantilever and a moveable surface. Forces are measured by the deflection of a laser beam reflected on the backside of the cantilever. (B) An unfolding force-extension trace of the N domain fusion construct (see inset). At low forces the N domains unfolds (blue) followed by the unfolding of the filamin domains (yellow). The contour length gains are determined by WLC fits and the average value of the N domain from multiple experiments is displayed. (C,D) Same as Fig. B with fusion constructs comprising the N and M domain or the C domain, respectively. (E) While the characterization of the individual domains works well, full length traces are hard to obtain. Here one of the few traces, which looks very similar to the full length construct in optical tweezers is shown.

C.3 ADDITIONAL HSP90 FOLDING DATA

C.3.1 *Refolding of Hsp90 in rapidly-pulled cycles*

Figure C.3 shows unfolding and refolding force-extension cycles at a pulling speed of 500 nm/s. Unlike the slowly-pulled traces (see Figure 8.1), Hsp90 doesn't always populate the native state with all domains fully folded in consecutively-pulled cycles.

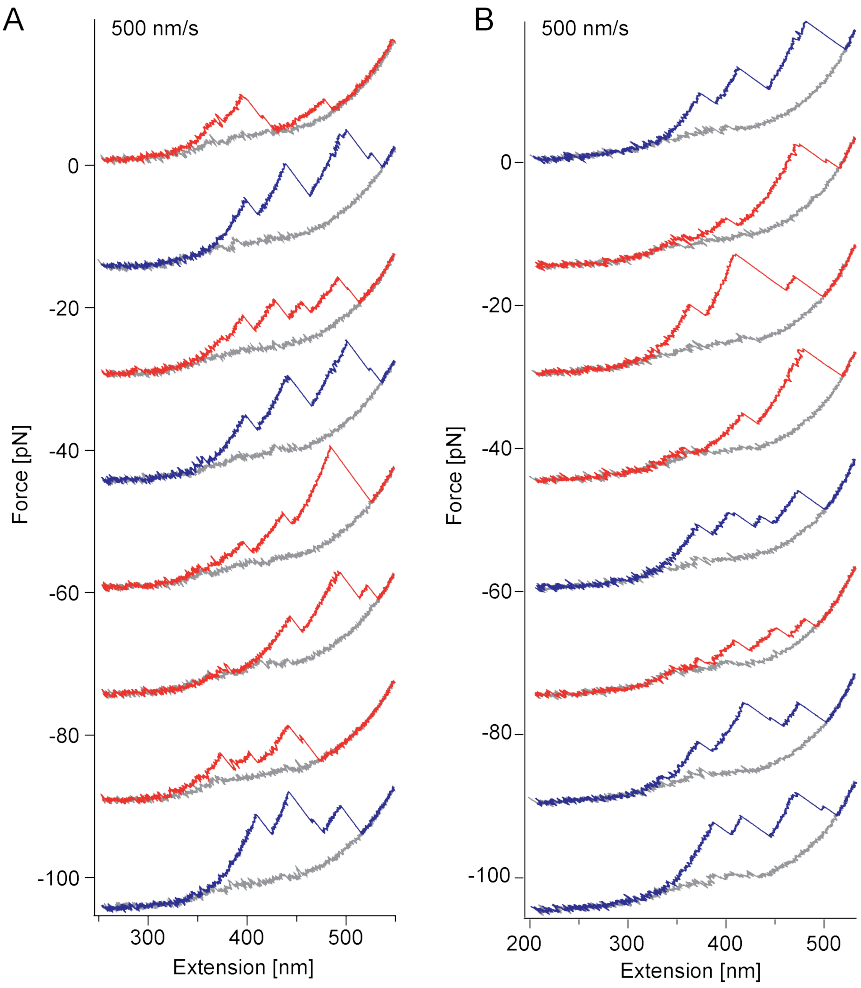

Figure C.3: Consecutively-pulled, offsetted, 500 nm/s constant velocity cycles of Hsp90-diUbi for two example proteins (A) and (B). The unfolding traces at this speed show the fully folded states (blue) as well as non-native or partially native states (red). Refolding traces are shown in gray. The refolding probabilities of such traces are used to estimate the overall refolding rate of the Hsp90 monomer in Section 8.1.

C.3.2 *Additional N domain folding data*

Figure C.4 shows more examples of unfolding and refolding cycles of the isolated N domain and compares them to simulated refolding force-extension traces. The Monte Carlo simulation assumes three states connected by the rates given in Figure 8.4 D. The transition state of the rates from and into the misfolded states is chosen in a way to fulfill the constrains of the model. The assumed model with the data obtained from the double-jump ramp experiments can reproduce the force-extension traces reasonable well. It is obvious that the simulation can't capture the heterogeneity of the misfolded states.

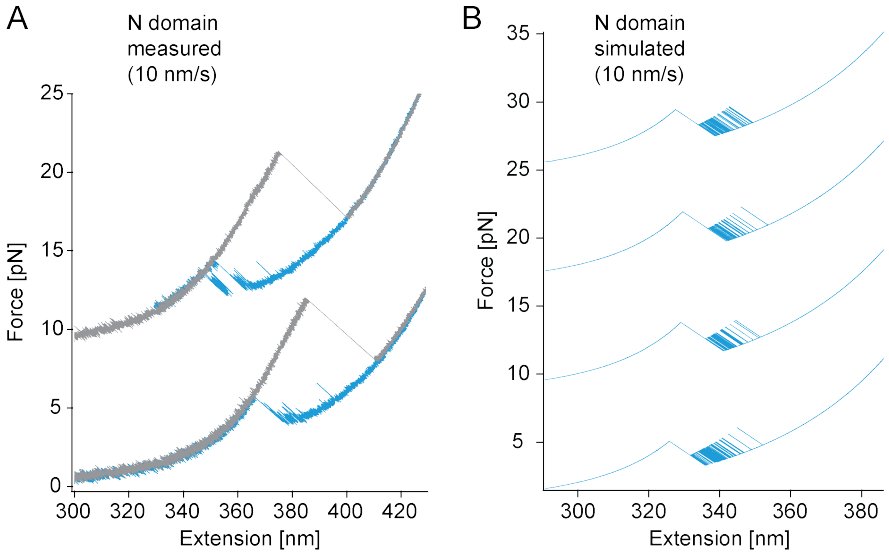

Figure C.4: (A) Slowly-pulled, offsetted (10 nm/s) unfolding (gray) and refolding (blue) traces of N-diUbi. (B) Simulated constant velocity (10 nm/s) refolding traces of the N domain describe the overall refolding behavior well. The rates for this simulation are those obtained from the refolding probabilities measured by double-jump ramp experiments. Simulation traces are offsetted.

Figure C.5 A shows a small part of a double-jump ramp experiment of the isolated N domain. Apart from the unfolding ramps, which are used to extract the refolding probabilities, also the observation of the protein's folding behavior at constant trap distance during the waiting time is possible, if forces are high enough to obtain a signal (see Figure C.5 B). The dynamic conformational search and the population of different intermediate states is clearly visible until the N domains reach the stable native state. Figure C.5 C shows an example of such a trace for a very high waiting force. In this case the molecule populates mainly the unfolded state and small excursions to intermediate states with different contour length are observed.

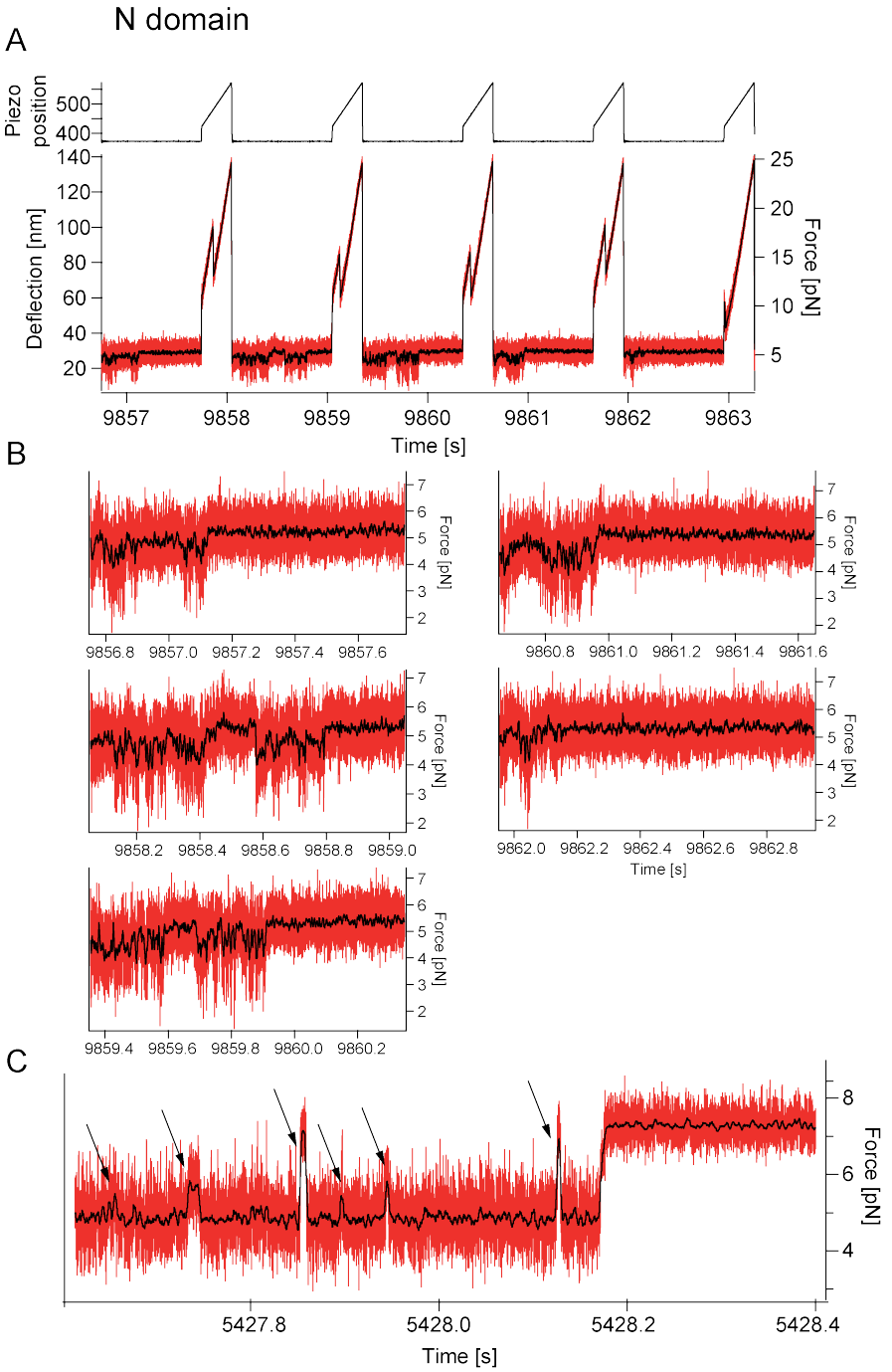

Figure C.5: (A) Example of a double-jump ramp experiment of the N domain. The measured piezo signal is displayed on top of the measured force trajectory. (B) Zooms to the force trajectory after the jump down show the dynamic refolding process. If the native state is reached, in this case at around 5.4 pN, the molecule remains stable. However, a stable state doesn't necessarily mean a natively-folded state, as it the case for the central, right-hand trace. (C) At high waiting forces, mainly the unfolded state is populated and the heterogeneity in contour lengths of the misfolded states (arrows) can be clearly seen.

C.3.2.1 *Connectivity of misfolded states*

In Figure C.6 one of the constant distance traces of Figure C.5 B is assigned by HMM assuming four equidistant states. The state at lowest force represents the completely unfolded state. On a closer look many transitions between intermediates with different contour lengths are observed. On the resolution of the experiment they don't visit the completely unfolded state, which suggests that interconversion between different misfolded states is possible.

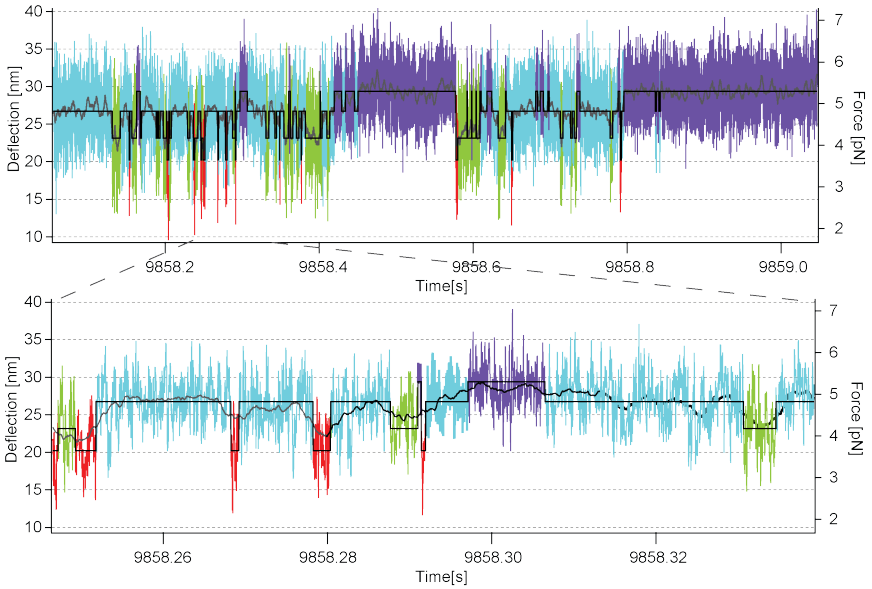

Figure C.6: The points of a constant distance trajectory of a not-yet-folded N domain are assigned by HMM assuming four states (colors). The red state corresponds to the contour length expected for the completely unfolded domain. The zoom shows that the different intermediate states can be populated without visiting the red state. This suggests that interconversion of the misfolded states is possible.

C.3.3 *Additional M domain folding data*

Figure C.7 shows measured and simulated refolding force-extension traces, relaxed with a velocity of 10 nm/s. For the Monte Carlo simulation the model and the rates described in Figure 8.6 D are used. As for the N domain (see Figure C.4) the simulation can describe the overall shape of the refolding traces well, but intrinsically doesn't capture the on-pathway intermediate.

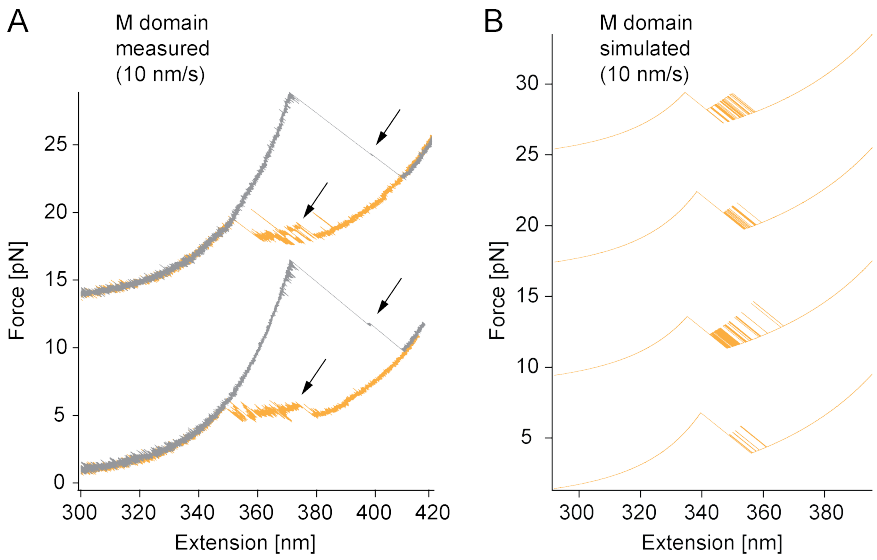

Figure C.7: (A) Example unfolding (gray) and refolding (orange) offsetted force-extension traces of M-diUbi at 10 nm/s. Black arrows mark the on-pathway intermediate present during unfolding and refolding. (B) Monte-Carlo simulation of 10 nm/s refolding traces using the rates determined by double-jump ramp experiments capture the main features of the measured refolding traces, but not the heterogeneity of the short-lived misfolding intermediates and the on-pathway intermediate. Simulation traces are offsetted.

Figure C.8 A shows a small part of a double-jump ramp experiment of the isolated M domain. Apart from the unfolding ramp, the dynamic protein folding process and the population of the intermediates during the waiting time at constant distance are observed (see Figure C.8 B). At high waiting forces the molecule populates mainly the unfolded state and the on-pathway or the misfolded intermediates are only visited for short times (see Figure C.8 C).

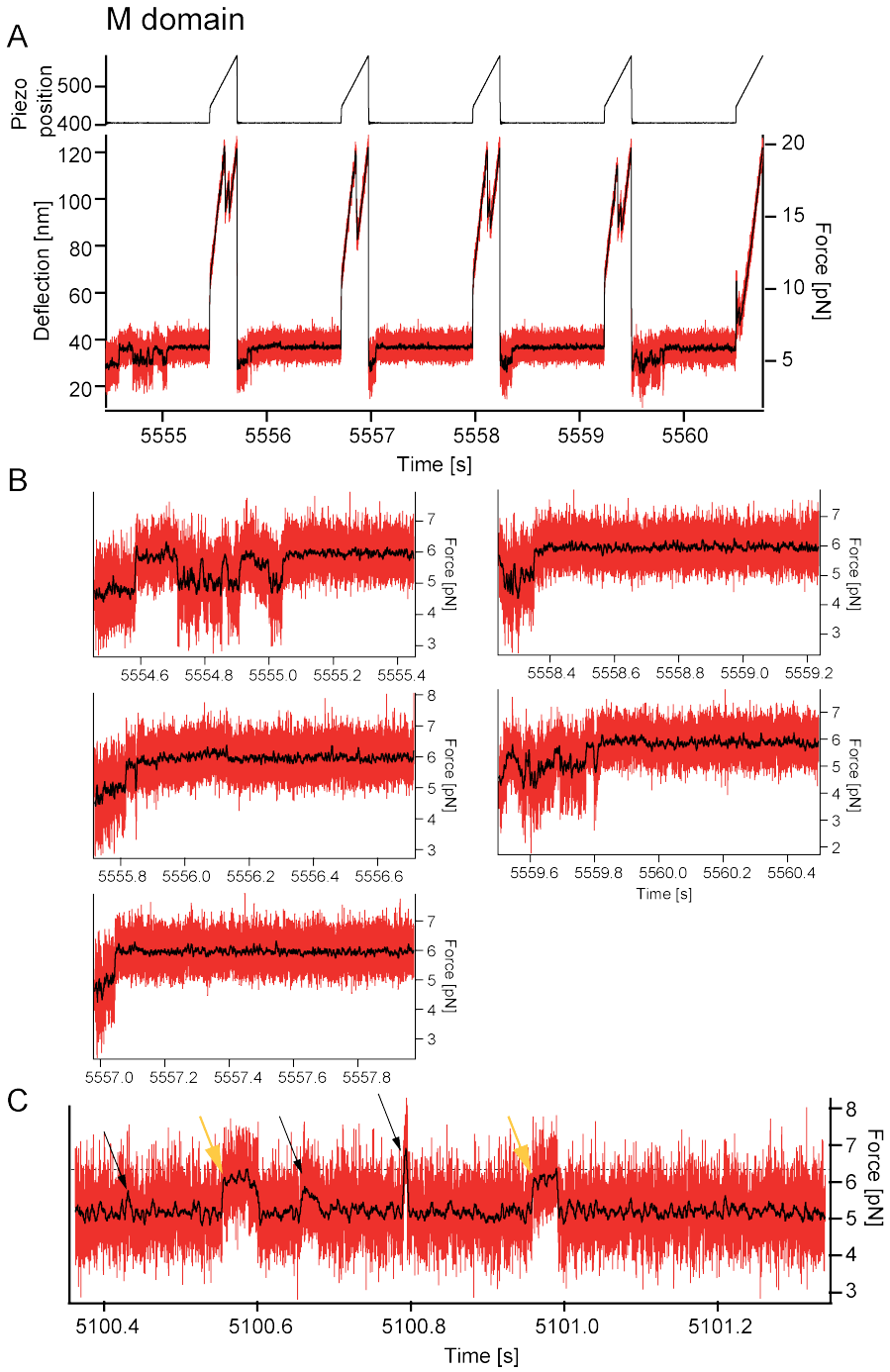

Figure C.8: (A) Example of a double-jump ramp experiment of the M-diUbi construct. The measured piezo signal is displayed on top of the measured force trajectory. (B) Zooms to the force trajectory after the jump down show the dynamic refolding process. If the native state is reached, in this case at around 6 pN, the molecule remains stable. However, a stable state doesn't necessarily mean a natively-folded state, as is the case for the central, right-hand trace. (C) At high waiting forces, mainly the unfolded state is populated, but also the on-pathway intermediate, which seems to fold in a two-state manner (yellow arrows), and different misfolded states are observed (black arrows).

C.3.4 *M domain boundaries*

Although the C-terminal domain boundary is typically assumed to be around residue 525 [2], it is found that an M domain construct, which comprises aminoacids until residue 527 (MΔ10-diUbi) behaves differently to one with aminoacids until 537 (M-diUbi). Different experiments with MΔ10-diUbi are shown in Figure C.9. Comparing them to the experiments of the longer variant shown in Figure 8.5, indicates that the C-terminal part of the M domain, which also gives rise to the on-pathway intermediate, is less stable for the shorter variant. This suggests that some aminoacids between residues 527-537 are still involved in domain formation.

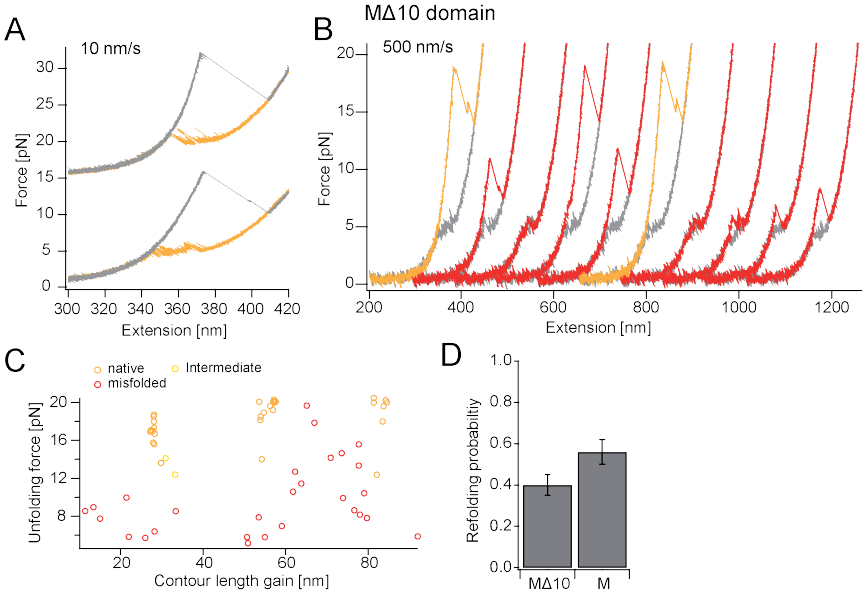

Figure C.9: (A–C) Fig. A shows offsetted force-extension unfolding (gray) and refolding (orange) of MΔ10-diUbi at a speed of 10 nm/s. Fig. B shows offsetted force-extension traces at a pulling speed of 500 nm/s. Orange unfolding traces show the natively-folded domain, red unfolding traces non-native events. Unfolding forces and contour length gains of 46 unfolding traces are gathered in a scatterplot shown in Fig. C. All these experiments indicate that the on-pathway intermediate is less populated and therefore less stable compared to M-diUbi (see Figure 8.5). This is most obvious comparing the scatterplots that show far less events where only the on-pathway intermediate is folded. (D) Also the probabilities of observing the fully-native state in constant velocity cycles with a velocity of 500 nm/s suggest lower stability for the MΔ10-diUbi construct. Its refolding probability is 40 ± 5 %, while that of M-diUbi is 56 ± 6 %. Probabilities are calculated from 304/414 unfolding traces of 10/11 molecules, respectively.

C.3.5 *Influence of subdomain structure on folding*

The model described in Section 8.6.4 treats the M domain as one folding unit. Although the measured data are well described, this view is not entirely correct, since the on-pathway folding intermediate (see Section 8.6.2) is neglected.

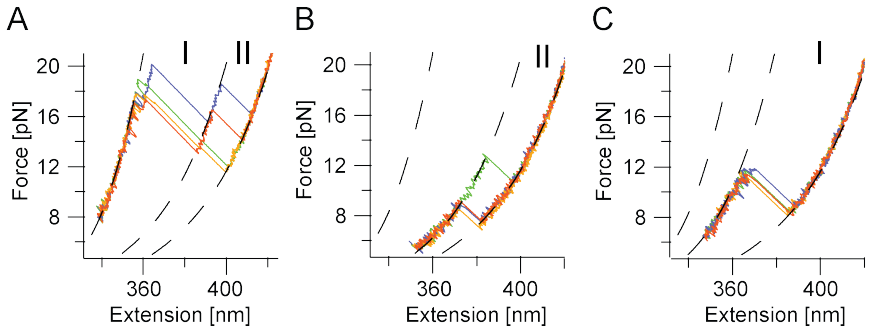

Figure C.10: Unfolding traces of M-diUbi after refolding at high waiting forces in double-jump ramp experiments. The native state with both subdomains (I and II) (A) or the natively-folded on-pathway intermediate comprising the C-terminal subdomain II (B) are observed. However, sometimes also the N-terminal subdomain I alone (C) is measured, indicating that it is stable on its own if properly folded. For each state four example traces are shown.

Figure C.10 shows unfolding force-extension traces of double-jump ramp experiments at a high refolding force, where complete refolding is rarely possible. Mainly the folded M domain intermediate is obtained, which seems to fold in a two-state manner in constant velocity (see Figure 8.5 A) and in constant distance experiments (see Figure C.8 C). However sometimes the other part of the M domain is observed, which also forms a stable intermediate on its own. If both subdomains are folded, they stabilize each other and higher unfolding forces are necessary to disrupt the M domain.

Most misfolding intermediates have contour lengths gains larger than the individual subdomains (see Figure 8.5 C), implying that both subdomains are involved in their formation. Furthermore in many unfolding traces the smaller subdomain is completely folded, while the other subdomain is in a misfolded state (see Figure 8.5 B). It can be speculated that the larger subdomain is more misfolding-prone and may form misfolding nuclei, which can misfold with the unfolded, smaller subdomain. In other words, if the smaller subdomain manages to fold, possible misfolding intermediates are prevented. This is in line with the reduced folding speed of the MΔ10-diUbi construct (see Section C.3.4).

C.3.6 *Additional C domain folding data*

Figure C.11 displays additional slowly pulled and relaxed molecules of the isolated C domain. In agreement with Figure 8.7 A, transitions between the folded and the unfolded states without noticeably populating misfolded states are observed.

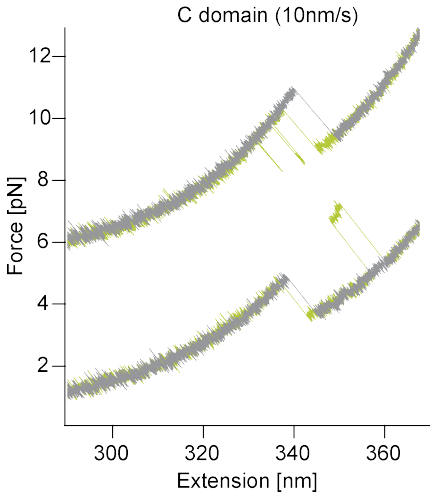

Figure C.11: Unfolding (gray) and refolding (green) offsetted force-extension traces of C-diUbi pulled and relaxed with a velocity of 10 nm/s.

Figure C.12 A shows an example section of a double-jump ramp experiment of the isolated C domain. Figure C.12 B shows enlargements of the trajectories after the jump down. In these traces the molecule either stays in the unfolded state or transitions into the native state. In longer traces, equilibrium transitions between the folded and the unfolded states are observed (see Figure C.12 C). Such experiments with even longer traces are presented and analyzed in detail in Section 8.7.4.

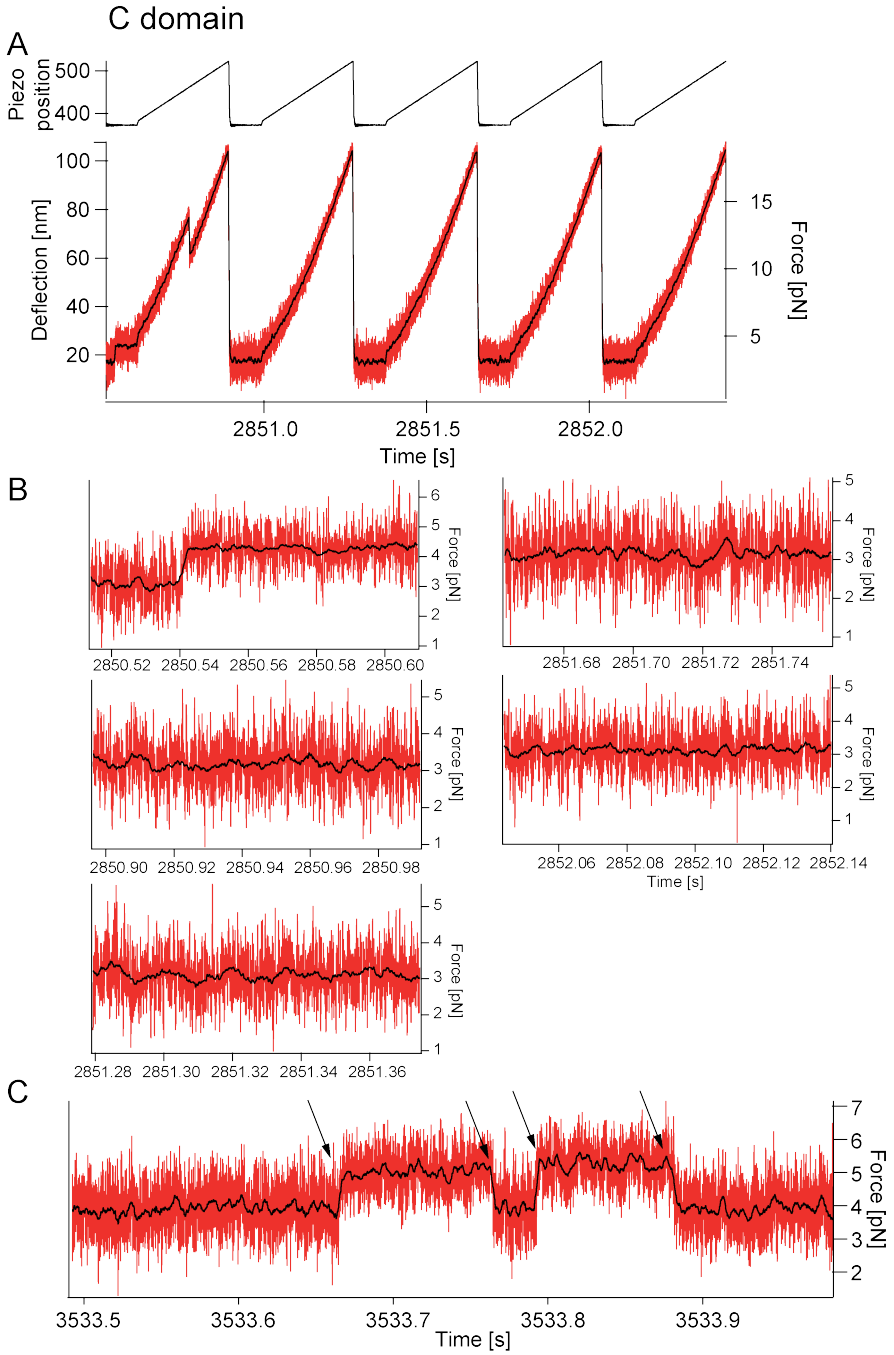

Figure C.12: (A) Example of a double-jump ramp experiment of the C-diUbi construct. The measured piezo signal is displayed on top of the measured force trajectory. (B) Close up views of the force signal of Fig. A after the jump down show no population of misfolded states. (C) In longer traces the C domain can unfold and refold during the waiting time, therefore the consideration of the unfolding rate in the model is important.

C.3.7 *Refolding of the Hsp90 monomer in double-jump ramp experiments*

Example force-extension traces of the double-jump ramp experiments with the Hsp90 monomer at two different refolding forces are shown in Figure C.13. From such traces the scatterplots shown in Figure 8.10 are extracted.

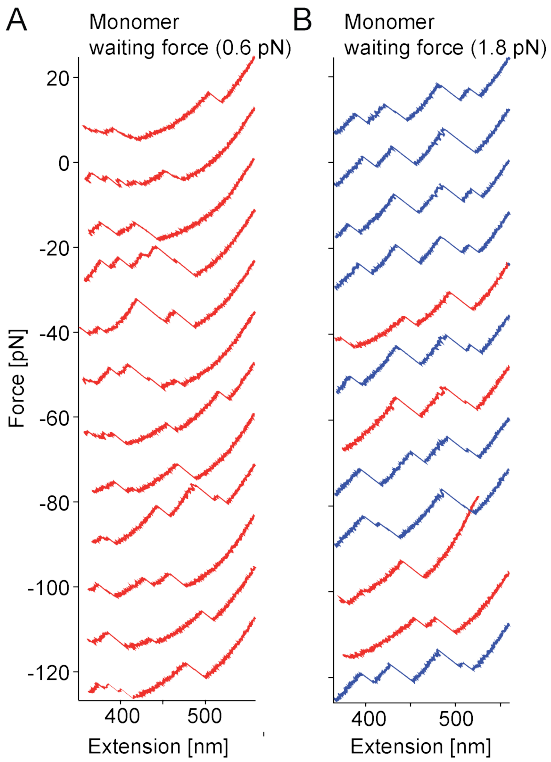

Figure C.13: Example traces of double-jump ramp experiments with Hsp90-diUbi at low waiting forces (A) and high waiting forces (B). Completely folded Hsp90 monomers are colored in blue, others in red. Hardly any natively-folded trace is found at low waiting forces. The traces are offsetted.

Figure C.14 gives additional information on the refolding properties of the Hsp90 monomer obtained by double-jump ramp experiments.

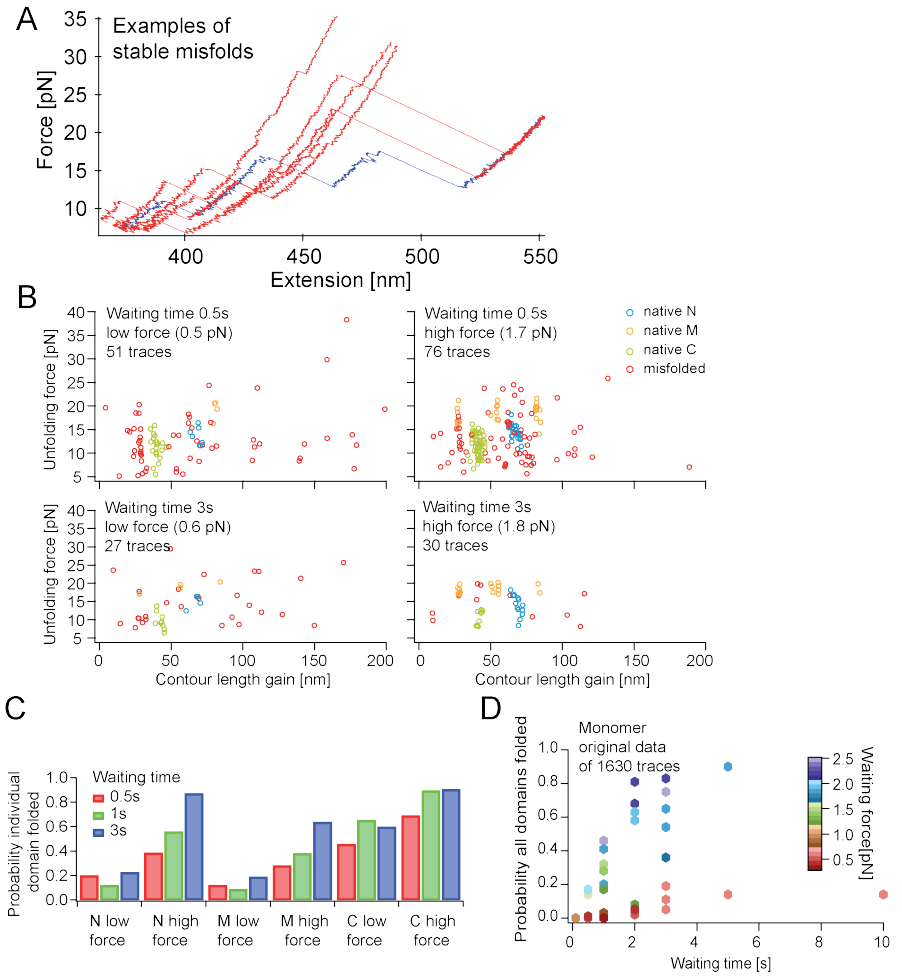

Figure C.14: (A) In double-jump ramp experiments with Hsp90-diUbi, often very stable misfolded intermediates are observed (red unfolding traces). Some of them need multiple cycles to unfold again. As a reference a native trace is shown in blue. (B) Scatterplots of unfolding events as in Figure 8.10 for waiting times of 0.5 s and 3 s. (C) The refolding probability of the individual domains within the Hsp90 monomer dependent on the waiting time and the waiting force. Example data from the scatterplots of Fig. B and Figure 8.10 show that the C domain is influenced strongly by cross-domain misfolds and that the N and M domains are trapped at low waiting forces. (D) The raw data of the probability to observe the completely folded Hsp90 after the double-jump for different waiting times and waiting forces. The averaged data in Figure 8.11 are calculated from these probabilities.

C.4 ADDITIONAL CHARGED LINKER DATA

Figure C.15 shows the full length constant velocity traces of the substitution constructs Sub211-263 and Sub211-272, whose Charged Linker regions are shown in Figure 9.2. Furthermore constant distance traces at low force, where the Charged Linker fluctuations are observed for wild type and scatterplots characterizing the unfolding of the Hsp90 domains for these constructs are shown.

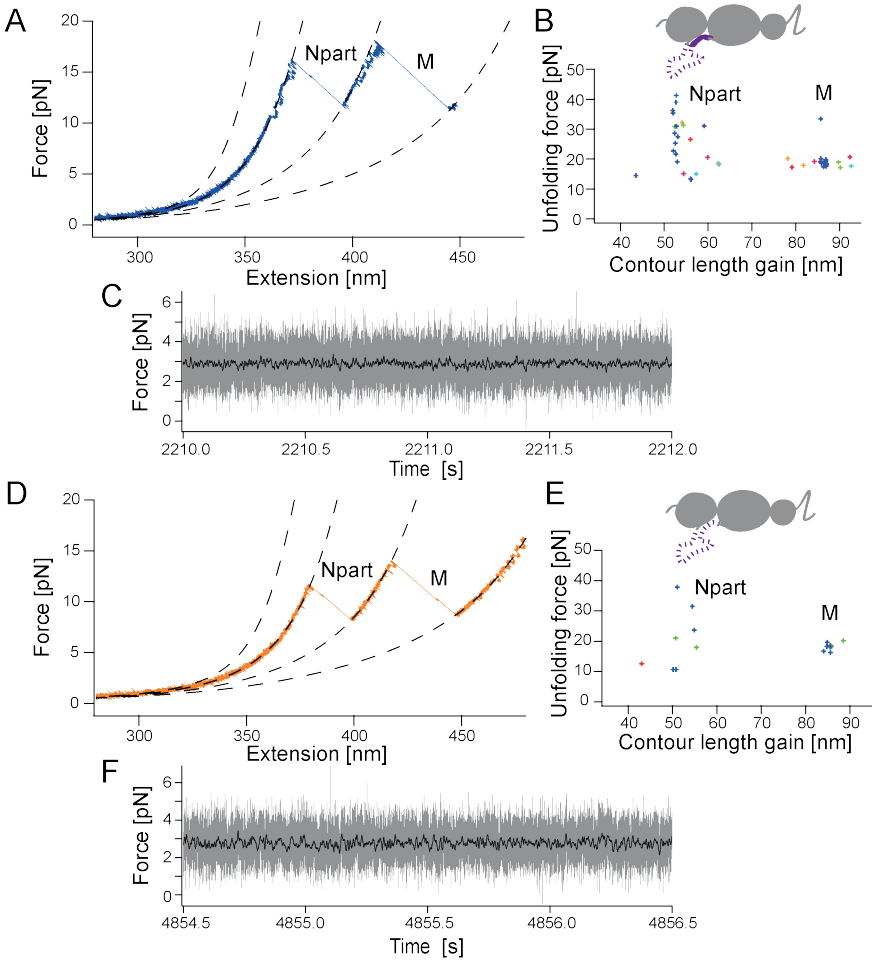

Figure C.15: (A) Example constant velocity unfolding trace at a speed of 10 nm/s of the construct Sub211-263 with the partially replaced Charged Linker. (B) A scatterplot of the domain unfolding events for different molecules (colors) shows similar stabilities and length gains as the wild type (compare to Figure 7.3). (C) A constant distance trajectory of the natively-folded construct at low forces shows no Charged Linker transitions. (D–F) Same as Figs. A–C for the full replacement construct Sub211-272.

Figure C.16 shows the full length constant velocity traces of the domain deletion constructs ΔN and ΔMC, whose Charged Linker regions are shown in Figure 9.3. Furthermore constant distance traces at low force, where the Charged Linker fluctuations are observed for wild type and scatterplots characterizing the unfolding of the Hsp90 domains for these constructs are shown.

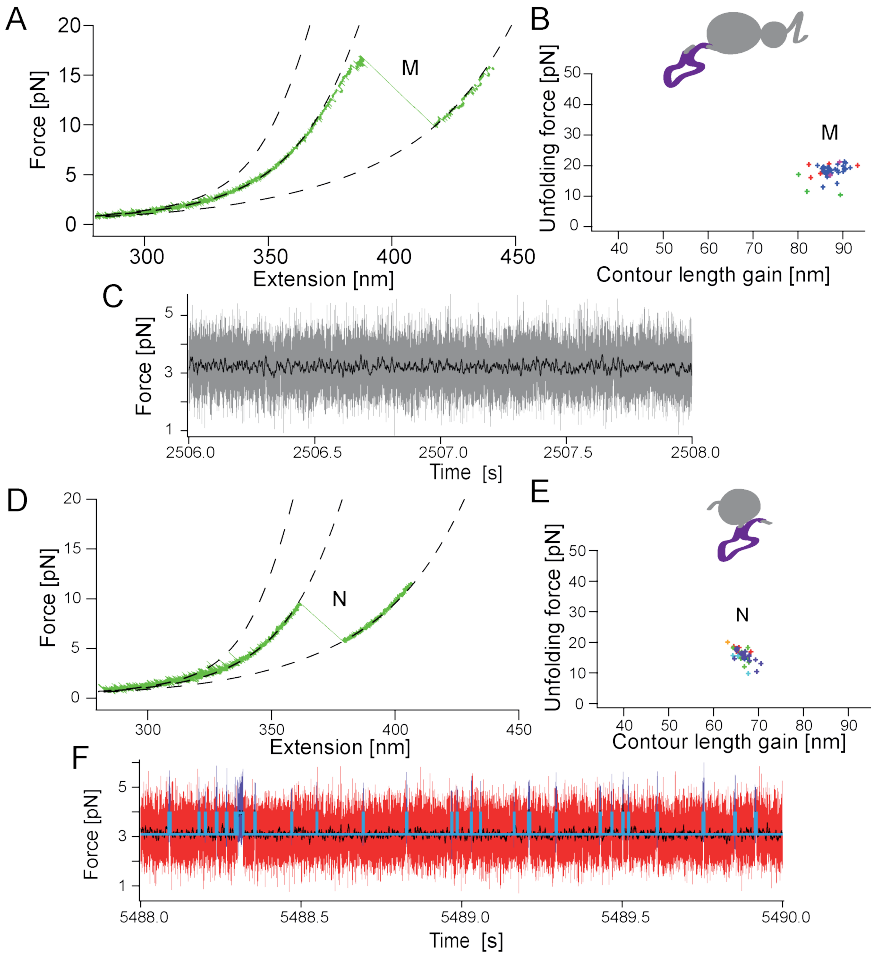

Figure C.16: (A) Example constant velocity unfolding trace at a speed of 10 nm/s of the construct without the N domain ΔN. (B) A scatterplot of the domain unfolding events for different molecules (colors) shows comparable stabilities and length gains as similar constructs (compare to Figure 7.4). (C) A constant distance trace of the natively-folded construct at low forces show no Charged Linker transitions. (D–F) Same as Figs. A–C for the construct lacking the M and C domains, ΔMC. Here the docked state of the Charged Linker fluctuations is observed in constant velocity and constant distance traces (red undocked, blue docked).

Figure C.17 summarizes the thermodynamic and kinetic properties of the Charged Linker obtained by constant distance experiments (see Section 9.3) in an energy landscape picture. The height of the energy barrier is estimated using Equation 6.12 with an attempt frequency of $10^8\,s^{-1}$.

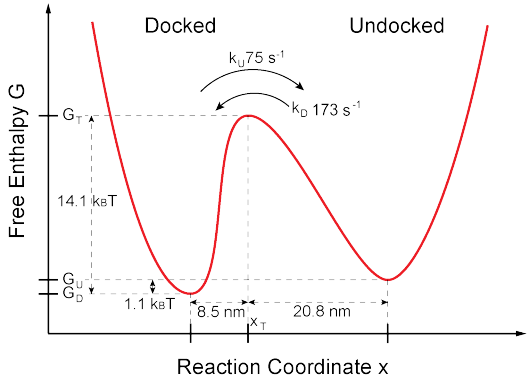

Figure C.17: The energy landscape of the Charged Linker summarizes the measured free energy and rates between the docked and the undocked states along the reaction coordinate x. The height of the transition state is estimated using the Arrhenius equation.

C.5 ADDITIONAL DIMER DATA

Figure C.18 shows additional force-extension traces from different experiments of the yeast Hsp90 dimer pulled at both N domains, as the one shown in Figure 10.1.

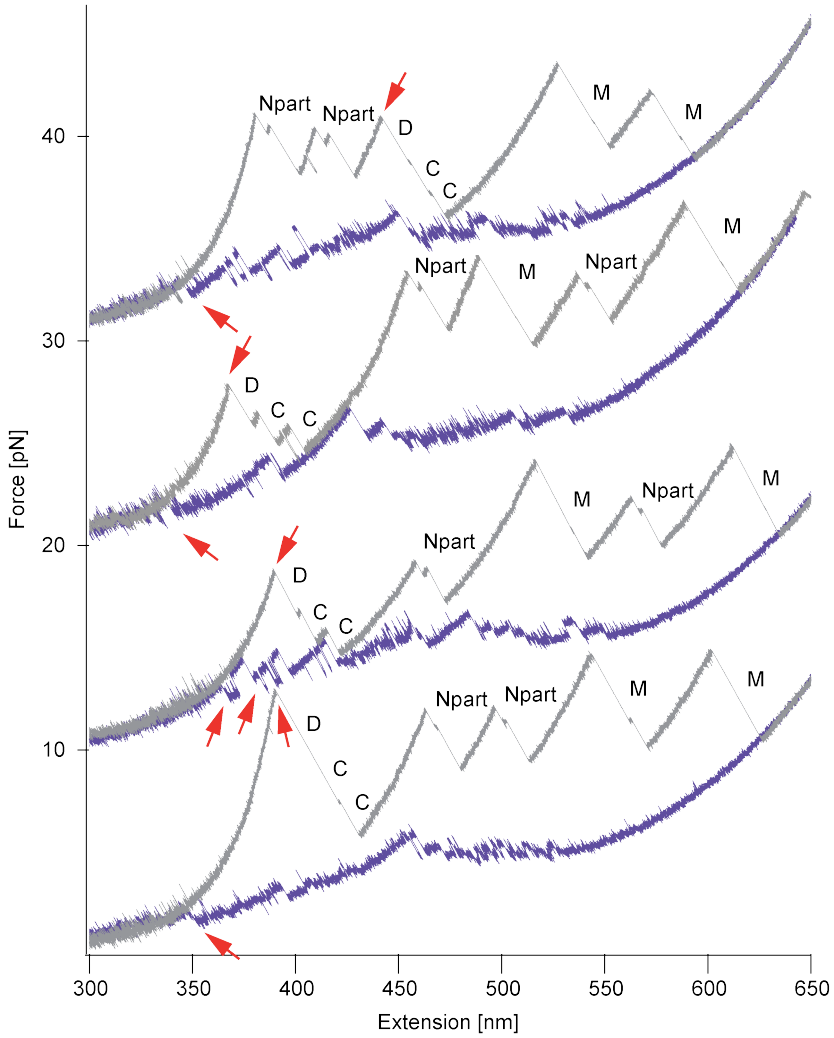

Figure C.18: Offsetted force-extension traces of the dimeric construct Hsp90zip_61C_715C pulled (gray) and relaxed (purple) with a velocity of 10 nm/s. All traces are from different molecules and association and dissociation events are marked by arrows. In the first trace dissociation of the dimer occurs after unfolding of the N domains, ruling out contributions of the N domains to C-terminal dimerization.

Detailed analysis of the dissociation forces of the dimer suggests two dimerized states with different stability (see Section 10.4). The slowly-pulled dimer shown in Figure C.19 supports this notion.

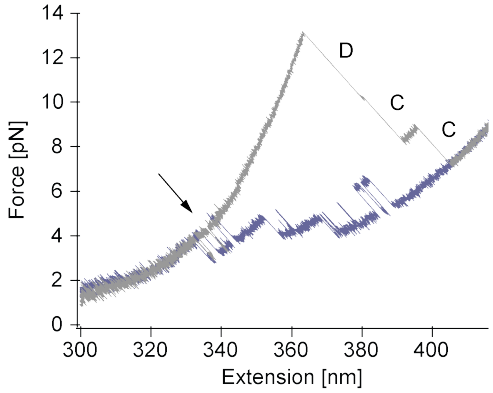

Figure C.19: Force-extension trace of the dimeric construct Hsp90zip-_452C_715C at a speed of 10 nm/s. The dimer dissociates and associates rapidly (arrow) in unfolding (gray) and refolding (purple) traces. While one dissociation transition is probable, two successive events as in the unfolding trace and the refolding trace are very unlikely. This suggests that different dimerized state exists.

D

HTPG, THE E. COLI HSP90

Here selected optical tweezers experiments with *E. coli* Hsp90 (HtpG) are presented. The structural properties of the HtpG monomer are compared to those of yeast Hsp90. Furthermore, initial experiments with HtpG dimers and experiments suggesting interactions between folded HtpG and unfolded HtpG are shown.

HtpG is short for High temperature protein G.

D.1 STRUCTURE OF HTPG

Figure D.1 shows a crystal structure obtained for HtpG [133] and example force-extension traces of a monomeric HtpG construct flanked by ubiquitins pulled at two different velocities. For details on the construct see Section A.3.

Similarly to the yeast monomer, the three domains of HtpG are identified in force-extension traces and some unstructured parts have to be assumed to fit the trace correctly (see Section 6.1.1). Before the N domains unfold they populate multiple intermediates. One very prominent intermediate at approximately half the contour length has significant stability. The heterogeneous unfolding pattern of the N domain is shown for different force-extension traces in Figure D.2 A. The M domain unfolds reproducibly via three intermediates, while the C domain doesn't show any. To verify the domain assignment a construct without the C domain is investigated. In accordance with the assignment, it doesn't show the unfolding event of the C domain (see Figure D.2 B).

To what extent incomplete refolding of the N domain is an issue hasn't be examined yet.

Figure D.1 D shows a scatterplot of unfolding forces versus contour length gains of the three domains. Average values thereof are listed in Table D.1. It is important to note that, because of the preflipping, the contour length gains obtained for the N domain depend on the fitting region. Fitting regions are chosen at the lowest forces possible, resulting in the longest contour length gains possible.

This may lead to an overestimation if some of the events observed at low forces are related to structures, which are not part of the N domain.

DOMAIN	CONTOUR LENGTH	UNFOLDING FORCE
N	$77.8\,\text{nm} \pm 2.6\,\text{nm}$	$11.9\,\text{pN} \pm 1.3\,\text{pN}$
M	$81.4\,\text{nm} \pm 1.1\,\text{nm}$	$21.0\,\text{pN} \pm 1.9\,\text{pN}$
C	$30.3\,\text{nm} \pm 1.3\,\text{nm}$	$20.8\,\text{pN} \pm 4.1\,\text{pN}$

Table D.1: Average contour length gains and unfolding forces of HtpG-diUbi obtained from 52 unfolding traces of 4 different molecules at a pulling speed of 500 nm/s. Errors are SD.

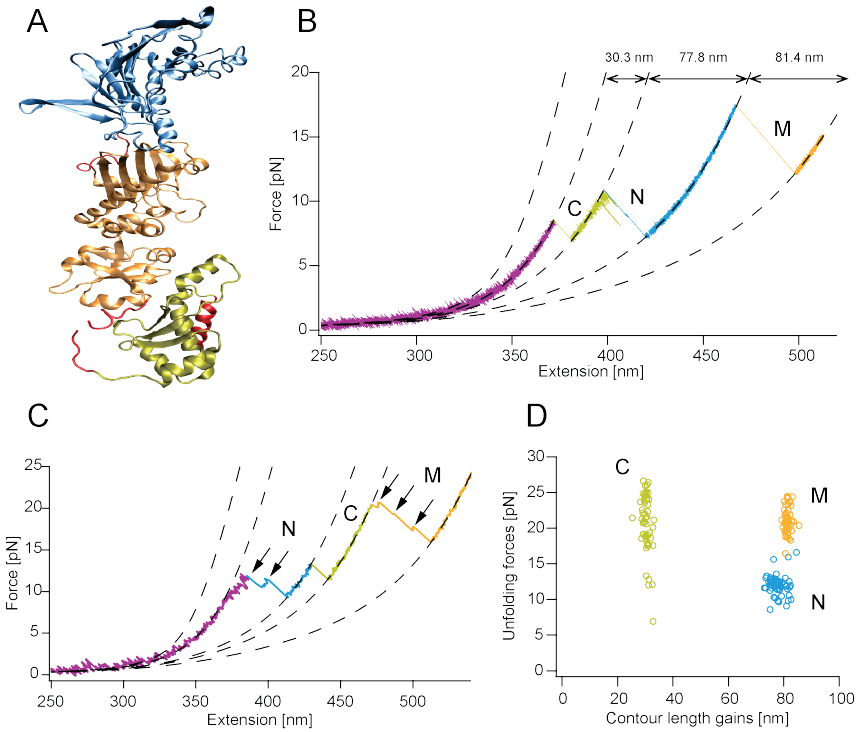

Figure D.1: (A) A monomer of the HtpG crystal structure is shown (pdb id: 2iop). Blue, orange and green colored regions correspond to the folded parts of the N, M and C domains determined by optical tweezers. The contour length gains obtained suggest that the regions shown in red are unstructured (see Table D.2). (B) An example unfolding force-extension trace of HtpG-diUbi at a slow pulling velocity of 10 nm/s shows three main peaks, which correspond to the unfolding of the three domains. Unstructured regions are taken into account assuming an additional contour length of 30 nm while fitting the WLCs (dashed lines). (C) Same experiment as in Fig. B with a pulling speed of 500 nm/s. At this speed the different unfolding intermediates of the N and M domain are clearly visible (black arrows). (D) A scatterplot shows the unfolding events of 52 unfolding traces from 4 independent experiments at a speed of 500 nm/s. The three populations correspond to the three domains.

The unfolding events are mapped onto the crystal structure, as described in Section 6.1.2 and the measured contour length gains are compared to those expected from the compact domains (see Table D.2). The contour length gains show good agreement and suggest unstructured regions between the domains. To match the contour length gain of the C domain, the last alpha helix is assumed to be unstable in the monomeric form, as it is in yeast (see Section 7.4). The folded and unstructured parts are colored in the crystal structure of Figure D.1 A.

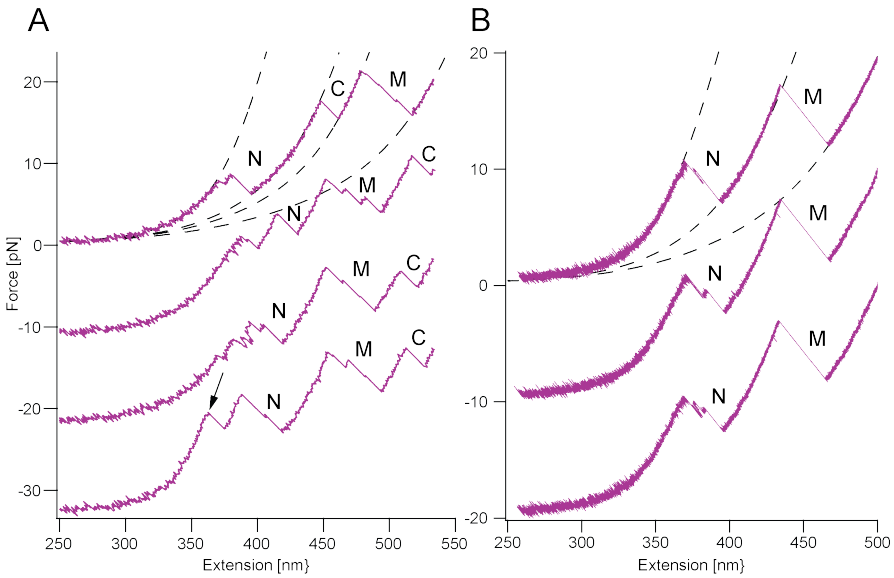

Figure D.2: (A) Offsetted, unfolding force-extension traces of HtpG-diUbi at a pulling speed of 500 nm/s show the different unfolding patterns of the N domain. Some traces also show an additional peak, like the bottommost trace, which might be a compact native state. (B) Offsetted, unfolding force-extension traces of HtpGΔC-diUbi at a pulling speed of 10 nm/s show the absence of the C domain. This construct contains less unstructured regions compared to the full length protein.

D.1.1 *Comparison of yeast and E. coli Hsp90*

As explained in the introduction (see Chapter 2) yeast and *E. coli* Hsp90s share many similarities but also show significant differences in biochemical experiments. Structurewise the major differences are the absence of the extended linker region between the N and M domains and the absence of the unstructured aminoacids at the C terminus. The crystal structures of both organisms show that the overall domain architecture is conserved [2, 133] and the secondary structure alignment shows that they are 40% homologous [41].

DOMAIN	DISTANCE IN CS	EXPECTED	MEASURED
N (1-228)	3.97 nm	78.85 nm	77.8 nm ± 2.6 nm
M (242-485)	5.36 nm	82.97 nm	81.4 nm ± 1.1 nm
C (509-606)	3.04 nm	32.00 nm	30.3 nm ± 1.3 nm

Table D.2: The theoretically expected contour length gains of the different HtpG domains are calculated from the number of aminoacids folded (regions in brackets) and the initial distances of the crystal structure (pdb id: 2iop). They are compared to the contour length gains obtained from experiments (see Table D.1).

In mechanical optical tweezers experiments the three domains of HtpG are identified. Apart from the mentioned structural differences the pattern is very similar to that of yeast. Nevertheless there are many subtle details that differ. The N domain shows less stability and slightly longer contour length, which is likely due to a continuation of the beta sheet at the C terminus (see Figure D.1 A). In yeast this region is already part of the unstable Charged Linker (see Chapter 9), which suggests that the additional flexibility provided by the Charged Linker in yeast has evolved. Furthermore, preflipping of the N domain seems to involve even more states than in yeast. The M domain is very similar and always unfolds via three intermediates, which is one more than in yeast. The first and shortest one is likely the N-terminal alpha helix, while the other two correspond to the subdomains described in the introduction. The C domain shows no unfolding intermediates and has a shorter contour length, as expected from the crystal structure. How and if these mechanical differences are related to the differences in biological function is as yet unclear.

Since the subdomains of the M domain have similar contour length gains they are difficult to distinguish, but they have a preferential unfolding sequence.

It is found that the HtpG monomer also contains unstructured regions, since the sum of contour length gains of the domains are smaller than expected. Candidates for the unstructured regions are the linkers between the domains as shown in Figure D.1 A. The linker between the M and C domains seems to comprise tens of unstructured aminoacids, which is somewhat surprising because HtpG is thought to be a more rigid protein than its eukaryotic counterpart [63].

As in yeast, nucleotides (ATP, ADP, AMP-PNP) don't show any measurable stabilization of the N domain compared to the nucleotide free (apo) state (data not shown).

D.2 HTPG DIMER EXPERIMENTS

Figure D.3 shows initial experiments with dimeric *E.coli* Hsp90 constructs pulled at both N domains as described in Section A.3. Like in yeast each domain is unfolded twice and a dimer dissociation event at high force is observed.

These traces were measured as part of the master's thesis of Maximilian Steinmaßl.

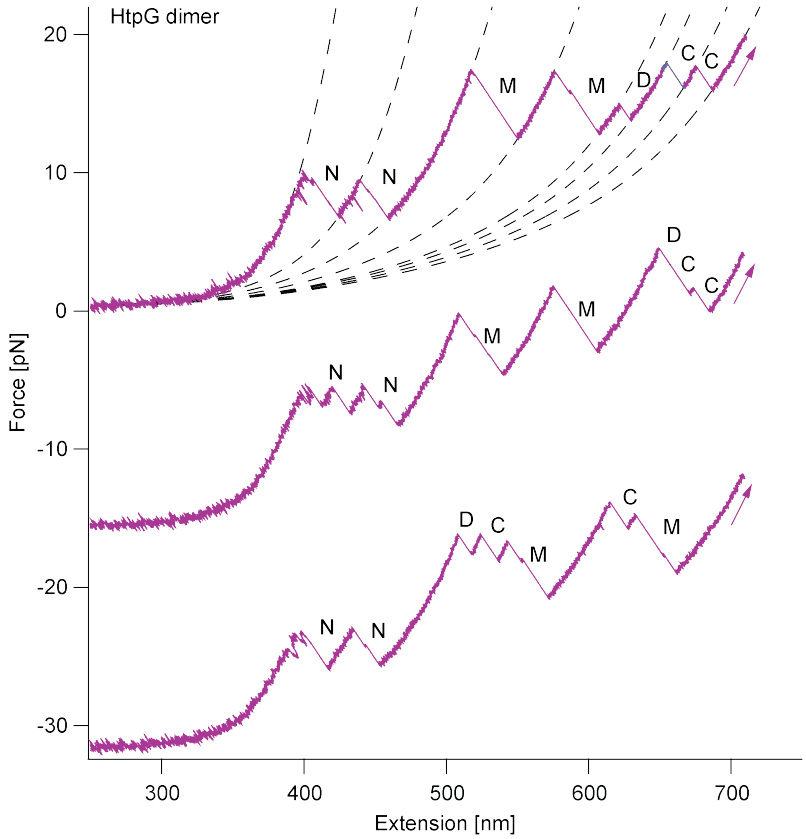

Figure D.3: Offsetted unfolding force-extension traces of dimeric HtpG_3C_634C pulled at a speed of 100 nm/s show the individual domains observed for the monomer twice (N, M and C; see also Section D.1). An additional event marks the dissociation of the C domains (D).

D.3 CHAPERONING HTPG WITH HTPG

Figure D.4 compares unfolding and refolding traces of HtpG monomers (see also Figure D.1) in the absence and presence of HtpG wild type in solution. Interestingly, if HtpG is added, the refolding traces seem to undergo additional transitions, which is best seen if the peptide is completely unfolded. In addition, the unfolding traces don't show the native pattern and seem to contain more transitions as normal, indicating that refolding in the previous cycle hasn't been accomplished, due to the presence of HtpG.

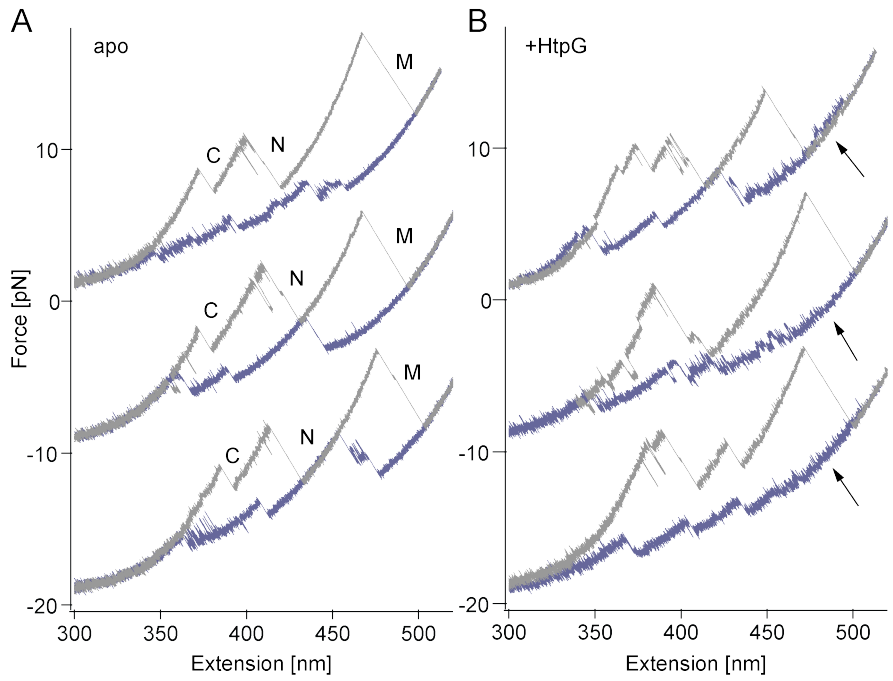

Figure D.4: Offsetted force-extension unfolding (gray) and refolding (purple) traces of the HtpG-diUbi construct without (A) and with HtpG wild type in solution (B). The refolding traces in the presence of HtpG already show transitions at high forces (arrows), indicating that the folded HtpG in solution interacts with the unfolded HtpG in the trap. Also unfolding traces have more features, suggesting incompletely refolded HtpG. All traces are from different experiments.

BIBLIOGRAPHY

[1] A. T. Alexandrescu and D. Shortle. Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *J Mol Biol*, 242(4):527–46, 1994. (Cited on page 124.)

[2] M. M. Ali, S. M. Roe, C. K. Vaughan, P. Meyer, B. Panaretou, P. W. Piper, C. Prodromou, and L. H. Pearl. Crystal structure of an hsp90-nucleotide-p23/sba1 closed chaperone complex. *Nature*, 440(7087):1013–7, 2006. (Cited on pages 7, 10, 17, 51, 87, 88, 99, 102, 112, 114, 115, 144, and 157.)

[3] S. Arrhenius. Zur theorie der chemischen reaktionsgeschwindigkeit. *Z. Phys. Chem.*, 4:S. 226–248, 1889. (Cited on page 34.)

[4] A. Ashkin. Acceleration and trapping of particles by radiation pressure. *Physical Review Letters*, 24(4):156, 1970. (Cited on page 21.)

[5] A. Ashkin and J. M. Dziedzic. Optical trapping and manipulation of viruses and bacteria. *Science*, 235(4795):1517–20, 1987. (Cited on page 21.)

[6] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt Lett*, 11(5):288, 1986. (Cited on page 21.)

[7] A. Ashkin, J. M. Dziedzic, and T. Yamane. Optical trapping and manipulation of single cells using infrared laser beams. *Nature*, 330(6150):769–71, 1987. (Cited on page 21.)

[8] R. L. Baldwin. On-pathway versus off-pathway folding intermediates. *Fold Des*, 1(1):R1–8, 1996. (Cited on pages 59 and 79.)

[9] U. Bastolla and L. Demetrius. Stability constraints and protein evolution: the role of chain length, composition and disulfide bonds. *Protein Eng Des Sel*, 18(9):405–15, 2005. (Cited on pages 81 and 83.)

[10] S. Batey and J. Clarke. Apparent cooperativity in the folding of multidomain proteins depends on the relative rates of folding of the constituent domains. *Proc Natl Acad Sci U S A*, 103(48):18113–8, 2006. (Cited on page 59.)

[11] G. I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–27, 1978. (Cited on page 36.)

[12] M. B. Borgia, A. Borgia, R. B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, and J. Clarke. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature*, 474(7353):662–5, 2011. (Cited on page 82.)

[13] K. A. Borkovich, F. W. Farrelly, D. B. Finkelstein, J. Taulien, and S. Lindquist. hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol*, 9(9):3919–30, 1989. (Cited on pages 7 and 121.)

[14] T. Bornschlogl, G. Woehlke, and M. Rief. Single molecule mechanics of the kinesin neck. *Proc Natl Acad Sci U S A*, 106(17):6992–7, 2009. (Cited on pages 105 and 122.)

[15] D. J. Brockwell and S. E. Radford. Intermediates: ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol*, 17(1):30–7, 2007. (Cited on pages 55 and 78.)

[16] M. Brunnbauer, F. Mueller-Planitz, S. Kosem, T. H. Ho, R. Dombi, J. C. Gebhardt, M. Rief, and Z. Okten. Regulation of a heterodimeric kinesin-2 through an unprocessive motor domain that is turned processive by its partner. *Proc Natl Acad Sci U S A*, 107(23):10460–5, 2010. (Cited on page 21.)

[17] C. Cecconi, E. A. Shank, C. Bustamante, and S. Marqusee. Direct observation of the three-state folding of a single protein molecule. *Science*, 309(5743):2057–2060, 2005. (Cited on page 21.)

[18] C. Cecconi, E. A. Shank, F. W. Dahlquist, S. Marqusee, and C. Bustamante. Protein-dna chimeras for single molecule mechanical folding studies with the optical tweezers. *European Biophysics Journal with Biophysics Letters*, 37(6):729–738, 2008. (Cited on pages 23 and 128.)

[19] S. Cheng, M. Cetinkaya, and F. Grater. How sequence determines elasticity of disordered proteins. *Biophys J*, 99(12):3863–9, 2010. (Cited on page 28.)

[20] H. S. Chung, J. M. Louis, and W. A. Eaton. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc Natl Acad Sci U S A*, 106(29):11837–44, 2009. (Cited on page 55.)

[21] H. S. Chung, S. Piana-Agostinetti, D. E. Shaw, and W. A. Eaton. Structural origin of slow diffusion in protein folding. *Science*, 349(6255):1504–10, 2015. (Cited on page 35.)

[22] C. N. Cunningham, D. R. Southworth, K. A. Krukenberg, and D. A. Agard. The conserved arginine 380 of hsp90 is not a catalytic residue, but stabilizes the closed conformation required for atp hydrolysis. *Protein Sci*, 21(8):1162–71, 2012. (Cited on page 9.)

[23] V. Daggett and A. Fersht. The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol*, 4(6):497–502, 2003. (Cited on page 78.)

[24] P. De Los Rios, A. Ben-Zvi, O. Slutsky, A. Azem, and P. Goloubinoff. Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proc Natl Acad Sci U S A*, 103(16):6166–71, 2006. (Cited on page 85.)

[25] A. A. Deniz, S. Mukhopadhyay, and E. A. Lemke. Single-molecule biophysics: at the interface of biology, physics and chemistry. *J R Soc Interface*, 5(18):15–45, 2008. (Cited on page 16.)

[26] H. Dietz, F. Berkemeier, M. Bertz, and M. Rief. Anisotropic deformation response of single protein molecules. *Proc Natl Acad Sci U S A*, 103(34):12724–8, 2006. (Cited on page 54.)

[27] K. A. Dill and J. L. MacCallum. The protein-folding problem, 50 years on. *Science*, 338(6110):1042–6, 2012. (Cited on page 5.)

[28] D. E. Dollins, J. J. Warren, R. M. Immormino, and D. T. Gewirth. Structures of grp94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. *Mol Cell*, 28(1):41–56, 2007. (Cited on pages 7 and 9.)

[29] O. K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys Rev Lett*, 96(10):108101, 2006. (Cited on page 38.)

[30] O. K. Dudko, G. Hummer, and A. Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc Natl Acad Sci U S A*, 105(41):15755–15760, 2008. (Cited on pages 38 and 41.)

[31] R. Dutta and M. Inouye. Ghkl, an emergent atpase/kinase superfamily. *Trends Biochem Sci*, 25(1):24–8, 2000. (Cited on page 7.)

[32] I. O. Ebong, N. Morgner, M. Zhou, M. A. Saraiva, S. Daturpalli, S. E. Jackson, and C. V. Robinson. Heterogeneity and dynamics in the assembly of the heat shock protein 90 chaperone complexes. *Proc Natl Acad Sci U S A*, 108(44):17939–44, 2011. (Cited on page 10.)

[33] P. J. Elms, J. D. Chodera, C. Bustamante, and S. Marqusee. The molten globule state is unusually deformable under mechanical force. *Proc Natl Acad Sci U S A*, 109(10):3796–801, 2012. (Cited on page 37.)

[34] S. Enoki, K. Saeki, K. Maki, and K. Kuwajima. Acid denaturation and refolding of green fluorescent protein. *Biochemistry*, 43(44):14238–48, 2004. (Cited on page 55.)

[35] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys J*, 72(4):1541–55, 1997. (Cited on page 36.)

[36] A. N. Fedorov and T. O. Baldwin. Cotranslational protein folding. *J Biol Chem*, 272(52):32715–8, 1997. (Cited on page 84.)

[37] J. C. M. Gebhardt, T. Bornschlogla, and M. Rief. Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci U S A*, 107(5):2013–2018, 2010. (Cited on page 38.)

[38] O. Genest, J. R. Hoskins, J. L. Camberg, S. M. Doyle, and S. Wickner. Heat shock protein 90 from escherichia coli collaborates with the dnak chaperone system in client protein remodeling. *Proc Natl Acad Sci U S A*, 108(20):8206–11, 2011. (Cited on page 12.)

[39] O. Genest, M. Reidy, T. O. Street, J. R. Hoskins, J. L. Camberg, D. A. Agard, D. C. Masison, and S. Wickner. Uncovering a region of heat shock protein 90 important for client binding in e. coli and chaperone function in yeast. *Mol Cell*, 49(3):464–73, 2013. (Cited on page 13.)

[40] D. H. Goldman, C. M. Kaiser, A. Milin, M. Righini, J. Tinoco, I., and C. Bustamante. Ribosome. mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science*, 348(6233):457–60, 2015. (Cited on page 84.)

[41] C. Graf, C. T. Lee, L. Eva Meier-Andrejszki, M. T. Nguyen, and M. P. Mayer. Differences in conformational dynamics within the hsp90 chaperone family reveal mechanistic insights. *Front Mol Biosci*, 1:4, 2014. (Cited on pages 114, 115, and 157.)

[42] C. Graf, M. Stankiewicz, G. Kramer, and M. P. Mayer. Spatially and kinetically resolved changes in the conformational dynamics of the hsp90 chaperone machine. *EMBO J*, 28(5):602–13, 2009. (Cited on pages 10 and 115.)

[43] O. Hainzl, M. C. Lapina, J. Buchner, and K. Richter. The charged linker region is an important regulator of hsp90 function. *Journal of Biological Chemistry*, 284(34):22559–22567, 2009. (Cited on page 87.)

[44] J. H. Han, S. Batey, A. A. Nickson, S. A. Teichmann, and J. Clarke. The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol*, 8(4):319–30, 2007. (Cited on pages 54 and 82.)

[45] P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after kramers. *Reviews of Modern Physics*, 62(2):251–341, 1990. (Cited on page 35.)

[46] S. F. Harris, A. K. Shiau, and D. A. Agard. The crystal structure of the carboxy-terminal dimerization domain of htpg, the escherichia coli hsp90, reveals a potential substrate binding site. *Structure*, 12(6):1087–97, 2004. (Cited on pages 9, 13, and 114.)

[47] P. O. Heidarsson, M. M. Naqvi, M. R. Otazo, A. Mossa, B. B. Kragelund, and C. Cecconi. Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1. *Proc Natl Acad Sci U S A*, 111(36):13069–74, 2014. (Cited on page 79.)

[48] M. Hessling, K. Richter, and J. Buchner. Dissection of the atp-induced conformational cycle of the molecular chaperone hsp90. *Nat Struct Mol Biol*, 16(3):287–93, 2009. (Cited on pages 94, 96, 99, 102, and 115.)

[49] W. Hu, B. T. Walters, Z. Y. Kan, L. Mayne, L. E. Rosen, S. Marqusee, and S. W. Englander. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci U S A*, 110(19):7684–9, 2013. (Cited on page 55.)

[50] C. Hyeon, G. Morrison, D. L. Pincus, and D. Thirumalai. Refolding dynamics of stretched biopolymers upon force quench. *Proc Natl Acad Sci U S A*, 106(48):20288–93, 2009. (Cited on page 82.)

[51] C. International Human Genome Sequencing. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–45, 2004. (Cited on page 3.)

[52] D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, and A. V. Finkelstein. Contact order revisited: influence of protein size on the folding rate. *Protein Sci*, 12(9):2057–62, 2003. (Cited on page 55.)

[53] M. Jahn, J. Buchner, T. Hugel, and M. Rief. Folding and assembly of the large molecular machine hsp90 studied in single-molecule experiments. *Proc Natl Acad Sci U S A*, 113(5):1232–7, 2016. (Cited on page 55.)

[54] M. Jahn, A. Rehn, B. Pelz, B. Hellenkamp, K. Richter, M. Rief, J. Buchner, and T. Hugel. The charged linker of the molecular chaperone hsp90 modulates domain contacts and biological function. *Proc Natl Acad Sci U S A*, 111(50):17881–6, 2014. (Cited on pages 87 and 127.)

[55] U. Jakob, H. Lilie, I. Meyer, and J. Buchner. Transient interaction of hsp90 with early unfolding intermediates of citrate synthase. implications for heat shock in vivo. *J Biol Chem*, 270(13):7288–94, 1995. (Cited on page 55.)

[56] U. Jakob, I. Meyer, H. Bugl, S. Andre, J. C. Bardwell, and J. Buchner. Structural organization of procaryotic and eucaryotic hsp90. influence of divalent cations on structure and function. *J Biol Chem*, 270(24):14412–9, 1995. (Cited on page 7.)

[57] A. N. Kapanidis, T. A. Laurence, N. K. Lee, E. Margeat, X. Kong, and S. Weiss. Alternating-laser excitation of single molecules. *Acc Chem Res*, 38(7):523–33, 2005. (Cited on page 132.)

[58] G. E. Karagoz, A. M. Duarte, E. Akoury, H. Ippel, J. Biernat, T. Moran Luengo, M. Radli, T. Didenko, B. A. Nordhues, D. B. Veprintsev, C. A. Dickey, E. Mandelkow, M. Zweckstetter, R. Boelens, T. Madl, and S. G. Rudiger. Hsp90-tau complex reveals molecular basis for specificity in chaperone action. *Cell*, 156(5):963–74, 2014. (Cited on page 12.)

[59] T. V. Karpinets, D. J. Greenwood, C. E. Sams, and J. T. Ammons. Rna:protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biol*, 4:30, 2006. (Cited on page 84.)

[60] S. Khorasanizadeh, I. D. Peters, and H. Roder. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol*, 3(2):193–205, 1996. (Cited on page 59.)

[61] E. Kirschke, D. Goswami, D. Southworth, P. R. Griffin, and D. A. Agard. Glucocorticoid receptor function regulated by coordinated action of the hsp90 and hsp70 chaperone cycles. *Cell*, 157(7):1685–97, 2014. (Cited on pages 13 and 115.)

[62] H. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940. (Cited on page 35.)

[63] K. A. Krukenberg, F. Forster, L. M. Rice, A. Sali, and D. A. Agard. Multiple conformations of e. coli hsp90 in solution: insights into the conformational dynamics of hsp90. *Structure*, 16(5):755–65, 2008. (Cited on page 158.)

[64] S. C. Kuo and M. P. Sheetz. Force of single kinesin molecules measured with optical tweezers. *Science*, 260(5105):232–4, 1993. (Cited on page 21.)

[65] K. J. Laidler and M. C. King. The development of transition-state theory. *Journal of Physical Chemistry*, 87(15):2657–2664, 1983. (Cited on page 34.)

[66] T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr Opin Struct Biol*, 23(1):58–65, 2013. (Cited on pages 4 and 78.)

[67] L. J. Lapidus, W. A. Eaton, and J. Hofrichter. Measuring the rate of intramolecular contact formation in polypeptides. *Proc Natl Acad Sci U S A*, 97(13):7220–5, 2000. (Cited on page 35.)

[68] L. A. Lavery, J. R. Partridge, T. A. Ramelot, D. Elnatan, M. A. Kennedy, and D. A. Agard. Structural asymmetry in the closed state of mitochondrial hsp90 (trap1) supports a two-step atp hydrolysis mechanism. *Mol Cell*, 53(2):330–43, 2014. (Cited on page 10.)

[69] M. C. Leake, D. Wilson, M. Gautel, and R. M. Simmons. The elasticity of single titin molecules using a two-bead optical tweezers assay. *Biophys J*, 87(2):1112–35, 2004. (Cited on page 28.)

[70] N. K. Lee, A. N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright, and S. Weiss. Accurate fret measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys J*, 88(4):2939–53, 2005. (Cited on page 16.)

[71] J. Li, K. Richter, and J. Buchner. Mixed hsp90-cochaperone complexes are important for the progression of the reaction cycle. *Nat Struct Mol Biol*, 18(1):61–6, 2011. (Cited on pages 10 and 99.)

[72] J. Li, K. Richter, J. Reinstein, and J. Buchner. Integration of the accelerator aha1 in the hsp90 co-chaperone cycle. *Nat Struct Mol Biol*, 20(3):326–31, 2013. (Cited on pages 10, 11, 102, and 133.)

[73] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–20, 2011. (Cited on page 4.)

[74] O. R. Lorenz, L. Freiburger, D. A. Rutz, M. Krause, B. K. Zierer, S. Alvira, J. Cuellar, J. M. Valpuesta, T. Madl, M. Sattler, and J. Buchner. Modulation of the hsp90 chaperone cycle by a stringent client protein. *Mol Cell*, 53(6):941–53, 2014. (Cited on page 12.)

[75] J. F. Louvion, R. Warth, and D. Picard. Two eukaryote-specific regions of hsp82 are dispensable for its viability and signal transduction functions in yeast. *Proc Natl Acad Sci U S A*, 93(24):13937–42, 1996. (Cited on page 87.)

[76] W. Luo, W. Sun, T. Taldone, A. Rodina, and G. Chiosis. Heat shock protein 90 in neurodegenerative diseases. *Mol Neurodegener*, 5:24, 2010. (Cited on page 13.)

[77] R. A. Maillard, G. Chistol, M. Sen, M. Righini, J. Tan, C. M. Kaiser, C. Hodges, A. Martin, and C. Bustamante. Clpx(p) generates mechanical force to unfold and translocate its protein substrates. *Cell*, 145(3):459–69, 2011. (Cited on page 85.)

[78] J. F. Marko and E. D. Siggia. Statistical mechanics of supercoiled dna. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, 52(3):2912–2938, 1995. (Cited on page 27.)

[79] J. F. Marko and E. D. Siggia. Stretching dna. *Macromolecules*, 28(26):8759–8770, 1995. (Cited on page 27.)

[80] A. Mashaghi, G. Kramer, P. Bechtluft, B. Zachmann-Brand, A. J. Driessen, B. Bukau, and S. J. Tans. Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature*, 500(7460):98–101, 2013. (Cited on page 85.)

[81] M. P. Mayer. Gymnastics of molecular chaperones. *Mol Cell*, 39(3):321–31, 2010. (Cited on page 7.)

[82] M. P. Mayer and L. Le Breton. Hsp90: breaking the symmetry. *Mol Cell*, 58(1):8–20, 2015. (Cited on page 12.)

[83] C. Mayr, K. Richter, H. Lilie, and J. Buchner. Cpr6 and cpr7, two closely related hsp90-associated immunophilins from saccharomyces cerevisiae, differ in their functional properties. *J Biol Chem*, 275(44):34140–6, 2000. (Cited on page 114.)

[84] S. H. McLaughlin, H. W. Smith, and S. E. Jackson. Stimulation of the weak atpase activity of human hsp90 by a client protein. *J Mol Biol*, 315(4):787–98, 2002. (Cited on page 115.)

[85] S. H. McLaughlin, L. A. Ventouras, B. Lobbezoo, and S. E. Jackson. Independent atpase activity of hsp90 subunits creates a flexible assembly platform. *J Mol Biol*, 344(3):813–26, 2004. (Cited on pages 54 and 115.)

[86] P. Meyer, C. Prodromou, C. Liao, B. Hu, S. M. Roe, C. K. Vaughan, I. Vlasic, B. Panaretou, P. W. Piper, and L. H. Pearl. Structural basis for recruitment of the atpase activator aha1 to the hsp90 chaperone machinery. *EMBO J*, 23(6):1402–10, 2004. (Cited on page 10.)

[87] M. Mickler, M. Hessling, C. Ratzke, J. Buchner, and T. Hugel. The large conformational changes of hsp90 are only weakly coupled to atp hydrolysis. *Nat Struct Mol Biol*, 16(3):281–286, 2009. (Cited on pages 10, 47, 93, 94, 95, 102, 115, 121, and 131.)

[88] L. Milanesi, J. P. Waltho, C. A. Hunter, D. J. Shaw, G. S. Beddard, G. D. Reid, S. Dev, and M. Volk. Measurement of energy landscape roughness of folded and unfolded proteins. *Proc Natl Acad Sci U S A*, 109(48):19563–8, 2012. (Cited on page 4.)

[89] P. Mishra and D. N. Bolon. Designed hsp90 heterodimers reveal an asymmetric atpase-driven mechanism in vivo. *Mol Cell*, 53(2):344–50, 2014. (Cited on page 10.)

[90] J. R. Moffitt, Y. R. Chemla, D. Izhaky, and C. Bustamante. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc Natl Acad Sci U S A*, 103(24):9006–11, 2006. (Cited on page 22.)

[91] B. K. Müller, E. Zaychikov, C. Brauchle, and D. C. Lamb. Pulsed interleaved excitation. *Biophys J*, 89(5):3508–22, 2005. (Cited on page 132.)

[92] D. F. Nathan and S. Lindquist. Mutational analysis of hsp90 function - interactions with a steroid-receptor and a protein-kinase. *Molecular and Cellular Biology*, 15(7):3917–3925, 1995. (Cited on page 7.)

[93] K. C. Neuman and A. Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods*, 5(6):491–505, 2008. (Cited on page 16.)

[94] T. A. Nieminen, V. L. Y. Loke, A. B. Stilgoe, G. Knoner, A. M. Branczyk, N. R. Heckenberg, and H. Rubinsztein-Dunlop. Optical tweezers computational toolbox. *Journal of Optics a-Pure and Applied Optics*, 9(8):S196–S203, 2007. (Cited on page 21.)

[95] L. Oberbarnscheidt, R. Janissen, and F. Oesterhelt. Direct and model free calculation of force-dependent dissociation rates from force spectroscopic data. *Biophys J*, 97(9):L19–21, 2009. (Cited on pages 41 and 108.)

[96] A. F. Oberhauser, P. E. Marszalek, M. Carrion-Vazquez, and J. M. Fernandez. Single protein misfolding events captured by atomic force microscopy. *Nat Struct Biol*, 6(11):1025–8, 1999. (Cited on page 82.)

[97] W. M. Obermann, H. Sondermann, A. A. Russo, N. P. Pavletich, and F. U. Hartl. In vivo function of hsp90 is dependent on atp binding and atp hydrolysis. *J Cell Biol*, 143(4):901–10, 1998. (Cited on pages 9 and 114.)

[98] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem*, 48:545–600, 1997. (Cited on page 83.)

[99] J. N. Onuchic and P. G. Wolynes. Theory of protein folding. *Curr Opin Struct Biol*, 14(1):70–5, 2004. (Cited on page 4.)

[100] B. Panaretou, C. Prodromou, S. M. Roe, R. O'Brien, J. E. Ladbury, P. W. Piper, and L. H. Pearl. Atp binding and hydrolysis are essential to the function of the hsp90 molecular chaperone in vivo. *Embo Journal*, 17(16):4829–4836, 1998. (Cited on pages 9, 114, and 133.)

[101] B. Panaretou, G. Siligardi, P. Meyer, A. Maloney, J. K. Sullivan, S. Singh, S. H. Millson, P. A. Clarke, S. Naaby-Hansen, R. Stein, R. Cramer, M. Mollapour, P. Workman, P. W. Piper, L. H. Pearl, and C. Prodromou. Activation of the atpase activity of hsp90 by the stress-regulated cochaperone aha1. *Molecular Cell*, 10(6):1307–1318, 2002. (Cited on page 99.)

[102] L. H. Pearl, C. Prodromou, and P. Workman. The hsp90 molecular chaperone: an open and shut case for treatment. *Biochem J*, 410(3):439–53, 2008. (Cited on page 13.)

[103] Q. Peng, J. Fang, M. Wang, and H. Li. Kinetic partitioning mechanism governs the folding of the third fniii domain of tenascin-c: evidence at the single-molecule level. *J Mol Biol*, 412(4):698–709, 2011. (Cited on page 82.)

[104] F. Persson, I. Barkefors, and J. Elf. Single molecule methods with applications in living cells. *Curr Opin Biotechnol*, 24(4):737–44, 2013. (Cited on page 16.)

[105] M. Pirchi, G. Ziv, I. Riven, S. S. Cohen, N. Zohar, Y. Barak, and G. Haran. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat Commun*, 2:493, 2011. (Cited on page 35.)

[106] K. W. Plaxco and D. Baker. Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proc Natl Acad Sci U S A*, 95(23):13591–6, 1998. (Cited on pages 35, 80, and 81.)

[107] I. Popa, J. M. Fernandez, and S. Garcia-Manyes. Direct quantification of the attempt frequency determining the mechanical unfolding of ubiquitin protein. *J Biol Chem*, 286(36):31072–9, 2011. (Cited on page 35.)

[108] W. B. Pratt and D. O. Toft. Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone ma-

chinery. *Exp Biol Med (Maywood)*, 228(2):111–33, 2003. (Cited on page 7.)

[109] C. Prodromou, B. Panaretou, S. Chohan, G. Siligardi, R. O'Brien, J. E. Ladbury, S. M. Roe, P. W. Piper, and L. H. Pearl. The atpase cycle of hsp90 drives a molecular 'clamp' via transient dimerization of the n-terminal domains. *EMBO J*, 19(16):4383–92, 2000. (Cited on page 114.)

[110] C. Prodromou, S. M. Roe, R. O'Brien, J. E. Ladbury, P. W. Piper, and L. H. Pearl. Identification and structural characterization of the atp/adp-binding site in the hsp90 molecular chaperone. *Cell*, 90(1):65–75, 1997. (Cited on pages 7, 54, and 115.)

[111] C. Prodromou, S. M. Roe, P. W. Piper, and L. H. Pearl. A molecular clamp in the crystal structure of the n-terminal domain of the yeast hsp90 chaperone. *Nat Struct Biol*, 4(6):477–82, 1997. (Cited on page 114.)

[112] L. Pullen and D. N. Bolon. Enforced n-domain proximity stimulates hsp90 atpase activity and is compatible with function in vivo. *J Biol Chem*, 286(13):11091–8, 2011. (Cited on pages 9, 114, and 115.)

[113] P. X. Qi, T. R. Sosnick, and S. W. Englander. The burst phase in ribonuclease a folding and solvent dependence of the unfolded state. *Nat Struct Biol*, 5(10):882–4, 1998. (Cited on page 59.)

[114] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. (Cited on page 31.)

[115] C. Ratzke, F. Berkemeier, and T. Hugel. Heat shock protein 90's mechanochemical cycle is dominated by thermal fluctuations. *Proc Natl Acad Sci U S A*, 109(1):161–6, 2012. (Cited on pages 10, 47, 54, and 115.)

[116] C. Ratzke, B. Hellenkamp, and T. Hugel. Four-colour fret reveals directionality in the hsp90 multicomponent machinery. *Nat Commun*, 5:4192, 2014. (Cited on page 132.)

[117] C. Ratzke, M. Mickler, B. Hellenkamp, J. Buchner, and T. Hugel. Dynamics of heat shock protein 90 c-terminal dimerization is an important part of its conformational cycle. *Proc Natl Acad Sci U S A*, 107(37):16101–16106, 2010. (Cited on page 114.)

[118] C. Ratzke, M. N. T. Nguyen, M. P. Mayer, and T. Hugel. From a ratchet mechanism to random fluctuations evolution of hsp90's mechanochemical cycle. *Journal of Molecular Biology*, 423(3):462–471, 2012. (Cited on pages 12 and 115.)

[119] M. Retzlaff, F. Hagn, L. Mitschke, M. Hessling, F. Gugel, H. Kessler, K. Richter, and J. Buchner. Asymmetric activation of the hsp90 dimer by its cochaperone aha1. *Molecular Cell*, 37(3):344–354, 2010. (Cited on pages 10, 99, and 102.)

[120] K. Richter, P. Muschler, O. Hainzl, and J. Buchner. Coordinated atp hydrolysis by the hsp90 dimer. *J Biol Chem*, 276(36):33689–96, 2001. (Cited on pages 9, 99, 109, 114, and 115.)

[121] K. Richter, J. Reinstein, and J. Buchner. N-terminal residues regulate the catalytic efficiency of the hsp90 atpase cycle. *J Biol Chem*, 277(47):44905–10, 2002. (Cited on pages 52 and 115.)

[122] K. Richter, J. Soroka, L. Skalniak, A. Leskovar, M. Hessling, J. Reinstein, and J. Buchner. Conserved conformational changes in the atpase cycle of human hsp90. *J Biol Chem*, 283(26):17757–65, 2008. (Cited on pages 9 and 10.)

[123] K. Richter, S. Walter, and J. Buchner. The co-chaperone sba1 connects the atpase reaction of hsp90 to the progression of the chaperone cycle. *J Mol Biol*, 342(5):1403–13, 2004. (Cited on pages 10 and 99.)

[124] L. Rognoni, T. Most, G. Zoldak, and M. Rief. Force-dependent isomerization kinetics of a highly conserved proline switch modulates the mechanosensing region of filamin. *Proc Natl Acad Sci U S A*, 111(15):5568–73, 2014. (Cited on page 109.)

[125] L. Rognoni, J. Stigler, B. Pelz, J. Ylanne, and M. Rief. Dynamic force sensing of filamin revealed in single-molecule experiments. *Proc Natl Acad Sci U S A*, 109(48):19679–19684, 2012. (Cited on pages 21, 45, and 78.)

[126] A. Rohl, J. Rohrberg, and J. Buchner. The chaperone hsp90: changing partners for demanding clients. *Trends Biochem Sci*, 38(5):253–62, 2013. (Cited on pages 10 and 12.)

[127] C. A. Ross and M. A. Poirier. Protein aggregation and neurodegenerative disease. *Nat Med*, 10 Suppl:S10–7, 2004. (Cited on page 4.)

[128] R. Roy, S. Hohng, and T. Ha. A practical guide to single-molecule fret. *Nat Methods*, 5(6):507–16, 2008. (Cited on page 16.)

[129] M. Schlierf, F. Berkemeier, and M. Rief. Direct observation of active protein folding using lock-in force spectroscopy. *Biophys J*, 93(11):3989–98, 2007. (Cited on pages 37 and 38.)

[130] I. Schwaiger, M. Schleicher, A. A. Noegel, and M. Rief. The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep*, 6(1):46–51, 2005. (Cited on page 59.)

[131] Y. Seol, J. Li, P. C. Nelson, T. T. Perkins, and M. D. Betterton. Elasticity of short dna molecules: theory and experiment for contour lengths of 0.6-7 microm. *Biophys J*, 93(12):4360–73, 2007. (Cited on page 28.)

[132] E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee, and C. Bustamante. The folding cooperativity of a protein is controlled by its chain topology. *Nature*, 465(7298):637–40, 2010. (Cited on page 55.)

[133] A. K. Shiau, S. F. Harris, D. R. Southworth, and D. A. Agard. Structural analysis of e. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell*, 127(2):329–40, 2006. (Cited on pages 7, 9, 17, 102, 114, 155, and 157.)

[134] M. Shtilerman, G. H. Lorimer, and S. W. Englander. Chaperonin function: folding by forced unfolding. *Science*, 284(5415):822–5, 1999. (Cited on page 85.)

[135] G. Siligardi, B. Hu, B. Panaretou, P. W. Piper, L. H. Pearl, and C. Prodromou. Co-chaperone regulation of conformational switching in the hsp90 atpase cycle. *Journal of Biological Chemistry*, 279(50):51989–51998, 2004. (Cited on pages 10 and 99.)

[136] D. R. Southworth and D. A. Agard. Species-dependent ensembles of conserved conformational states define the hsp90 chaperone atpase cycle. *Mol Cell*, 32(5):631–40, 2008. (Cited on pages 12 and 115.)

[137] J. Stigler and M. Rief. Calcium-dependent folding of single calmodulin molecules. *Proc Natl Acad Sci U S A*, 109(44):17814–17819, 2012. (Cited on page 21.)

[138] J. Stigler and M. Rief. Hidden markov analysis of trajectories in single-molecule experiments and the effects of missed events. *Chemphyschem*, 13(4):1079–1086, 2012. (Cited on pages 40 and 92.)

[139] J. Stigler, F. Ziegler, A. Gieseke, J. C. Gebhardt, and M. Rief. The complex folding network of single calmodulin molecules. *Science*, 334(6055):512–6, 2011. (Cited on pages 21, 23, 45, 55, 79, 91, and 129.)

[140] T. O. Street, L. A. Lavery, and D. A. Agard. Substrate binding drives large-scale conformational changes in the hsp90 molecu-

lar chaperone. *Mol Cell*, 42(1):96–105, 2011. (Cited on pages 13, 115, and 124.)

[141] T. O. Street, L. A. Lavery, K. A. Verba, C. T. Lee, M. P. Mayer, and D. A. Agard. Cross-monomer substrate contacts reposition the hsp90 n-terminal domain and prime the chaperone activity. *J Mol Biol*, 415(1):3–15, 2012. (Cited on page 101.)

[142] T. Suhane, S. Laskar, S. Advani, N. Roy, S. Varunan, D. Bhattacharyya, S. Bhattacharyya, and M. K. Bhattacharyya. Both the charged linker region and atpase domain of hsp90 are essential for rad51-dependent dna repair. *Eukaryot Cell*, 14(1):64–77, 2015. (Cited on page 87.)

[143] K. Svoboda and S. M. Block. Biological applications of optical forces. *Annu Rev Biophys Biomol Struct*, 23:247–85, 1994. (Cited on page 21.)

[144] M. Taipale, I. Krykbaeva, M. Koeva, C. Kayatekin, K. D. Westover, G. I. Karras, and S. Lindquist. Quantitative analysis of hsp90-client interactions reveals principles of substrate recognition. *Cell*, 150(5):987–1001, 2012. (Cited on pages 7 and 12.)

[145] J. K. Tamura and M. Gellert. Characterization of the atp binding site on escherichia coli dna gyrase. affinity labeling of lys-103 and lys-110 of the b subunit by pyridoxal 5'-diphospho-5'-adenosine. *Journal of Biological Chemistry*, 265(34):21342–9, 1990. (Cited on page 133.)

[146] S. F. Tolic-Norrelykke, E. Schaffer, J. Howard, F. S. Pavone, F. Julicher, and H. Flyvbjerg. Calibration of optical tweezers with positional detection in the back focal plane. *Review of Scientific Instruments*, 77(10), 2006. (Cited on page 22.)

[147] J. Trepel, M. Mollapour, G. Giaccone, and L. Neckers. Targeting the dynamic hsp90 complex in cancer. *Nat Rev Cancer*, 10(8):537–49, 2010. (Cited on page 13.)

[148] C. J. Tsai, Z. E. Sauna, C. Kimchi-Sarfaty, S. V. Ambudkar, M. M. Gottesman, and R. Nussinov. Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol*, 383(2):281–91, 2008. (Cited on page 84.)

[149] S. Tsutsumi, M. Mollapour, C. Graf, C. T. Lee, B. T. Scroggins, W. Xu, L. Haslerova, M. Hessling, A. A. Konstantinova, J. B. Trepel, B. Panaretou, J. Buchner, M. P. Mayer, C. Prodromou, and L. Neckers. Hsp90 charged-linker truncation reverses the functional consequences of weakened hydrophobic contacts in the n domain. *Nat Struct Mol Biol*, 16(11):1141–7, 2009. (Cited on pages 87 and 121.)

[150] S. Tsutsumi, M. Mollapour, C. Prodromou, C. T. Lee, B. Panaretou, S. Yoshida, M. P. Mayer, and L. M. Neckers. Charged linker sequence modulates eukaryotic heat shock protein 90 (hsp90) chaperone activity. *Proc Natl Acad Sci U S A*, 109(8):2937–42, 2012. (Cited on pages 87 and 101.)

[151] M. Tsytlonok and L. S. Itzhaki. The how's and why's of protein folding intermediates. *Arch Biochem Biophys*, 531(1-2):14–23, 2013. (Cited on pages 59 and 78.)

[152] Y. von Hansen, A. Mehlich, B. Pelz, M. Rief, and R. R. Netz. Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using bayesian inference. *Review of Scientific Instruments*, 83(9), 2012. (Cited on page 21.)

[153] B. T. Walters, L. Mayne, J. R. Hinshaw, T. R. Sosnick, and S. W. Englander. Folding of a large protein at high structural resolution. *Proc Natl Acad Sci U S A*, 110(47):18898–903, 2013. (Cited on page 79.)

[154] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. Stretching dna with optical tweezers. *Biophys J*, 72(3):1335–46, 1997. (Cited on pages 21, 27, and 28.)

[155] A. L. Watters, P. Deka, C. Corrent, D. Callender, G. Varani, T. Sosnick, and D. Baker. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell*, 128(3):613–24, 2007. (Cited on page 4.)

[156] N. Wayne and D. N. Bolon. Dimerization of hsp90 is required for in vivo function. design and analysis of monomers and dimers. *J Biol Chem*, 282(48):35386–95, 2007. (Cited on page 114.)

[157] N. Wayne and D. N. Bolon. Charge-rich regions modulate the anti-aggregation activity of hsp90. *J Mol Biol*, 401(5):931–9, 2010. (Cited on page 87.)

[158] S. J. Wheelan, A. Marchler-Bauer, and S. H. Bryant. Domain size distributions can predict domain boundaries. *Bioinformatics*, 16(7):613–8, 2000. (Cited on pages 3 and 69.)

[159] S. H. White and R. E. Jacobs. Statistical distribution of hydrophobic residues along the length of protein chains. implications for protein folding and evolution. *Biophys J*, 57(4):911–21, 1990. (Cited on page 82.)

[160] L. Whitesell and S. L. Lindquist. Hsp90 and the chaperoning of cancer. *Nat Rev Cancer*, 5(10):761–72, 2005. (Cited on page 13.)

[161] G. Wildegger and T. Kiefhaber. Three-state model for lysozyme folding: triangular folding mechanism with an energetically

trapped intermediate. *J Mol Biol*, 270(2):294–304, 1997. (Cited on page 59.)

[162] C. J. Wilson, P. Das, C. Clementi, K. S. Matthews, and P. Wittung-Stafshede. The experimental folding landscape of monomeric lactose repressor, a large two-domain protein, involves two kinetic intermediates. *Proc Natl Acad Sci U S A*, 102(41):14563–8, 2005. (Cited on page 55.)

[163] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314(5801):1001–1004, 2006. (Cited on page 35.)

[164] J. O. Wrabl, J. Gu, T. Liu, T. P. Schrank, S. T. Whitten, and V. J. Hilser. The role of protein conformational fluctuations in allostery, function, and evolution. *Biophys Chem*, 159(1):129–41, 2011. (Cited on page 101.)

[165] Y. Wu, E. Kondrashkina, C. Kayatekin, C. R. Matthews, and O. Bilsel. Microsecond acquisition of heterogeneous structure in the folding of a tim barrel protein. *Proc Natl Acad Sci U S A*, 105(36):13367–72, 2008. (Cited on page 79.)

[166] F. Xia, D. Thirumalai, and F. Grater. Minimum energy compact structures in force-quench polyubiquitin folding are domain swapped. *Proc Natl Acad Sci U S A*, 108(17):6963–8, 2011. (Cited on page 82.)

[167] W. Xu, M. Mollapour, C. Prodromou, S. Wang, B. T. Scroggins, Z. Palchick, K. Beebe, M. Siderius, M. J. Lee, A. Couvillon, J. B. Trepel, Y. Miyata, R. Matts, and L. Neckers. Dynamic tyrosine phosphorylation modulates cycling of the hsp90-p50(cdc37)-aha1 chaperone machine. *Mol Cell*, 47(3):434–43, 2012. (Cited on page 12.)

[168] W. Y. Yang and M. Gruebele. Folding at the speed limit. *Nature*, 423(6936):193–7, 2003. (Cited on page 35.)

[169] H. Yu, A. N. Gupta, X. Liu, K. Neupane, A. M. Brigley, I. Sosova, and M. T. Woodside. Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *Proc Natl Acad Sci U S A*, 109(36):14452–14457, 2012. (Cited on page 78.)

[170] R. Zhao, M. Davey, Y. C. Hsu, P. Kaplanek, A. Tong, A. B. Parsons, N. Krogan, G. Cagney, D. Mai, J. Greenblatt, C. Boone, A. Emili, and W. A. Houry. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell*, 120(5):715–27, 2005. (Cited on page 7.)

[171] G. Zoldak, J. Stigler, B. Pelz, H. Li, and M. Rief. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. *Proc Natl Acad Sci U S A*, 110(45):18156–61, 2013. (Cited on page 40.)

[172] A. Zuehlke and J. L. Johnson. Hsp90 and co-chaperones twist the functions of diverse client proteins. *Biopolymers*, 93(3):211–7, 2010. (Cited on page 13.)

[173] A. D. Zuehlke and J. L. Johnson. Chaperoning the chaperone: a role for the co-chaperone cpr7 in modulating hsp90 function in saccharomyces cerevisiae. *Genetics*, 191(3):805–14, 2012. (Cited on page 87.)

[174] R. Zwanzig, A. Szabo, and B. Bagchi. Levinthal's paradox. *Proc Natl Acad Sci U S A*, 89(1):20–2, 1992. (Cited on page 3.)