



PATRICK MÖBMER

SINGLE-MOLECULE FORCE SPECTROSCOPY STUDY  
OF THE GLUCOCORTICOID RECEPTOR  
LIGAND BINDING DOMAIN

HSP70/HSP40 CHAPERONE INTERACTION  
HORMONE BINDING DYNAMICS

DISSERTATION AT THE PHYSICS DEPARTMENT E22  
TECHNISCHE UNIVERSITÄT MÜNCHEN



Technische Universität München  
Fakultät für Physik

SINGLE-MOLECULE FORCE SPECTROSCOPY STUDY  
OF THE GLUCOCORTICOID RECEPTOR  
LIGAND BINDING DOMAIN

HSP70/HSP40 CHAPERONE INTERACTION  
HORMONE BINDING DYNAMICS

Patrick Mößmer

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.

Vorsitz:  
Prof. Dr. Martin Zacharias

Prüfer der Dissertation:  
1. Prof. Dr. Matthias Rief  
2. Prof. Dr. Hendrik Dietz

Die Dissertation wurde am 16.03.2022 bei der Technischen Universität München  
eingereicht und durch die Fakultät für Physik am 10.06.2022 angenommen.



*“You seem awfully smiley for someone whose experiments must be failing 99% of the time.”*

A professor at the University of Santa Barbara, when I told him that I was working on my PhD in single-molecule force spectroscopy.

# Abstract

The Glucocorticoid Receptor (GR) is an important transcription factor linked to a variety of biological functions and diseases. It is one of the most stringent physiological clients of the Hsp90/Hsp70/Hsp40 chaperone system.

The Hsp70/Hsp40 chaperone system (Hsp70/40) constitutes the central component of the cellular surveillance network of chaperones. It contributes to a multitude of cellular processes including protein folding, translocation of polypeptides through membranes, prevention and solubilisation of aggregates, regulation of signaling molecules and degradation of aberrant proteins.

Many details of the GR's signaling pathway and its chaperone cycle as well as the general molecular mechanisms by which Hsp70/40 performs its chaperoning tasks on substrate proteins remain unknown.

In this study, single-molecule force spectroscopy by optical tweezers was used to investigate several aspects of the GR's cycle as a transcription factor. First, the hormone binding behavior of the GR's ligand binding domain (GR-LBD) was examined. An N-terminal secondary structure element was identified, which acts as a "lid" that regulates hormone binding. Next, the interaction of the GR-LBD with Hsp70/40 was studied. After ligand dissociation, it was observed in real-time how Hsp70/40 unfolds the complete GR-LBD in a stepwise manner, typically via up to 5 chaperone-induced unfolding intermediates. Each unfolding step involves binding of an Hsp70 to the GR-LBD and subsequent ATP hydrolysis, stimulated by Hsp40. The kinetics of chaperone mediated unfolding were shown to depend on chaperone concentrations as well as on the presence of the nucleotide exchange factor BAG1. It was found that Hsp70/40 can stabilize new unfolding intermediates, which is evidence that Hsp70/40 can directly interact with the folded core of the protein when working as an unfoldase. In particular, Hsp70/40 induced a new long-lived unfolding intermediate at an unfolded contour length of 32 nm, which coincides with a high-probability Hsp70 binding site predicted by several algorithms. This supports a novel unfolding mechanism where Hsp70 can directly bind to folded protein structures and unfold them upon ATP hydrolysis.

Control experiments were conducted using four different variants of the Hsp40 co-chaperone, namely Ydj1, Hdj2, Hdj1 and the J-Domain of Ydj1. In combination with Hsp70, the complete and stepwise unfolding of GR-LBD was observed for all Hsp40 variants. In the absence of Hsp70, there were differences in the effects of the four Hsp40s on GR-LBD.

Along the way, decisive modifications of the oxygen scavenging system and buffering conditions were established, which brought about drastic improvements of the biochemical assay, enabling the GR-LBD to refold more reliably and thus improving both the quality and quantity of available data.

This work thereby provides important new insights into the function of GR-LBD and its chaperone cycle, as well as the general chaperoning mechanism of Hsp70/40.

# Zusammenfassung

Der Glucocorticoidrezeptor (GR) ist ein wichtiger Transkriptionsfaktor, welcher mit einer Vielzahl biologischer Funktionen und Krankheiten in Verbindung steht. Er ist eines der stringentesten Substratproteine des Hsp90/Hsp70/Hsp40-Systems molekularer Chaperone.

Das Hsp70/Hsp40 Chaperone System (Hsp70/40) stellt das Herzstück des zellulären Überwachungsnetzwerks molekularer Chaperone dar. Es trägt zu zahlreichen zellulären Prozessen bei, die von Proteinfaltung über Translokation von Polypeptiden durch Membranen, Prävention und Auflösung von Aggregaten, Regulation von Signalmolekülen, bis hin zum Abbau fehlgefalteter Proteine reichen.

Viele Details der Signalübertragung und der Chaperone-Interaktion des GR, sowie auch die allgemeinen molekularen Mechanismen, durch die Hsp70/40 seine Aufgaben als Chaperone an Substratproteinen ausübt, sind noch unbekannt.

In dieser Arbeit wurde Einzelmolekül-Kraftspektroskopie mit einer optischen Pinzette angewandt, um diverse Aspekte des Kreislaufs des GR in seiner Rolle als Transkriptionsfaktor zu untersuchen. Zunächst wurde die Hormonbindung der Ligandenbindungsdomäne des GR (GR-LBD) erforscht. Ein N-terminales Sekundärstrukturelement wurde identifiziert, welches als eine Art "Deckel" fungiert, der die Hormonbindung reguliert. Im Anschluss daran wurde die Interaktion zwischen GR-LBD und Hsp70/40 untersucht. Nach der Dissoziation des Liganden wurde in Echtzeit beobachtet, wie Hsp70/40 die gesamte GR-LBD schrittweise entfaltet, wobei typischerweise bis zu 5 chaperone-induzierte Entfaltungsintermediate bevölkert wurden. Jeder Entfaltungsschritt beinhaltet das Binden eines Hsp70 an die GR-LBD, sowie anschließende ATP Hydrolyse, welche durch Hsp40 stimuliert wird. Es wurde gezeigt, dass die Kinetik der chaperone-induzierten Entfaltung sowohl von den Konzentrationen der Chaperone, als auch von dem Vorhandensein des Nukleotidaustauschfaktors BAG1 abhängt. Es wurde herausgefunden, dass Hsp70/40 neue Entfaltungsintermediate stabilisieren kann, was ein Beleg dafür ist, dass Hsp70/40 als Unfoldase direkt mit dem gefalteten Kern des Proteins interagieren kann. Insbesondere induzierte Hsp70/40 ein neues langlebiges Entfaltungsintermediat bei 32 nm entfalteter Konturlänge, welches mit einer von mehreren Algorithmen vorhergesagten hochwahrscheinlichen Hsp70-Bindestelle übereinstimmt. Diese Erkenntnis unterstützt einen neuen Entfaltungsmechanismus, bei dem Hsp70 direkt an gefaltete Proteinstrukturen binden und diese durch ATP-Hydrolyse entfalten kann.

Es wurden Kontrollexperimente mit vier verschiedenen Varianten der Hsp40 Co-Chaperone durchgeführt, nämlich Ydj1, Hdj2, Hdj1 und die J-Domain von Ydj1. In Kombination mit Hsp70 wurde die vollständige und schrittweise Entfaltung der GR-LBD für alle vier Hsp40 Varianten beobachtet. In Abwesenheit von Hsp70 wurden Unterschiede in der Wirkung der vier Hsp40s auf die GR-LBD festgestellt.

Im Zuge dieser Experimente wurden entscheidende Modifikationen des Oxygen Scavenging Systems und der Pufferbedingungen implementiert, welche eine drastische Verbesserung des biochemischen Assays zur Folge hatten. Die GR-LBD konnte in der Folge zuverlässiger zurückfalten, was sowohl die Qualität als auch die Menge der verfügbaren Daten verbesserte.

Die vorliegende Arbeit liefert daher neue Erkenntnisse zur Funktion der GR-LBD und ihres Chaperone-Kreislaufs, sowie zum Wirkungsmechanismus von Hsp70/40.

# Publications

Some of the ideas, results, figures and tables contained in this thesis have appeared previously in the following publications:

**Active unfolding of the Glucocorticoid Receptor by the Hsp70/Hsp40 chaperone system in single-molecule mechanical experiments.** Patrick Moessmer, Thomas Suren, Ulrike Majdic, Vinay Dahiya, Daniel Rutz, Johannes Buchner, Matthias Rief. Accepted for publication in Proceedings of the National Academy of Sciences (03/2022).

**Single-molecule force spectroscopy reveals folding steps associated with hormone binding and activation of the glucocorticoid receptor.** Suren, T., Rutz, D., Mößmer, P., Merkel, U., Buchner, J., Rief, M. (2018). Proceedings of the National Academy of Sciences, ISSN 0027-8424

**The switch from client holding to folding in the Hsp70/Hsp90 chaperone machineries is regulated by a direct interplay between co-chaperones.** Vinay Dahiya, Daniel Andreas Rutz, Patrick Moessmer, Moritz Mühlhofer, Janis Lawatscheck, Matthias Rief, Johannes Buchner (2022). Molecular Cell, ISSN 1097-2765

# Acknowledgements

First of all, I would like to thank Prof. Matthias Rief for being a dedicated, knowledgeable and encouraging supervisor with an open-door policy. He always provided helpful discussions, constructive feedback and resourceful ideas. Also, he created a relaxed and comfortable, yet productive work atmosphere at his lab, enabling cutting-edge science.

Thomas Suren did the GR project's pioneering work and this thesis heavily builds on his achievements. Thanks for helping me get started at the lab and patiently transferring your profound knowledge of the GR and lab techniques to me. Also, thanks for all the amazing parties at "Villa Schandfleck".

Special thanks go to Ulrike Majdic, who took great care of the biochemical side of the project. Over the years, she purified a variety of proteins for me. You make challenging tasks look effortless and always maintain your great mood while doing so. It is fair to say that this work would not have been possible without you.

A big thank you goes to Kasia Tych for maintaining the "Ducktrap" setup for years. I am glad that we went from being colleagues to being friends.

Thanks to my colleagues and labmates over the years, Anubhuti Singh, Mihai Bodescu, Marco Grison, Leone Rosetti, Daniela Bauer, Johannes Stigler, Andreas Weiß, Ulli Merkl, Abhigyan Sengubta, Bin Li, Fabio Falleroni, Katharina Augustin and Andreas Walbrun for valuable discussions, technical help, and always keeping me great company. One could not wish for a nicer working environment.

A lab is always only as good as its TAs. Therefore I would like to express my great gratitude to Gabriele Chmel, Monika Rusp and Karin Vogt for being always helpful, attentive and affable whenever I needed their support.

Thanks to Nicole Baker, Elke Fehsenfeld and Anne Ploß for taking great care of all the administrative and bureaucratic tasks.

I was very fortunate to meet Vera Hechtel early during my PhD at the Rief group. Conducting experiments at the optical trap and taking breaks was much more fun when you were around. Thank you for becoming my close friend over the last years.

It was Magnus Bauer who suggested that I apply at the Rief group in the first place. I learned and adapted a lot from you, both scientifically and personally. And I was always impressed by your cordiality, versatility and profundity in all areas. I hope that, even though you are at Stanford now and might go to outer space with ESA, we will run many more Marathons together and remain comrades for life.

I would like to thank my family and friends for their kindness, support and encouragement throughout the years.

# Contents

<b>Introduction</b>	<b>1</b>
<b>Theoretical and Experimental Background</b>	<b>3</b>
<b>1 Biochemical Background</b>	<b>3</b>
1.1 Biochemistry of Proteins . . . . .	3
1.2 Molecular Structure of the GR-LBD . . . . .	4
1.3 Physiology of GR Signaling . . . . .	5
1.4 Heat Shock Proteins . . . . .	6
1.5 The Chaperone Cycle of the GR . . . . .	7
1.6 The Hsp70/Hsp40 Chaperone System . . . . .	8
1.6.1 The Purpose of Hsp70s . . . . .	8
1.6.2 Ubiquity and Conservation of Hsp70s Across Species . . . . .	9
1.6.3 Structure of Hsp70s . . . . .	9
1.6.4 Hsp70's Allosteric Cycle during Substrate and Nucleotide Interaction . . . . .	10
1.6.5 Substrate Specificity of Hsp70s . . . . .	10
1.7 Affinity . . . . .	11
<b>2 Optical Trapping</b>	<b>14</b>
2.1 History of Optical Trapping as an Experimental Method . . . . .	14
2.2 The Physics Underlying Optical Trapping . . . . .	17
2.2.1 Mie Regime . . . . .	17
2.2.2 Rayleigh Regime . . . . .	19
2.2.3 In Between the Mie and the Rayleigh Regime . . . . .	20
2.3 The Optical Trap Setup Used in this Work . . . . .	20
<b>3 Experimental Approach</b>	<b>22</b>
3.1 Stepwise Assembly of the Dumbbell Assay . . . . .	22
3.1.1 Oligo Attachment via Cysteine-Maleimide Chemistry . . . . .	23
3.1.2 DNA Handles Hybridization . . . . .	24
3.1.3 Incubation of the Protein-Oligo-Handle Construct with Streptavidin Beads . . . . .	24
3.1.4 Assembly of the Full Dumbbell Assay by Collision with Anti-Digoxigenin Beads . . . . .	25
3.2 Measurement Modes at the Optical Trap . . . . .	26
3.2.1 Stretch-Relax Experiments . . . . .	26
3.2.2 Passive-Mode Experiments . . . . .	27
<b>4 Data Analysis</b>	<b>28</b>
4.1 Wormlike-Chain Model . . . . .	28
4.2 Hidden Markov Models . . . . .	29
4.3 Extracting Equilibrium Free Energies from Passive Mode Traces . . . . .	31
4.4 Transition Rates . . . . .	32
4.5 Single-Exponential Lifetimes of States . . . . .	33

<b>Results</b>	<b>35</b>
<b>5 GR-LBD Basics</b>	<b>35</b>
5.0.1 Introduction to Passive-Mode Experiments with GR-LBD . . .	35
5.0.2 Variation of DEX Concentration in Passive-Mode Experiments	35
5.0.3 Variation of Force in Passive-Mode Experiments . . . . .	38
5.0.4 Introduction to Stretch-Relax Experiments with GR-LBD . .	39
<b>6 Structural Identification of the "Lid" Governing Hormone Binding</b>	<b>43</b>
6.1 Tethering between an Internal and a Terminal Cysteine . . . . .	43
6.2 Competition Assay between Lid and Peptide Chain . . . . .	46
6.3 Linker Construct . . . . .	49
<b>7 Improvement of the Biochemical Assay of GR-LBD Experiments</b>	<b>52</b>
7.1 DTT Abolishes the "Ligand Binding Incompetent State" of GR-LBD	52
7.2 Optimization of the Oxygen Scavenging System . . . . .	55
7.2.1 Misfolding Tendency of GR-LBD . . . . .	55
7.2.2 Glucose Oxidase/Glucose/Catalase . . . . .	56
7.2.3 Protocatechuate-3,4-dioxygenase/Protocatechuic Acid . . . .	57
7.2.4 Pyranose Oxidase/Glucose/Catalase . . . . .	58
7.2.5 Scavenging Systems Discussion . . . . .	58
<b>8 Interaction of the GR-LBD with Hsp70/40</b>	<b>63</b>
8.1 Hsp70/40 Induces Complete and Stepwise Unfolding of GR-LBD . .	63
8.2 Lid Flipping Remains Unaffected by Chaperones . . . . .	63
8.3 Kinetics of the Chaperone-Induced GR-LBD Unfolding . . . . .	65
8.3.1 Variation of Hsp70 Concentration while Holding Hsp40 Con-	
centration Constant . . . . .	65
8.3.2 Variation of Hsp40 Concentration while Holding Hsp70 Con-	
centration Constant . . . . .	66
8.4 Discussion of Chaperone-Induced Unfolding Kinetics . . . . .	68
8.5 Hsp70/40 Inhibits Refolding of GR-LBD . . . . .	70
8.5.1 Hsp70/40 Inhibiting Refolding of GR-LBD in Stretch-Relax	
Experiments . . . . .	70
8.5.2 Hsp70/40 Inhibiting Refolding of GR-LBD in Passive-Mode	
Experiments . . . . .	70
8.6 BAG1 Counteracts the Unfolding by Hsp70/40 . . . . .	71
8.7 GR-LBD Can Refold at Low Hsp70/40 Concentrations . . . . .	72
8.8 Hsp70/40 Unfolding Works in the Absence of an External Force Bias	73
8.9 Hsp70 Binding Sites in GR-LBD Sequence - Experiment vs. Algorithms	75
8.9.1 The LIMBO Algorithm . . . . .	77
8.9.2 The ChaperISM Algorithm . . . . .	79
8.10 Intermediates in Chaperone-Induced and Force-Induced Unfoldings .	79
<b>Discussion</b>	<b>82</b>
<b>9 The Hsp70/40 Unfolding Mechanism</b>	<b>82</b>
9.1 Hsp70 as a Holdase . . . . .	82
9.2 Hsp70 as an Unfoldase . . . . .	82

9.3	The Hsp70/40-Induced Unfolding of GR-LBD . . . . .	83
9.4	When Exactly Do Hsp70 Binding and Hydrolysis Occur? . . . . .	83
9.5	The Overall Model of Hsp70/40-Induced Unfolding of GR-LBD . . .	85
9.6	Potential Unfolding Mechanisms of Hsp70/40 . . . . .	87
9.6.1	Ratchet . . . . .	87
9.6.2	Entropic Pulling . . . . .	88
9.6.3	Direct Interaction between Hsp70 and Folded GR-LBD . . .	88
<b>10</b>	<b>The Role of Hsp40</b>	<b>90</b>
10.1	Ydj1 . . . . .	90
10.2	Hdj2 . . . . .	93
10.3	Hdj1 . . . . .	94
10.4	J-Domain . . . . .	94
10.5	Hsp70/40 Combination for All Hsp40 Variants . . . . .	95
10.6	Discussion of the Four Hsp40 Variants and Their Effects . . . . .	96
<b>11</b>	<b>Conclusion and Outlook</b>	<b>98</b>
	<b>Appendix</b>	<b>99</b>
<b>12</b>	<b>Methods and Protocols</b>	<b>99</b>
12.1	Maleimide Oligos Attachment Protocol . . . . .	99
12.2	Azide Oligos Attachment Protocol . . . . .	100
12.3	Comparison of Maleimide Oligos vs. Azide Oligos . . . . .	101
12.4	DNA Handles Hybridization . . . . .	102
12.5	Quality Control of Oligo Attachment and DNA Handles Hybridization via Agarose Gel . . . . .	102
12.6	Preparation of the Optical Trap Measurements . . . . .	103
12.7	Global Fit Parameters . . . . .	103
12.8	Assembly of Dumbbell Assay . . . . .	104
12.8.1	Protein+Oligo+Handles (POH) . . . . .	104
12.8.2	Protein+Oligo+Handles' (POH') . . . . .	104
12.8.3	Incubation with Streptavidin Beads (S) . . . . .	104
12.8.4	Final Solution (F) . . . . .	105
12.9	Protein Expression and Sequences . . . . .	106
12.9.1	GR-LBD . . . . .	106
12.9.2	Ydj1 . . . . .	106
12.9.3	Human Hsp70 . . . . .	107
12.9.4	J-Domain . . . . .	108
12.9.5	Hdj2 . . . . .	108
12.10	ChaperISM Algorithm Prediction . . . . .	109
	<b>Bibliography</b>	<b>120</b>

# Introduction

The Glucocorticoid Receptor (GR) is a ligand-dependent transcription factor linked to a remarkable range of essential biological functions and processes [1] [2]. It is found in all vertebrates and expressed in nearly every cell of the human body [3] [4]. The GR is a key factor in immune and inflammatory response, metabolism, development, reproduction and stress response [5] [6]. As a result, it also plays a decisive role in a multitude of diseases and clinical conditions, including inflammatory disorders [7], anaphylactic shock [8], depression [9], cerebral edema [10], diabetes [11], asthma [12], leukemia [13], osteoporosis [14], organ transplants [15], septic shock [16], preterm delivery [17], chronic obstructive pulmonary disease [18] and rheumatoid arthritis [19].

Thanks to its versatile functionality, the GR constitutes an important drug target. The most common medical application is the use of glucocorticoids, which bind to the GR's ligand binding domain (GR-LBD) and thereby induce its functionality as a transcription factor. They have proven to be of vital importance in acute, long-term and prophylactic treatments in all kinds of medical scenarios.

Endogenous cortisone, which is the body's primary glucocorticoid, was first isolated in the late 1930s by Mason, Kendall and others [20] [21] [22]. In the mid-1940s it was first synthesized by Sarett [23]. Soon, the potential benefits of administering this hormone to rheumatic patients became apparent [24]. As Dudley Hart remarked in retrospect [25]:

“Therapy was now dated BC (before cortisol) or after (AC)!”

In 1950, Philip S. Hench, Edward C. Kendall, and Tadeusz Reichstein received the Nobel Prize in Physiology or Medicine *“for their discoveries relating to the hormones of the adrenal cortex, their structure, and biological effects.”* [26]

Today, cortisol counts among the most-prescribed medications worldwide [27]. In the treatment of critically ill COVID-19 patients during the present pandemic, it reduces the risk of all-cause mortality and duration of ventilation [28] [29].

While in short-term treatments glucocorticoids have a relatively low risk of side-effects, in long-term treatments, high doses of these drugs can lead to many adverse and potentially severe consequences [30] [31]. Again, these include a plethora of conditions, including musculoskeletal, metabolic, endocrine, cardiovascular, dermatologic and ophthalmologic side effects. GR-modulating drugs that selectively preserve the anti-inflammatory and immunosuppressive effects of glucocorticoids, but do not show the adverse side-effects induced by chronic glucocorticoid therapy, have long been sought without success [1].

It is a recurring bio-molecular principle that great functional and conformational versatility often comes at the cost of stability or reliability. The GR is notoriously prone to aggregation and misfolding in vitro and difficult to purify [32] [33]. Therefore, during its intricate cycle as a transcription factor, the GR relies on the assistance of a set of molecules called “chaperones” [34] [35]. Molecular chaperones are an evolutionarily heavily conserved set of proteins shared by prokaryotes as well as eukaryotes. They perform various tasks on thousands of client proteins, including de novo protein folding and maturation [36], protection from aggregation [37], dilution of aggregates [38], protein transport through membranes [39], protein complex

assembly and disassembly [40] and many more [41]. In human cell lines, 10% of the entire proteome is made up of chaperones [42]. The GR is one of the most stringent physiological clients of the Hsp90/Hsp70/Hsp40 chaperone system [33] [43], which renders it a paradigm candidate in order to examine the molecular mechanisms of chaperoning.

In this study, the method of choice was single-molecule force spectroscopy by optical tweezers. This experimental approach offers the unique possibility to observe the operation and interaction of single molecules in real-time, to study the folding behavior and energy landscape of the protein in question, while also receiving structural resolution. Given the general biochemical fragility of the GR, single-molecule optical trapping constitutes an elegant way to circumvent some (but by far not all) of the problems the GR poses in bulk experiments.

First, the folding and ligand interaction of a stabilized variant of the GR-LBD (F602S) were investigated [44]. Afterwards, building on our deep understanding of this protein, the first part of the GR's chaperone cycle, namely the GR-LBD's interaction with the Hsp70/Hsp40 system, was examined.

Hence, this study covers two areas of medical and biophysical interest at once: the specific function of the GR-LBD as well as the more general chaperoning mechanism of Hsp70/40. A deeper understanding of the GR's operation on the molecular level might lead to better and more refined clinical treatment possibilities. At the same time, investigating Hsp70/40 from the perspective of the GR as its obligate client might reveal important and even more broadly applicable insights into this universal maintenance and control network of organisms.

# Theoretical and Experimental Background

## 1 Biochemical Background

### 1.1 Biochemistry of Proteins

Proteins are large biomolecules that constitute the machinery of life in all prokaryotic and eukaryotic cells [45] [46] [47]. They contribute to practically all biological processes within organisms from DNA replication, metabolism, transport of molecules, response to stimuli, signaling and mechanical stabilization [48] [49].

Proteins consist of chains of amino acids, which are covalently connected via peptide bonds [50]. These are formed when the carboxyl group of one amino acid reacts with the amino group of another amino acid, releasing one water molecule during the reaction.

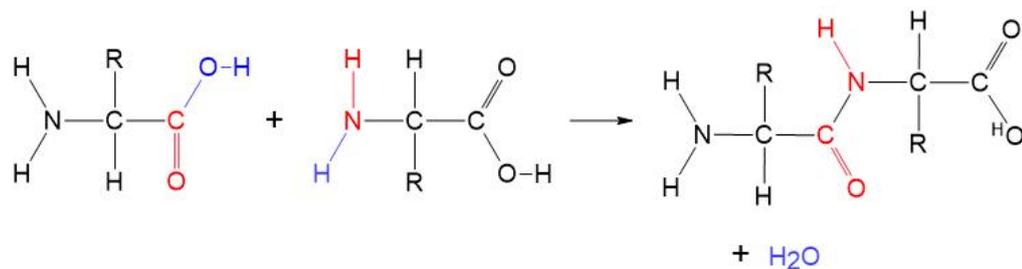


Figure 1: **Peptide bond formation:** The carboxyl group of the left amino acid (-COOH) reacts with the amino group of the right amino acid (-NH<sub>2</sub>), leaving the two amino acids covalently connected. One water molecule (H<sub>2</sub>O) is released during the reaction.

Therefore, the peptide chain exhibits directionality from the N-terminus (amino group) to the C-terminus (carboxyl group).

The sequence of amino acids in a particular protein is encoded in the DNA of its corresponding gene [48]. A triplet of nucleotides in the DNA, named codon, encodes one specific amino acid. There are 20 proteinogenic amino acids. While each triplet unambiguously encodes only one specific amino acid, the code is degenerate insofar as several triplets can encode the same amino acid. During transcription, a segment of DNA is copied into messengerRNA. In the following translation, the peptide chain is assembled by ribosomes according to their messengerRNA blueprint with the help of transferRNA. The resulting peptide chains must then fold into a complex three-dimensional structure.

According to Anfinsen's dogma, the natively folded structure of small globular proteins is determined merely by their amino-acid sequence [51]. This requires that, in its natural physiological environment, the native fold of a protein must be a unique, stable and kinetically accessible minimum of free energy. The native fold of a protein is defined by the interactions of the amino-acid residues of its peptide chain with each other and with their environment. Amino acid residues can be charged, hydrophobic, acidic or polar [45].

As Levinthal pointed out in his famous paradox [52], an unfolded peptide chain pos-

sesses a large number of degrees of freedom, which leads to an overwhelming number of possible conformations. Proteins can therefore not rely on mere sampling of all possible conformations until the energetic minimum has been reached. Even if the time spent in each conformation was on the picosecond scale, such a process would require timescales larger than the age of the universe, while most proteins typically fold within less than a second. The solution to this problem is the fact that the energy-landscape of protein folding is funnel-like [53]. It works via processes such as folding of local secondary structures, e.g. alpha-helices and beta-sheets, folding intermediates and, above all, shielding of hydrophobic amino acids in the core of the protein. The secondary and tertiary structure of the GR-LBD will be discussed in detail in the following section.

While many proteins fold on their own, some require the assistance of “molecular chaperones” (cf. section 1.4) to reach their native state. [54]

## 1.2 Molecular Structure of the GR-LBD

The GR consists of three main functional domains: the N-terminal domain (NTD), the DNA-binding domain (DBD) and the ligand binding domain (LBD) [3]:



Figure 2: The three domains of the GR.

It has been shown that the interaction between the GR and chaperones exclusively occurs via the GR’s ligand binding domain (GR-LBD) [34] [55]. As for many Hsp90 client proteins, purification of sufficient quantities of apo GR-LBD for in vitro experiments used to be challenging due to solubility problems and the incapability of GR-LBD to bind ligand in the absence of Hsp90 [56] [57]. In order to better understand the problematic purification of GR, Bledsoe et al. performed sequence alignment of GR with two related steroid receptors, the Progesterone Receptor (PR) and the Androgen Receptor (AR), which had successfully been expressed and purified from *E.coli* [56]. They were looking for residues that were hydrophobic in GR and hydrophilic in PR and AR, since these would likely contribute to solubility and aggregation problems. They were able to identify a single mutation in residue 602 from phenylalanine to serine (F602S) in helix 5 of GR-LBD, which significantly improved the expression of GR-LBD in the presence of Dexamethasone (DEX) while maintaining the functionality of wild-type GR-LBD. This GR-LBD F602S construct goes by the identifier 1m2z in the Protein Data Base [58] [56] and measures 257 amino acids in length. According to a crystal structure where DEX and TIF-2 are bound to the GR-LBD, it folds into a globular structure consisting of 11 alpha-helices and 4 short beta-sheets (Fig.3) [56].

In subsequent years, most major studies on the ligand binding behavior as well as the chaperone cycle of GR-LBD, such as Kirschke et al. [43] and Lorenz et al. [33], used the stabilized F602S construct, and we followed this approach. Throughout this thesis, whenever GR-LBD is written, the F602S construct is meant.

Since for our optical trapping experiments the protein needs to be tethered between DNA handles via maleimide chemistry (cf. section 3.1), we added cysteines at the N- and C-terminus of GR-LBD and substituted a surface cysteine at residue 638 by

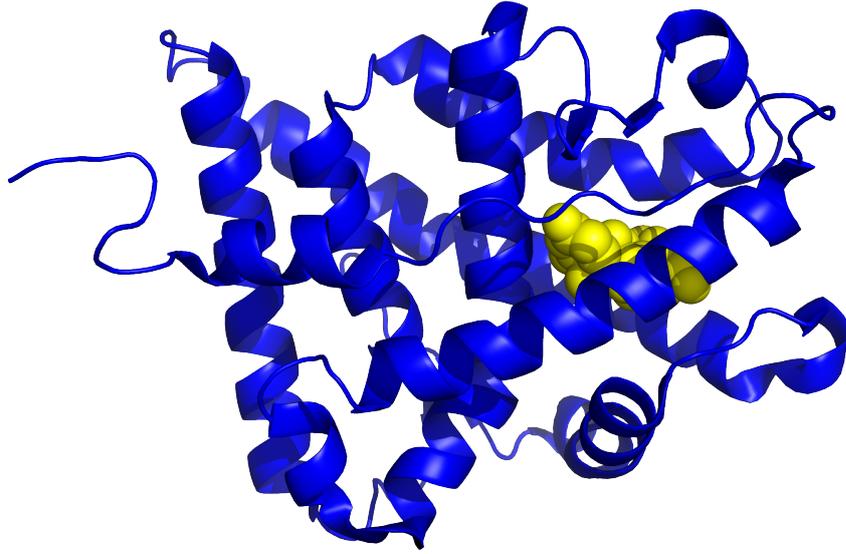


Figure 3: **Crystal structure of the GR-LBD:** The GR-LBD consists of 11  $\alpha$ -helices and 4 small  $\beta$ -strands. The ligand DEX is depicted in yellow. PDB-identifier: 1m2z [58] [56]

an aspartate (C638D). The resulting sequence used in our experiments, with highlighted mutations F602S (red), C638D (green) and cysteines at the termini (blue), is the following:

N-SACK-GR-LBD F602S/C638D-KCL-C (GR-LBD):  
 SA**C**AKQLTPTLVSLLEVIEPEVLYAGYDSSVPDSTWRIMTTLNMLGGRQV  
 IAAVKWAKAIPGFRNLHLDDQMTLLQYSWMS**L**MAFALGWRSYRQSSA  
 NLLCFAPDLIINEQRMTLP**D**MYDQCKHMLYVSSELHRLQVSYEEYLCM  
 KTTTTLLSSVPKDGLKSQELFDEIRMTYIKELGKAIVKREGNSSQNWQRF  
 YQLTKLLDSMHEVVENLLNYCFQTFLDKTMSIEFPEMLAEIITNQIPKY  
 SNGNIKKLLLFHQ**KCL**

### 1.3 Physiology of GR Signaling

The natural glucocorticoid of the human body is cholesterol. It is secreted by the *zona fasciculata* of the adrenal glands [59] [60]. The synthesis and release of cholesterol are under dynamic daily and short term regulation by the hypothalamic-pituitary-adrenal axis [61]. The availability of cholesterol in tissues is further regulated by globulin, which binds glucocorticoids in serum, and by locally expressed  $11\beta$ -hydroxysteroid dehydrogenase enzymes ( $11\beta$ -HSD) [59] [62].

An imbalance in glucocorticoid levels such as chronic elevation or deficiency can lead to pathological conditions such as Cushing's disease and Addison's disease [59].

Apo GR-LBD resides in the cytosol in a complex with Hsp90 and p23, ready to bind ligand. Upon ligand binding, the complex dissociates and the GR translocates to the nucleus, where it transiently binds to target genes and regulates their transcription both via transactivation and transrepression using various strategies [4]. One of the mechanisms employed by the GR to regulate transcription of a gene is to bind to glucocorticoid responsive elements (GREs) as a homodimer. When bound

to GREs in promoter or enhancer regions of certain genes, the GR can increase the transcription rate of these genes through direct interaction with the transcription machinery, co-activators or other transcription factors [63]. The GR can also repress transcription of target genes by binding as a monomer to specific DNA sequences termed negative GREs [64]. Apart from the GR's capability to directly bind to DNA sequences, it can also regulate the transcription of genes via protein-protein interaction with other transcription factors, downregulating their activity [65].

After fulfilling its task as a transcription factor, the GR returns to the cytoplasm, ready to bind ligand again [66].

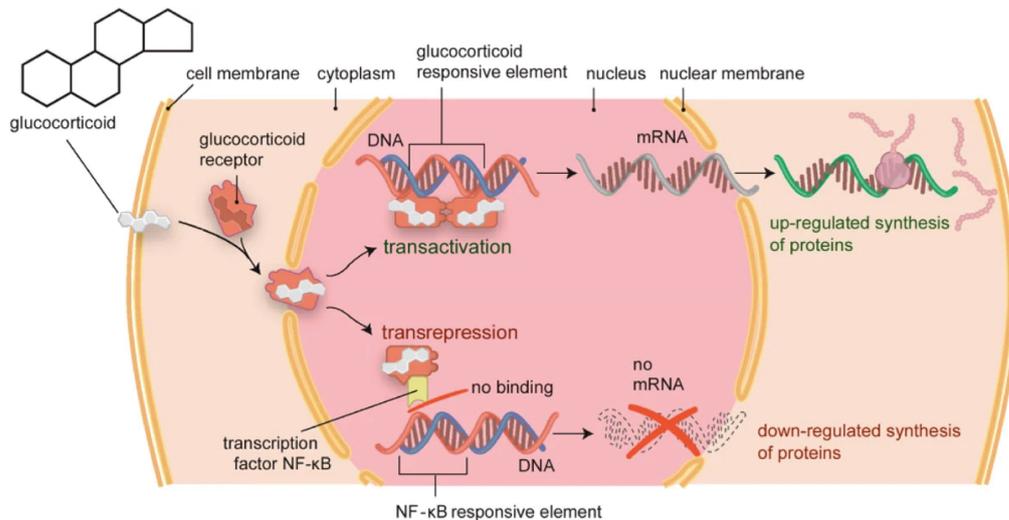


Figure 4: **The GR's cycle as a transcription factor:** The GR resides in the cytosol and, upon ligand binding, translocates to the nucleus to regulate the transcription of target genes both via transactivation and transrepression. Figure reprinted from [67] with permission. Copyright 2014 by Springer Nature.

During its intricate cycle of ligand binding, voyage into the nucleus, binding of DNA, transcription regulation, ligand dissociation and return to the cytoplasm, the GR heavily relies on the assistance of the Hsp90/Hsp70/Hsp40 chaperone system [43] [68] [59]. This interaction between GR and its chaperones was closely examined in this work. The next section will therefore introduce molecular chaperones in more detail.

## 1.4 Heat Shock Proteins

Heat shock proteins (Hsps) are a subgroup of molecular chaperones, which owe their name to the fact that their expression is upregulated under heat stress. They were discovered by accident in 1962, when a member of Ferruccio Ritossa's lab unintentionally increased the incubation temperature of *Drosophila* fruit flies. Upon subsequent examination of the *Drosophila* chromosomes, a characteristic "puffing" layer around them was observed, which was linked to the increased gene transcription of a so far unknown class of proteins [69].

Soon it was found that the upregulation of the expression of Hsps is actually a more general response to many kinds of stressful conditions, which, apart from heat stress, also include toxic, cold, starvation, inflammation or UV radiation stress.

Different Hsps are classified according to their molecular weight in kDa, such as

Hsp100 ( $\sim 100$  kDa), Hsp90, Hsp70, Hsp60, and Hsp40. The larger Hsp100, Hsp90 and Hsp70 contain an ATPase domain and, upon hydrolysis, can undergo conformational changes. The smaller Hsp60 and Hsp40 are often classified as co-chaperones, which regulate ATPase activity and substrate specificity of the larger Hsps. Hsps are evolutionarily highly conserved. All living organisms possess homologues of Hsps. They constitute the cellular protein surveillance system and therefore play essential roles in a multitude of biochemical processes including folding of proteins, transport of proteins through membranes, inhibition of aggregation, dissolution of aggregates, regulation of signaling molecules and many more [70] [71].

### 1.5 The Chaperone Cycle of the GR

The GR, like many other signaling molecules, strictly depends on the Hsp90/Hsp70/Hsp40 chaperone system for *in vivo* ligand binding and activity [66] [72] [73]. Hsp90 is known to form a complex with apo GR, which is stabilized by p23. In this complex, Hsp90 enables ligand binding by GR, while at the same time preventing aggregation. In order to end up in this complex with Hsp90 and p23, the GR has to undergo interaction with Hsp70/40 first, which then delivers it to Hsp90. It has been shown that Hsp90, Hsp70, Hsp40, Hop and p23 constitute a minimal chaperone system that guarantees proper GR functionality *in vitro* [43] [74]. The general chronology of chaperones entering and exiting this pathway has been studied in detail [75] [76].

So far, the agreed-upon chaperone-cycle of the GR goes as depicted in Fig.5.

After ligand dissociation, Hsp40, Hsp70, Hop, Hsp90 and p23 act on the GR sequentially. First, Hsp70 binds to GR-LBD in a process requiring both Hsp40 and ATP. Until the publication of this study, Hsp70 was thought to at least partially unfold the GR-LBD [43]. Hop mediates between the GR-LBD-Hsp70 complex and Hsp90. ATP hydrolysis by Hsp90 is necessary to release Hsp70 and Hop from this complex and form a stable new complex consisting of apo GR-LBD, Hsp90 and p23. In this state, GR-LBD awaits the binding of a new ligand to initiate its function as a transcription factor anew. This GR-LBD-Hsp90-p23 complex exhibits a particularly high affinity for ligand binding.

Despite this extensive knowledge concerning the chaperone cycle of GR, the mechanical and temporal details as well as the precise coordination between the chaperones still remain largely unknown. Focussing on the GR's association with Hsp70/40, the present study provides completely new insights into the interaction between chaperones and client molecules. It elucidates the details of the first part of the GR's chaperone cycle, namely the action of the Hsp70/40 chaperone system. Therefore, the following section will give a more detailed introduction to this particular chaperone system.

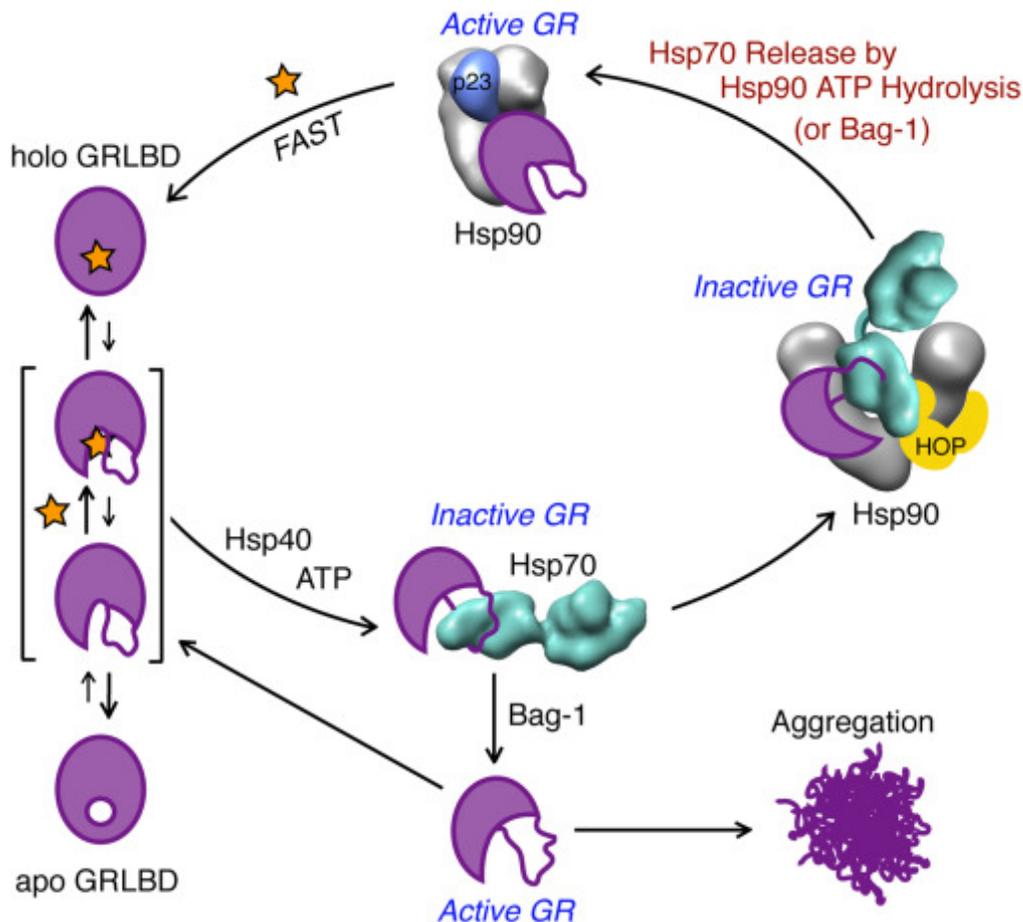


Figure 5: **The GR's cycle of ligand binding and chaperone interaction:** At the top, the GR is depicted in its active form, ligand-unbound and in a complex with Hsp90 and p23. This is the complex the GR assumes in the cytoplasm, waiting for a ligand to bind. Upon ligand binding, the GR fulfills its tasks as a transcription factor. Afterwards, Hsp40, Hsp70 and finally Hop, Hsp90 and p23 sequentially interact with GR-LBD and in an orchestrated maneuver return it to its active apo form. Figure reprinted from [43] with permission. Copyright 2014 by Elsevier.

## 1.6 The Hsp70/Hsp40 Chaperone System

### 1.6.1 The Purpose of Hsp70s

Hsp70s constitute the central component of the cellular surveillance network of chaperones and as such assist in a multitude of protein folding and remodelling processes [77] [78]. They contribute to all life stages of proteins ranging from translation to degradation and hence are indispensable for maintaining protein homeostasis. Hsp70s help with the folding of newly synthesized proteins, the translocation of polypeptides through membranes, the prevention and solubilisation of aggregates, the disassembly of protein complexes, the regulation of signaling molecules, as well as the degradation of aberrant proteins.

### 1.6.2 Ubiquity and Conservation of Hsp70s Across Species

Hsp70s are the most ubiquitous and conserved of all Hsps [79]. They constitute the predominant Hsp in humans, plants and bacteria [80]. Humans express at least 13 different Hsp70 homologues [81]. Across species, Hsp70s exhibit very high structural homology and conserved functional features. The similarity across species goes so far that *Drosophila* Hsp70 can, in fact, complement mammalian Hsp70 with respect to the protection from heat stress [82].

### 1.6.3 Structure of Hsp70s

Hsp70s are composed of highly conserved structural domains, as depicted in Fig.6:

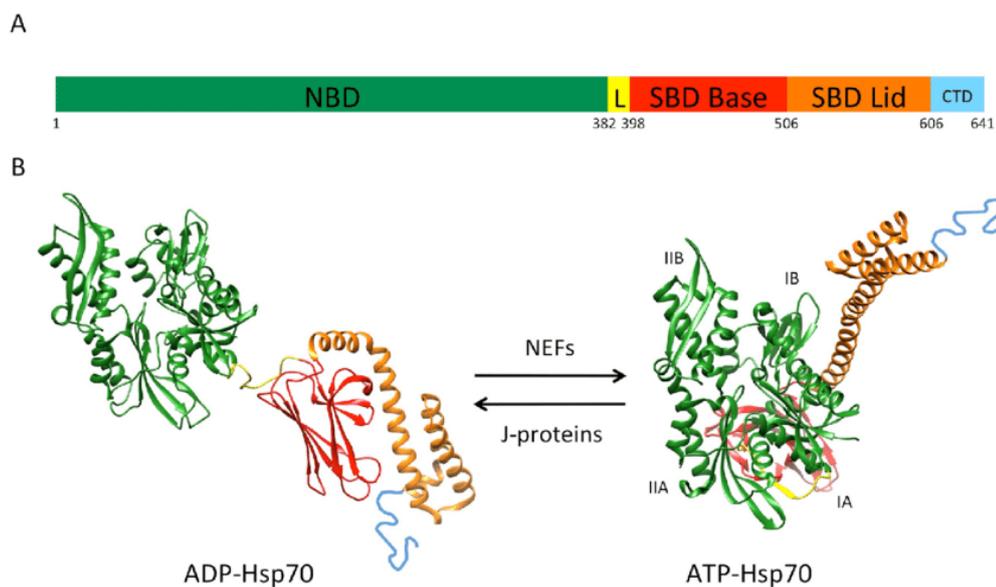


Figure 6: **The Structure of Hsp70:** **A)** Scheme of Hsp70's domain organization. Hsp70 comprises a nucleotide binding domain (NBD), a linker (L), a substrate binding domain (SBD) that consists of a base (SBD- $\beta$ ) and a corresponding lid (SBD- $\alpha$ ) and a C-terminal domain (CTD). **B)** Hsp70's structure in the closed ADP-bound conformation (left) and the open ATP-bound conformation (right). Figure reprinted from [83] with permission. Copyright 2018 by NLM.

As Fig.6 illustrates, Hsp70 consists of a 44kDa N-terminal nucleotide binding domain (NBD), a 15kDa substrate binding domain base (SBD- $\beta$ ) and a 10kDa C-terminal helical substrate binding domain lid (SBD- $\alpha$ ), which can close down on the SBD- $\beta$ . The nucleotide binding domain consists of four subdomains, which form two lobes divided by a deep cleft [77] [84]. The nucleotide binding pocket is situated at the bottom of this cleft. The binding of ATP leads to a rotation of the two lobes, which allosterically transmits to the other domains.

The substrate binding domain consists of an eight-stranded  $\beta$ -sandwich, which contains the hydrophobic substrate binding cavity.

In the nucleotide-free as well as in the ADP-bound state, the SBD- $\alpha$  is docked onto the SBD- $\beta$  to fully enclose the substrate binding cavity, while in the ATP-bound state it is not.

#### 1.6.4 Hsp70's Allosteric Cycle during Substrate and Nucleotide Interaction

Hsp70's chaperoning function involves fast association and timely release of substrate in order to promote folding and inhibit aggregation. This controlled association and dissociation requires the coordinated action of all of its domains, involving intricate allosteric changes that link the three events of ATP hydrolysis in the NBD, substrate binding in the SBD, and association of Hsp40 with Hsp70 [78].

The affinity and kinetics of Hsp70 substrate binding depend on the nucleotide state of its NBD. In the ATP bound state, with open substrate binding domain, the association and dissociation rates between substrate and SBD are high, while the overall affinity is low [78] [85]. In the ADP bound state, the dissociation rate and association rate between substrate and SBD are slowed down by several orders of magnitude, leading to a 9-400-fold higher affinity of Hsp70 for substrate in the ADP state as compared to the ATP state [78].

The binding of ATP to the NBD induces a rotation of the two lobes relative to each other. This opens a crevice at the bottom of the NBD, which allows binding of the interdomain linker and association of the NBD with the SBD. The SBD pushes on the lobes of the NBD, forcing the catalytic center into a conformation unsuitable for ATP hydrolysis [78]. Accordingly, basal ATP hydrolysis rates of Hsp70 without substrate binding are low (~1 ATP molecule/6-40min [78]).

Substrate binding in the SBD induces the release of SBD- $\beta$  and SBD- $\alpha$  from the NBD through a rearrangement of intramolecular hydrophobic and polar contacts [84]. After the relaxation of the SBD clamp on the NBD, its lobes can rotate back into a conformation that is more suitable for ATP hydrolysis. However, the release of SBD from NBD induced by substrate binding also makes the interdomain linker slip out of the previously occupied crevice. The linker then no longer keeps the lobes of the NBD in a suitable conformation for ATP hydrolysis, which is the reason why substrate alone only leads to a moderate increase in ATP hydrolysis rate. For maximally accelerated hydrolysis rate, Hsp70 requires both substrate binding and the presence of an Hsp40 co-chaperone. Hsp40's J-domain inhibits the slipping of the interdomain linker, thereby allowing Hsp70 to reach the optimal conformation for ATP hydrolysis. The combination of substrate binding to Hsp70's ATP-bound open conformation and the association of a J-domain protein leads to an acceleration of ATP hydrolysis rate of up to 15000-fold above the basal rate. Hydrolysis causes the helical lid SBD- $\alpha$  to dock onto the substrate binding pocket of SBD- $\beta$ , prohibiting substrate dissociation.

This complex allosteric interdomain coupling allows Hsp70's switching from a fast substrate association rate in the ATP bound state to a slow substrate dissociation rate in the ADP bound state, triggered by substrate binding and stimulated by a JDP. As a consequence of this non-equilibrium process, the effective affinity of Hsp70 for substrate lies orders of magnitude above the substrate affinity of the ADP-bound closed conformation and was named "ultra-affinity" [85].

#### 1.6.5 Substrate Specificity of Hsp70s

Hsp70s promiscuously interact with a variety of substrate proteins, including unfolded polypeptides, folding intermediates, natively folded proteins, misfolded proteins and protein aggregates [81]. In studies with peptides, they transiently bind to motifs of seven amino acids in client proteins, where the five central amino acids are

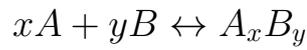
enriched in hydrophobic residues such as leucine, isoleucine, valine, phenylalanine and tyrosine, and the two flanking amino acids tend to be positively charged [86]. Such motifs frequently occur in all kinds of protein sequences, on average every 35-40 amino acids [86]. They are usually buried in the hydrophobic core of natively folded proteins. This explains why Hsp70s tend to bind to unfolded or misfolded proteins at multiple binding sites, since they exhibit exposed hydrophobic regions.

However, these studies with polypeptides are limited. While polypeptides can enter the substrate binding domain of Hsp70 in an extended conformation [87], the geometry of substrate binding can be potentially more complicated when Hsp70s bind to partially or natively folded proteins with secondary and tertiary structure. Also, Hsp40 co-chaperones play a crucial role in substrate interaction. J-Domain proteins (Hsp40s) have been shown to increase the number of Hsp70 binding sites [88].

### 1.7 Affinity

Binding affinity is a measure of the strength by which two (or more) molecules bind to each other. It is usually expressed by the equilibrium dissociation constant  $K_D$ , which has the unit of concentration. The smaller its value, the higher the affinity, and vice-versa.

Consider a chemical reaction where molecule  $A$  reacts with molecule  $B$  such that:



with  $x$  the number of molecules of type  $A$ ,  $y$  the number of molecules type  $B$ , and  $A_xB_y$  the complex built of  $x$  molecules of type  $A$  and  $y$  molecules of type  $B$ .

The binding affinity is then defined as:

$$K_D = \frac{[A]^x[B]^y}{[A_xB_y]}$$

with  $[A]$  the concentration of molecule  $A$ ,  $[B]$  the concentration of molecule  $B$  and  $[A_xB_y]$  the concentration of the complex  $A_xB_y$ .

In the frequently encountered case where only one molecule binds to one other molecule ( $x = y = 1$ ), the above formula for  $K_D$  reduces to:

$$K_D = \frac{[A][B]}{[AB]}$$

In this case,  $K_D$  has a simple biochemical interpretation: If  $[A] = K_D$ , then:

$$[A] = \frac{[A][B]}{[AB]}$$

$$1 = \frac{[B]}{[AB]}$$

$$[B] = [AB]$$

Which is equivalent to:

$$\frac{[AB]}{[B] + [AB]} = \frac{1}{2}$$

This means that, in the special case of  $x = y = 1$ ,  $K_D$  equals the concentration of molecule  $A$  at which half of the total number of molecules  $B$  are associated with molecule  $A$ .

In equilibrium, we also know:

$$[A] \cdot [B] \cdot k_{bind} = [AB] \cdot k_{diss}$$

with  $k_{bind}$  as the on-rate of molecules  $A$  and  $B$  and  $k_{diss}$  as the dissociation rate of the complex  $AB$ .

This immediately results in an additional interpretation of  $K_D$  as the ratio between dissociation rate and binding rate:

$$K_D = \frac{[A][B]}{[AB]} = \frac{k_{diss}}{k_{bind}}$$

In a single-molecule context such as in this work, we cannot define a "concentration" of the single molecule of interest, tethered in our dumbbell assay (cf. section 3.1).  $K_D$  then becomes not the concentration of molecule  $A$  at which 50 % of  $B$  molecules are bound to  $A$ , but rather the concentration at which molecule  $B$  is bound to molecule  $A$  50% of the time. This means that when  $[A] = K_D$ , the binding and dissociation rates are equal in equilibrium:  $k_{bind} = k_{diss}$

Determining the binding affinity between molecules  $A$  and  $B$  in single-molecule

experiments is possible by measuring the dissociation rate  $k_{diss}$  and binding rate  $k_{bind}$  of  $A$  to  $B$  at a certain concentration  $[A]$  and determining their ratio. While the binding rate of  $A$  to  $B$  depends on the concentration of free molecules of type  $A$  in solution, the dissociation rate does not.

## 2 Optical Trapping

### 2.1 History of Optical Trapping as an Experimental Method

Optical trapping as an experimental method to confine and manipulate particles dates back to the 1960s. Arthur Ashkin and others conducted pioneering work when they used the radiation pressure of an argon laser with Gaussian beam profile to trap and accelerate micron-sized transparent latex spheres in solution [89]. The principle is illustrated in Fig.7:

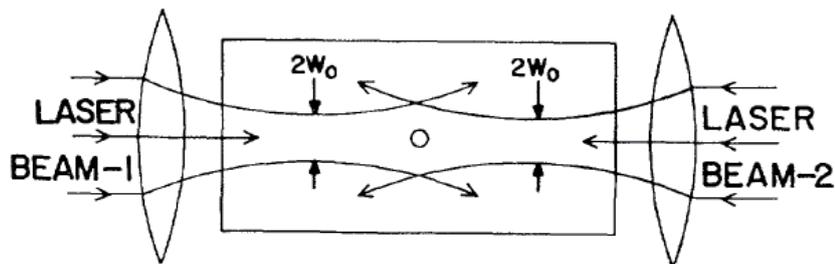


Figure 7: **First demonstration of optical trapping via radiation pressure:** Two focused laser beams of opposite direction form a stable optical well in which particles of high index of refraction can be trapped. Figure reprinted from [89] with permission. Copyright 1970 by APS physics.

Over the subsequent years, they further refined their technique, showing gravi-levitation of glass spheres [90] and trapping of individual atoms [91]:

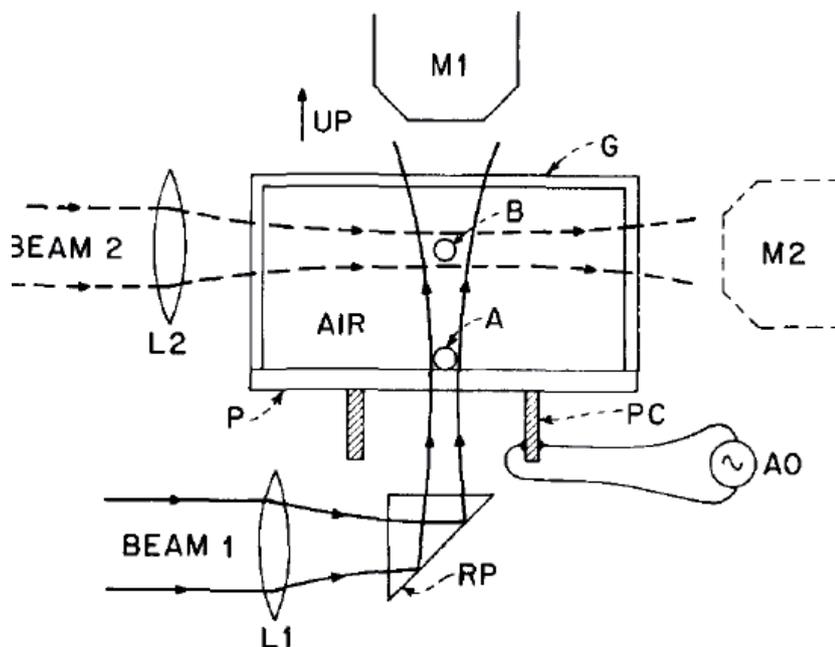


Figure 8: **Gravi-levitation of glass spheres:** Beam1 confines the glass spheres perpendicularly to the beam axis, while also pushing them up parallel to the beam axis. Gravity pulls the spheres down, thereby balancing out the upward radiation force. Figure reprinted from [90] with permission. Copyright 1971 by AIP.

In these early optical trapping experiments, the trapped object was drawn in perpendicularly towards the laser beam axis, but at the same time accelerated parallel to the laser beam axis in the direction of incident light by radiation forces. Confinement of the trapped object was therefore only possible perpendicular to the beam axis, but not parallel to it. In order to balance out the scattering force along the beam axis, either a second laser beam running in the opposite direction (such as in Fig.7) or gravity (such as in Fig.8) or electrostatic forces were necessary. Eventually, using a highly focused laser beam with Gaussian beam profile, stable trapping of small dielectric spheres in all three dimensions was achieved with just one laser beam [92]:

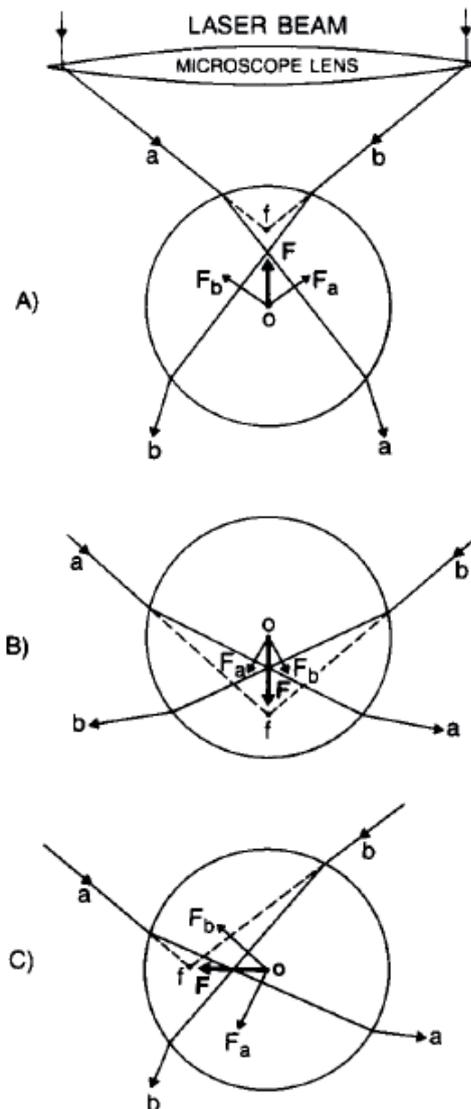


Figure 9: **Qualitative view of the optical trapping of a dielectric sphere using just one laser beam:** The sphere is confined both in axial as well as transverse direction from the beam. Displacement of the sphere in any direction out of the laser focus leads to a restoring force that pulls it back into the focus. The underlying physics will be discussed in section 2.2. Figure reprinted from [92] with permission. Copyright 1992 by Biophysical Journal.

Trapping in the direction parallel to the beam axis is possible as soon as the gradient force, which draws the trapped object towards the focus of the laser beam, surpasses the scattering force, which pushes the trapped object away from the focus in the direction of the incident light. This requires a high numerical aperture objective. Single-beam optical trapping from then on was the simplest and most versatile modus both conceptually and practically.

The possibility of trapping objects with a single laser beam in three dimensions further increased the range of possible applications of this method. Soon, trapping of biological samples such as single tobacco viruses as well as live, motile *E.coli* bacteria was demonstrated [93]. Later on, optical trapping was also used to apply forces to single biomolecules and study their mechanical properties. A remarkable range of optical trapping experiments have been conducted with all kinds of biological samples, including DNA [94], RNA polymerase [95], molecular motors such as kinesin and myosin [96] [97] and single proteins [98] [99].

The next generation of optical tweezers used the combination of two optical traps. In this configuration, both traps are produced by the same laser beam and separated according to their polarizations. In these traps, dielectric spheres ("beads") can be conveniently trapped and manipulated. Single molecules can then be chemically tethered between these beads. By moving one laser beam while holding the other fixed, it is possible to stretch and relax the tethered molecule or expose it to a predefined force bias. This approach is called a "dumbbell assay" and was also used in this work. It is illustrated in Fig.10. The experimental details will be discussed in section 3.1.

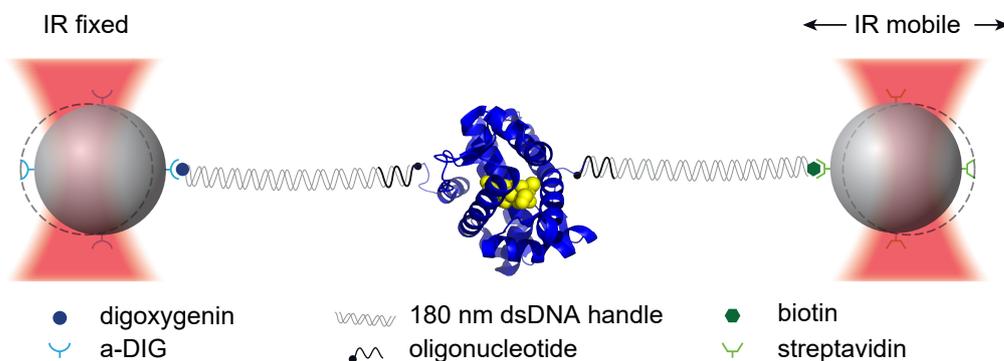


Figure 10: **Dumbbell Assay:** A single protein is tethered between two silica beads, which can be trapped and manipulated using highly focussed infrared laser beams. One of the traps is mobile, while the other one is fixed. This allows exerting force on the tethered molecule.

A dumbbell assay mechanically decouples the measurement from the rest of the system and thereby largely reduces noise and drift. Through differential detection of the beads, any perturbation or fluctuation concerning the laser beam affects both traps and can be eliminated [100]. The forces exerted by optical traps lie in the low picoNewton range and thereby count among the smallest possible forces that can be applied to samples in a controlled experimental environment. As a result, with this dumbbell approach, the achievable high resolution allowed experiments investigating for example complex protein folding networks [101] and subnanometre enzyme mechanics [102].

In 2018, Arthur Ashkin shared the Nobel Prize in physics for “optical tweezers and their application to biological systems” [103].

## 2.2 The Physics Underlying Optical Trapping

Optical trapping requires the interaction of light with dielectric objects via radiation pressure. It can be divided into different regimes depending on the size of the trapped object relative to the wavelength of the incident light.

### 2.2.1 Mie Regime

When the trapped object is much larger than the wavelength of the incident light ( $d \gg \lambda$ ), this is called the “Mie regime”. It reduces the theoretical description of optical trapping to geometric ray optics [92] [104]. It is in this limit that the principle of optical trapping can be understood most easily and intuitively. In Fig.11, a dielectric sphere is shown in two different positions relative to a laser beam with Gaussian profile:

- A: Laterally displaced from the beam axis of a parallel beam,
- B: Centered on the beam axis of a focussed beam, below the focus.

Two exemplary rays, ray1 and ray2, are depicted in each case. Both rays are refracted at the surface of the sphere when they enter and when they exit it. Photons carry momentum according to the formula  $p = \frac{h}{\lambda}$ , with  $h$  being the Planck constant and  $\lambda$  the wavelength of light. Therefore, if a ray changes its direction due to refraction, a change in momentum occurs. According to the law of conservation of momentum, the system’s total momentum must be conserved. Therefore, if a ray changes its momentum upon refraction on a dielectric sphere, the sphere must undergo an equal and opposite change in momentum. By Newton’s Second Law, the rate of change of momentum produces a force. Looking at Fig.11, this simple consideration about momentum conservation suffices to understand optical trapping of the sphere in three dimensions:

In Fig.11A, the sphere is displaced laterally from the beam axis. Since the beam has a Gaussian intensity profile, ray2, being closer to the center of the beam, has higher intensity and carries more momentum than ray1 (therefore, ray2 is drawn thicker than ray1). Both beams are refracted on both surfaces of the sphere (red paths). The green arrows show the resulting equal and opposite change in momentum that the sphere experiences due to each beam. At the bottom, the vector sum of the changes in momentum due to both rays is depicted. The net change in momentum that the sphere will experience is directed to the right and slightly downward, thereby driving the sphere into the center of the beam, i.e. confining it laterally on the beam axis. Whenever the sphere is displaced laterally from the beam axis, it will be exposed to this restoring force that draws it back into the center of the beam. Only when the center of the sphere is aligned with the beam axis, the forces resulting from the refraction of ray 1 and ray 2 will cancel out symmetrically, resulting in no net lateral force on the sphere.

In Fig.11B the sphere is centered on the beam axis, but positioned below the focus of the laser beam. Again, two illustrative light rays are displayed. With the sphere positioned symmetrically with respect to the beam axis, this time the two rays have equal intensity. The refractions of ray1 and ray2 on the surfaces of the

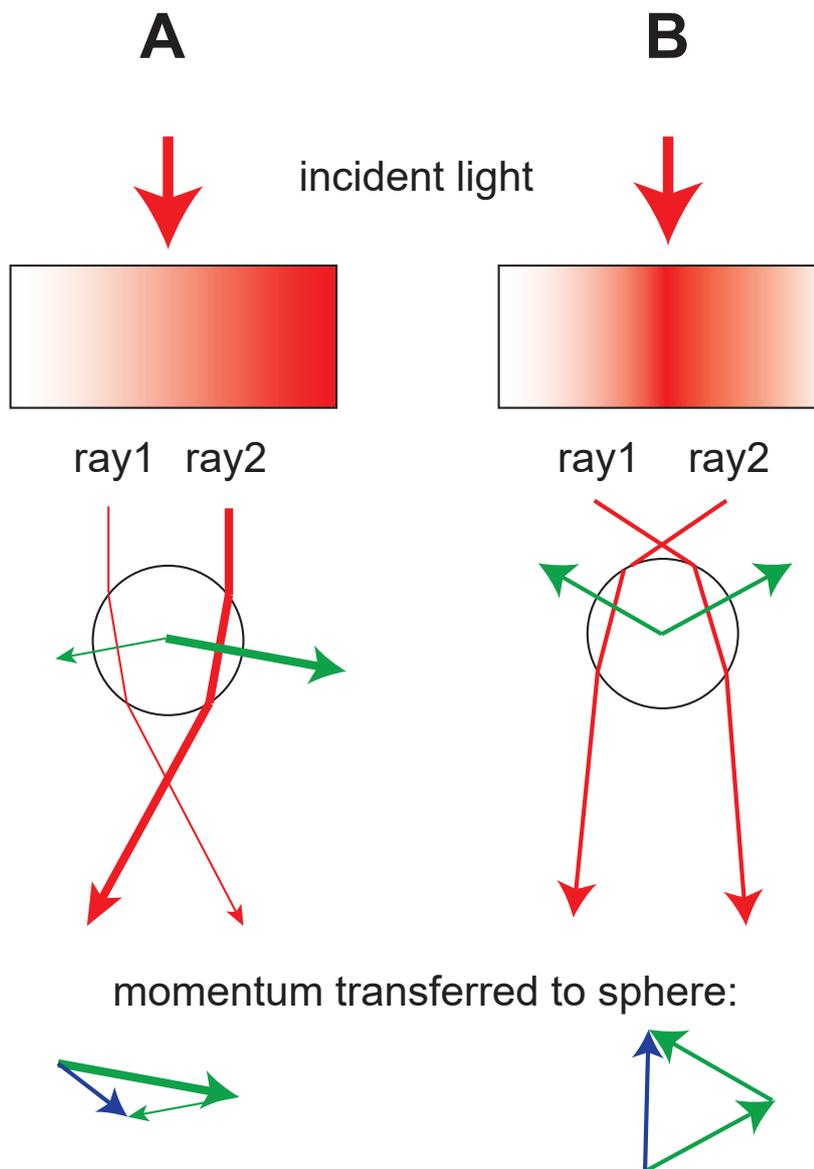


Figure 11: **Geometric optics picture of the gradient force:** **A)** A parallel beam of light with a gradient in intensity (darker color indicates higher intensity) shines on a transparent sphere. The sphere is transversely displaced from the beam center. Both exemplary rays transmit momentum to the sphere, when they are refracted at its surfaces. Since ray2 has a higher intensity than ray1, the resulting force on the sphere is directed towards the center of the beam. **B)** Now the beam of incident light is focussed. The sphere is located on the beam axis of highest intensity, but below the focus. The two rays are now equal in intensity. Their lateral components cancel each other out, leaving only a net transfer of momentum on the sphere directed upwards to the beam focus.

sphere lead to a change in momentum parallel to the direction of the incident light, with the lateral components of the two rays cancelling each other out. Hence, the sphere experiences a change in momentum opposite to the propagation of the light beam, which pushes it into the focus of the laser beam.

A third possible case where the sphere is situated on the beam axis but above the focus of the laser beam works in analogy to Fig.11B. Again, the sphere will be pushed into the focus of the laser beam.

Due to the radiation pressure in the direction of incident light, the actual location of stable trapping of the sphere is slightly below the focus of the laser beam. At this position, the restoring gradient force and the force due to radiation pressure cancel each other out.

### 2.2.2 Rayleigh Regime

If the trapped object is much smaller than the wavelength of incident light ( $d \ll \lambda$ , with  $d$  the diameter of the trapped object and  $\lambda$  the wavelength of light), this is called the ‘‘Rayleigh’’ regime. In this limit, the electric field can be approximated as uniform across the dielectric object, which allows the trapped object to be treated as an induced point dipole [104]. The forces acting on the dielectric again decompose into two components, namely the scattering force and the gradient force.

The scattering force arises due to the radiation pressure on the particle. It is directed along the propagation direction of the incident light and for a sphere of radius  $a$  is given by [105]:

$$F_{scatt} = \frac{I_0 \sigma n_m}{c}$$

with  $I_0$  the intensity of the incident light,  $\sigma$  the particle’s scattering cross-section,  $n_m$  the refraction index of the surrounding medium and  $c$  the speed of light. In the case of a spherical particle,  $\sigma$  is given by [105]:

$$\sigma = \frac{128\pi^5 a^6}{3\lambda^4} \left( \frac{m^2 - 1}{m^2 + 2} \right)^2$$

with  $a$  the radius of the trapped spherical object,  $\lambda$  the wavelength of the trapping laser, and  $m$  the ratio of the index of refraction of the particle to the index of the medium ( $\frac{n_p}{n_m}$ ).

The gradient force is the Lorentz force that acts on the induced dipole. It is given by [105]:

$$F_{grad} = \frac{2\pi\alpha}{cn_m^2} \nabla I_0$$

where  $\alpha$  is the polarizability of the spherical particle,  $c$  the speed of light,  $n_m$  the refraction index of the surrounding medium and  $I_0$  the intensity of the incident light.  $\alpha$  can be calculated using the expression [105]:

$$\alpha = n_m^2 a^3 \left( \frac{m^2 - 1}{m^2 + 2} \right)$$

where  $n_m$  the refraction index of the surrounding medium,  $a$  is the radius of the trapped sphere and  $m$  the ratio of the index of refraction of the particle to the index of the medium ( $\frac{n_p}{n_m}$ ).

The gradient force is proportional to and directed parallel to the gradient in energy density. The requirement for optical trapping with a single laser beam is that in the z-direction (direction of incident light), the gradient force is larger than the scattering force. Increasing the numerical aperture decreases the size of the focal spot and thereby increases the gradient strength. Therefore, in the Rayleigh regime, trapping forces in all directions increase with greater numerical aperture [104].

The effect of the scattering force is that it displaces the particle slightly out of the focus of the laser beam in the direction of incident light.

The peripheral rays in the incident light beam contribute disproportionately to the axial gradient force, while the central rays contribute mostly to the scattering force. This can be understood intuitively, considering that a light ray hitting the spherical bead in a perfectly centered (radial) position is not deflected at all, while a peripheral light ray hitting the surface of the sphere in a flat angle is deflected strongly. Therefore, slightly overfilling the back aperture of the objective such that the  $\frac{1}{e^2}$  intensity waist of the Gaussian beam matches the width of the aperture leads to improved trapping efficiency. Another way to improve axial trapping efficiency would be to use the TEM<sub>01</sub> mode of a laser, which is "donut"-shaped.

### 2.2.3 In Between the Mie and the Rayleigh Regime

The Mie and the Rayleigh regime cover the two extreme cases where the trapped particle is either much bigger ( $d \gg \lambda$ ) or much smaller ( $d \ll \lambda$ ) than the wavelength of the incident light.

However, in many experiments, including all experiments in this work, the scale of the trapped bead and the wavelength of the laser light used are about equal ( $d \approx \lambda$ ). In our experiments, the trapped beads have a diameter of 1  $\mu\text{m}$ , while the infrared laser has a wavelength of  $\lambda = 1064 \text{ nm}$ . In this region in between regimes, trapping still works, but the theoretical description becomes much more complicated [106] [107].

## 2.3 The Optical Trap Setup Used in this Work

The custom-built optical trap setup predominantly used in this work was designed by Benjamin Pelz [108]. A schematic overview of the setup is given in Fig.12:

The setup has been described in detail before [109] [110]. A brief summary is given here. The infrared laser is a Nd:YVO<sub>4</sub> solid state laser with a wavelength of  $\lambda = 1064 \text{ nm}$ . Its TEM<sub>00</sub> mode is used, yielding a Gaussian beam profile. A Faraday

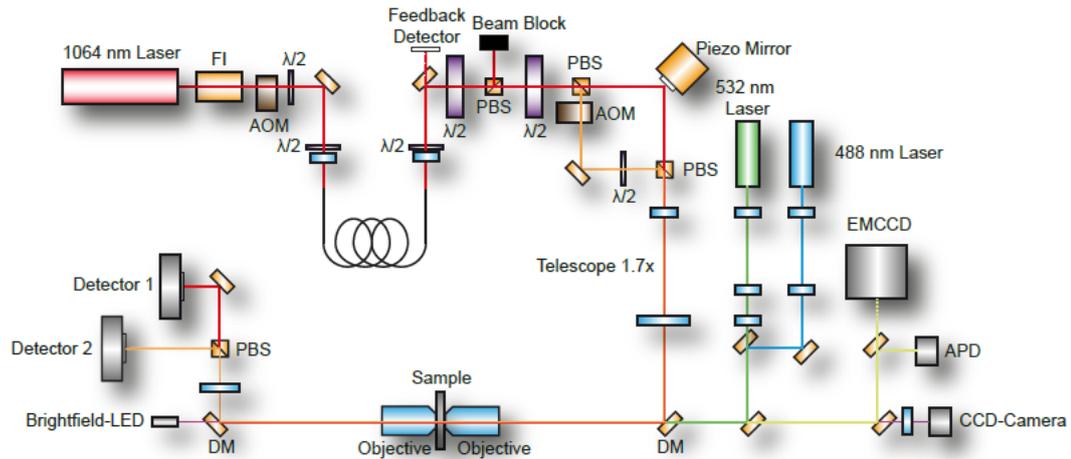


Figure 12: The "Ducktrap" optical tweezers setup used in this work [109] [110].

insulator (FI) prevents harmful back-reflections from optical elements in the beam path into the laser. The possibility of a feedback loop using an Acousto-Optical Modulator (AOM) and a feedback detector is built into the setup, but was not used in this work. A motorized  $\lambda/2$ -waveplate and a polarization beam splitter (PBS) are installed to adjust the overall power of the laser beam by disposing the desired amount of laser light into the beam block. Another beam splitter divides the laser beam into two beams of orthogonal polarization. The motorized  $\lambda/2$ -waveplate before this beamsplitter gives the possibility to adjust the relative power between the two traps. The beam creating the fixed trap passes through an AOM, which induces a slight frequency shift in the light of the fixed trap, preventing interference effects between the two beams later on in the setup. The piezo mirror in the beampath of the mobile trap is used to steer the laser beam of the mobile trap in the x-direction. Another polarizing beam splitter unites the two traps again into one beampath. A telescope is used to expand the beam in order to overfill the back aperture of the objective. Via a dichroic mirror, the two laser beams pass through an objective of high numerical aperture ( $NA = 1.27$ ), the sample, and the condenser, which is identical to the objective. Via another polarization beam splitter, the two laser beams are again split according to their polarization and end up in two photo-sensitive detectors, one for each trap. The back-focal plane of the condenser must be focused on the detectors. Light from a brightfield LED travels along the beampath in the direction opposite to the infrared laser light and is observed in the CCD camera. This brightfield illumination makes the two dichroic mirrors necessary and allows to see the micrometer-sized trapped beads. The possibility to include two additional laser beams for fluorescence experiments was included in the setup, but not used in this work.

### 3 Experimental Approach

This section provides a detailed description of the experimental approach used for single-molecule force spectroscopy in this study. Each step in the biochemical assembly of the dumbbell assay construct as well as the measurement modes at the optical trap are explained.

#### 3.1 Stepwise Assembly of the Dumbbell Assay

Fig.13 indicates all the steps necessary for the sequential assembly of the dumbbell assay:

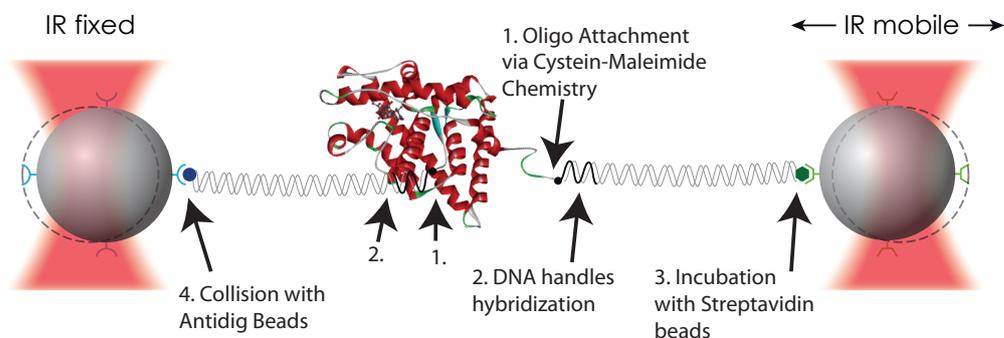


Figure 13: **Stepwise assembly of the dumbbell -assay:** 1. DNA oligo attachment to terminal cysteines via maleimide chemistry. 2. DNA handles hybridization to oligos. 3. Incubation with streptavidin beads. 4. Bringing the two optical traps into close proximity (referred to as "ditching" in our group) allows the attachment of the construct to the anti-digoxigenin beads.

The assembly proceeds from the center towards the edges, i.e. starting with the protein of interest in the middle and ending with the beads.

The protein is genetically modified such that it exhibits cysteines at its N- and C-terminus. Cysteines allow the use of maleimide chemistry to attach short maleimide-modified DNA oligos to the termini of the protein (section 3.1.1). The single-stranded DNA oligos connected to the termini of the protein can then hybridize via base pairing with DNA handles of 180 nm length (section 12.4). The DNA handles are double-stranded except of a short single-stranded stretch complementary to the sequence of the DNA oligos at one end. At their other end, they have either a biotin or digoxigenin attached. The construct is then incubated with streptavidin beads (section 3.1.3), such that the biotin-modified DNA handles can attach to the beads. In the last step, during the measurement at the optical trap, the two traps are used to bring a streptavidin bead and an anti-digoxigenin bead into close proximity, such that the digoxigenin and the anti-digoxigenin can form a bond and the dumbbell assay is completed (section 3.1.4).

The following sections explain the biochemistry of each of these individual steps of the assembly of this construct in more detail. Protocols concerning the experimental procedures are given in the SI Methods (section 12).

### 3.1.1 Oligo Attachment via Cysteine-Maleimide Chemistry

The first step in the assembly of the dumbbell assay is the attachment of DNA oligos to the terminal cysteines of the protein via maleimide-cysteine chemistry.

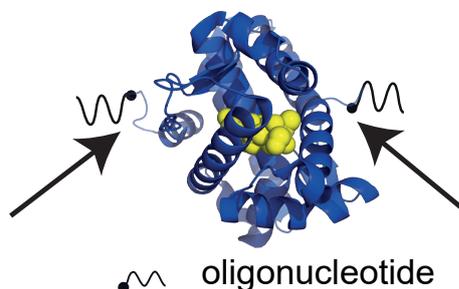


Figure 14: Oligo attachment via maleimide-cysteine chemistry.

The connection between the cysteine-modified termini of the GR-LBD and the DNA oligos is established using maleimide chemistry [111]. The thiol group of a cysteine can react with the maleimide as depicted in Fig.15.

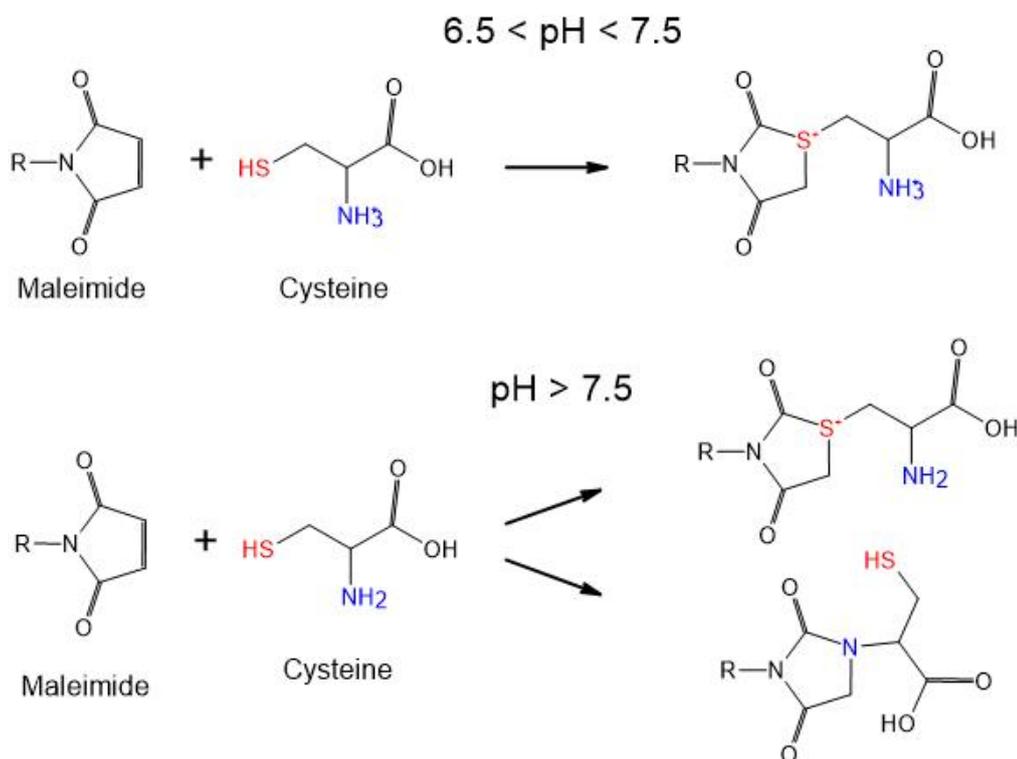


Figure 15: **Maleimide-cysteine reaction:** The thiol group (SH-, depicted in red) of a cysteine reacts with the maleimide and forms a stable connection. The reaction only works efficiently at a pH value between 6.5 and 7.5. In this pH range, the reaction rate of the maleimide with the thiol is about 1000 times faster than the reaction rate of the maleimide with the amide ( $\text{NH}_3^+$ ) depicted in blue). At a pH value above 7.5, the amine changes to  $\text{NH}_2$  and competes with the thiol group for binding to the maleimide.

The reaction works best at a pH between 6.5 and 7.5. If the pH is above 7.5, the maleimide can also bind to other sites than the thiol group of the cysteine, leading to unspecific attachment.

Depending on the experimental boundary conditions imposed by the respective protein, we used one of two different approaches for oligo attachment: in the first approach, the DNA oligos carry a maleimide at their end and are brought to react with the terminal cysteines of the GR-LBD construct (cf. SI section 12.1). In the alternative approach, azide-modified oligos were used. The connection between DNA oligo and cysteine was then created via a DBCO-maleimide linker (cf. SI section 12.2).

### 3.1.2 DNA Handles Hybridization

Before each measurement, the construct consisting of protein and oligos is incubated with the DNA handles for 1 hour to allow hybridization of handles and oligos.

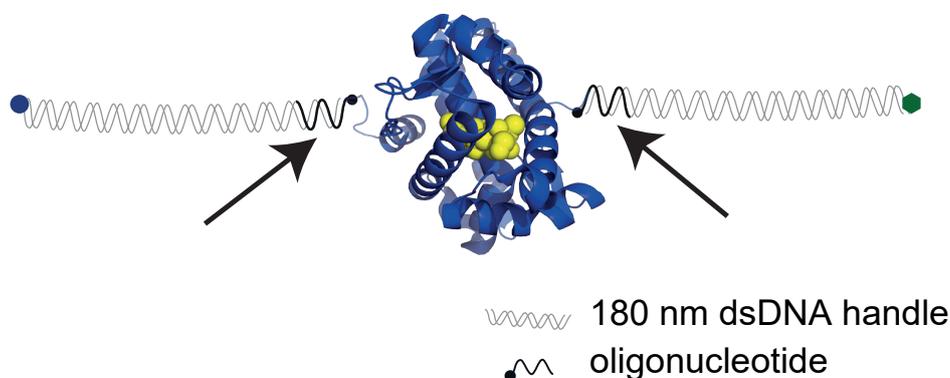


Figure 16: DNA handle hybridization to oligos

It is crucial to find a good ratio of handles to protein in this step. Adding too small an amount of DNA handles results in the majority of proteins only carrying one DNA handle, which is of no use to form a dumbbell-assay. Adding too large an amount of DNA handles will result in a large excess of non-reactive DNA handles, which during the incubation with beads will block too many binding sites on the beads and thereby lead to too low a density of protein constructs on the surface of the beads.

The experimental details of both the DNA handles incubation as well as the agarose gel control, are explained in SI section 12.4.

### 3.1.3 Incubation of the Protein-Oligo-Handle Construct with Streptavidin Beads

Directly before each measurement, the construct now consisting of protein, DNA oligos and DNA handles is incubated with streptavidin-coated beads for 20 minutes.

The reason why it is advantageous to use the streptavidin beads and not the anti-digoxigenin beads in this reaction lies in the nature of the biotin-streptavidin and dig-anti-digoxigenin bonds: The biotin-streptavidin bond is very stable once formed, but does not connect well under laser radiation [110]. The dig-anti-digoxigenin bond

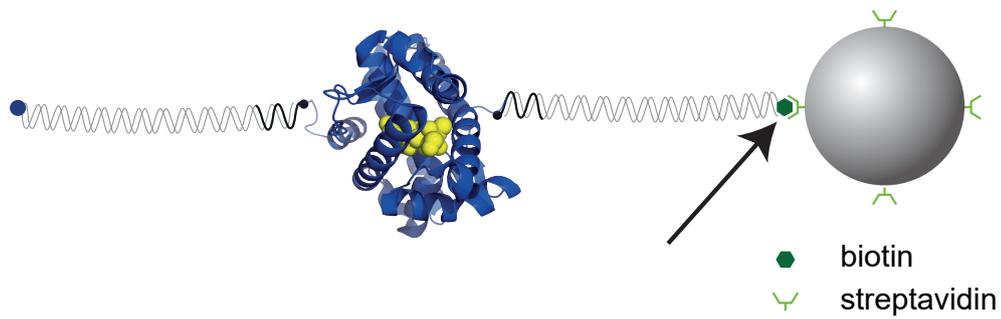


Figure 17: **Incubation with streptavidin-coated beads allows the formation of the Biotin-streptavidin bond between handles and beads.**

forms well under incident laser radiation. Therefore, stable tethering is more likely, if the biotin-streptavidin bond is established before the beads are trapped by laser beams.

### 3.1.4 Assembly of the Full Dumbbell Assay by Collision with Anti-Digoxigenin Beads

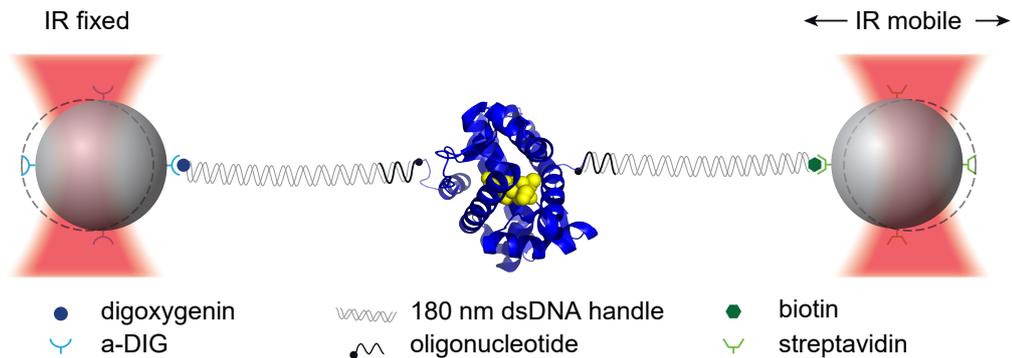


Figure 18: **Formation of the anti-dig-digoxigenin bond through the collision of the beads in the optical traps.**

The last step to reach the full dumbbell assembly is the addition of the anti-dig beads to the construct consisting of protein, DNA oligos, DNA handles and streptavidin beads. To this end, the beads need to be distinguishable during the measurement at the optical trap in order to pick one of each kind in the two traps. Therefore, the anti-dig beads are labelled fluorescently using Rhodamin and the streptavidin beads are non-fluorescent. A 488 nm laser is coupled into the same optical path as the infrared laser. This way, the two kinds of beads can be distinguished during the measurement through simultaneous imaging via the brightfield camera and the fluorescence camera. A streptavidin bead is then caught in one trap and an anti-dig bead in the other trap. The two traps are repeatedly brought into close proximity ("ditching") and moved away from each other until the formation of a tether becomes apparent through the typical DNA-protein wormlike-chain fingerprint (cf. section 4.1).

### 3.2 Measurement Modes at the Optical Trap

Optical trapping experiments conducted in this study can be loosely divided into two categories of measurement mode: "stretch-relax" experiments (in our research group often referred to as "constant velocity" experiments) and "passive mode" experiments (in our research group often referred to as "constant distance" experiments).

These two basic measurement modes are explained in more detail in the following sections.

#### 3.2.1 Stretch-Relax Experiments

In "stretch-relax" experiments, the position of one optical trap is held fixed, while the other optical trap is moving back and forth, thus continuously stretching and relaxing the tethered molecule. Typical speeds at which the mobile laser is moved are on the order of  $500 \text{ nm s}^{-1}$ . In this study, experiments were conducted at a range of speeds between  $2 \text{ nm s}^{-1}$  and  $5000 \text{ nm s}^{-1}$ .

The stretch-relax measurement mode results in traces such as the one shown in Fig.19.

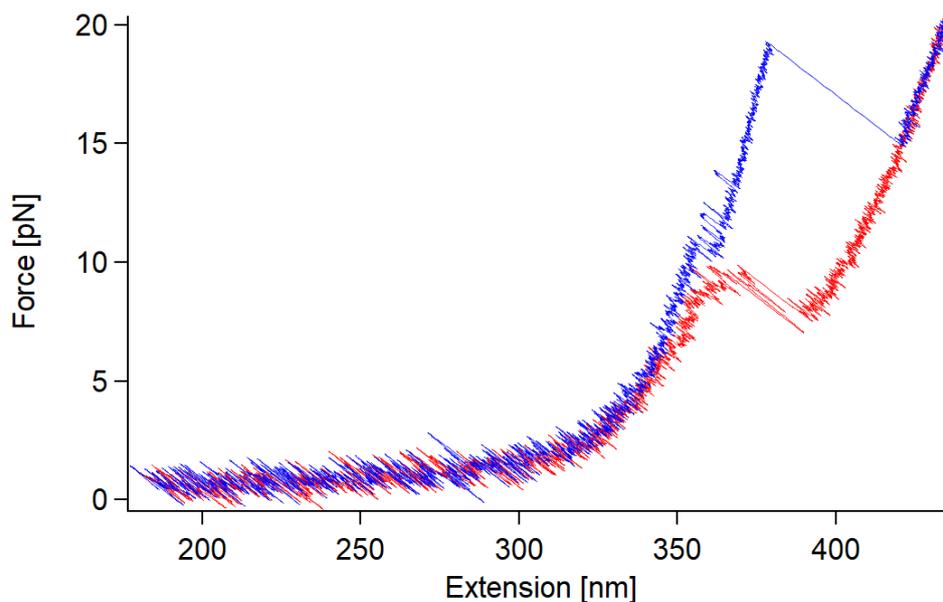


Figure 19: A typical stretch-relax cycle of GR-LBD.

These force-extension traces provide a unique unfolding "fingerprint" of each molecule. They are useful to identify whether a molecule was tethered correctly, but also provide information about contour length changes, unfolding and folding intermediates and more.

Note that the extension is not the distance between the two optical traps. Since the optical traps act like Hookean springs, when the tether is under tension, the two beads are displaced from the centers of the traps in proportion to the applied force. This deflection of the beads must be taken into account when determining the true extension of the molecule. Therefore, the extension is actually the distance between the traps minus the deflection of the beads out of the centers of the traps, i.e.:  $\text{extension} = \text{distance} - \text{deflection}$ .

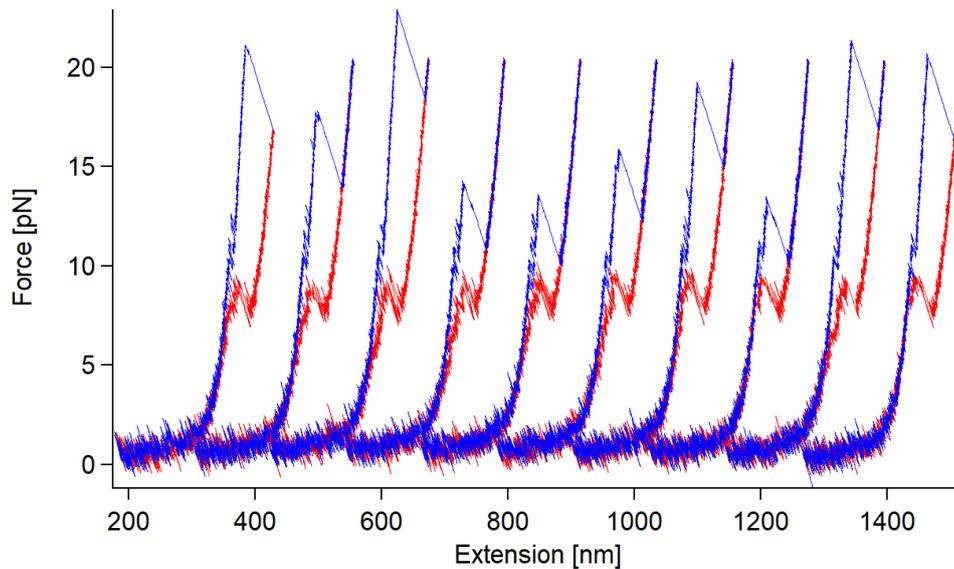


Figure 20: 10 consecutive stretch-relax cycles of the same molecule to illustrate the reproducibility of the protein "fingerprint" and the variability in unfolding force (highest peak of the trace).

### 3.2.2 Passive-Mode Experiments

In "passive mode" experiments, the positions of both optical traps are held at a fixed distance. The tethered molecule is thereby subjected to a predefined force bias set by this distance between the optical traps. The force will drop upon unfolding-induced lengthening of the molecule. Typically, a force range is chosen in which particular events of interest are induced, e.g. unfoldings or ligand dissociation. For the GR-LBD, the force range of interest was relatively narrow between 8 pN and 12 pN. Fig.21 shows a passive-mode trace with the characteristic flipping of GR-LBD indicative of the native holo state, multiple ligand dissociations and rebindings, and complete unfoldings, all at a predefined force of 11.5 pN in the native state.

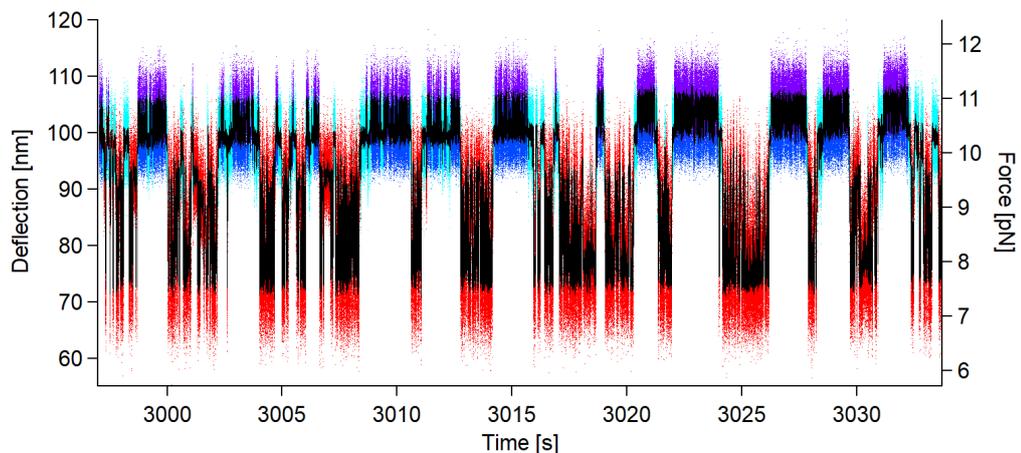


Figure 21: An example of a passive mode trace of GR-LBD.

## 4 Data Analysis

To analyse the raw data acquired at the optical trap setup, several mathematical models and techniques were applied. The following sections explain the most frequently used ones among them.

### 4.1 Wormlike-Chain Model

The Wormlike-Chain (WLC) model describes the force-extension relation of a polymer upon stretching. Extending the ends of a linear flexible polymer such as DNA leads to a reduction of the conformational entropy of the chain and therefore requires the application of a force.

Since exact solutions to the problem are difficult to obtain, Marko and Siggia worked out the following interpolation formula to describe the force-extension relation of a stretched polymer [112]:

$$F_{WLC}(x) = \frac{k_B T}{p} \left( \frac{1}{4(1 - \frac{x}{L})^2} - \frac{1}{4} + \frac{x}{L} \right)$$

Here,  $F$  is the applied force,  $x$  is the extension of the polymer,  $k_B$  is Boltzmann's constant,  $p$  is the chain's persistence length and  $L$  is the chain's contour length.

In the case of DNA molecules the elasticity deviates from a purely entropic WLC model at high forces due to enthalpic effects. To account for this deviation, an additional term involving a stretch modulus  $K$  can be included in the WLC model. This is called the extended Wormlike-Chain model (eWLC) [113]:

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

Throughout this thesis, stretch-relax traces of GR-LBD involved the stretching of a dumbbell assay (cf. section 3.1) consisting of two different types of biopolymer, namely DNA and unfolded peptide chain. Hence, to fit these traces, a combination of an eWLC model, describing the stretching of the DNA only, and a standard WLC model, describing the stretching of the unfolded peptide chain, was used:

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

$$F_{WLC\ prot}(x) = \frac{k_B T}{p_{prot}} \left( \frac{1}{4(1 - \frac{x}{L_{prot}})^2} - \frac{1}{4} + \frac{x}{L_{prot}} \right)$$

Again,  $F$  is the applied force,  $x$  is the extension of the polymer,  $k_B$  is Boltzmann's constant,  $p$  is the chain's persistence length and  $L$  is the chain's contour length.

Fig.22 shows the signature unfolding fingerprint of the GR-LBD along with WLC model fits.

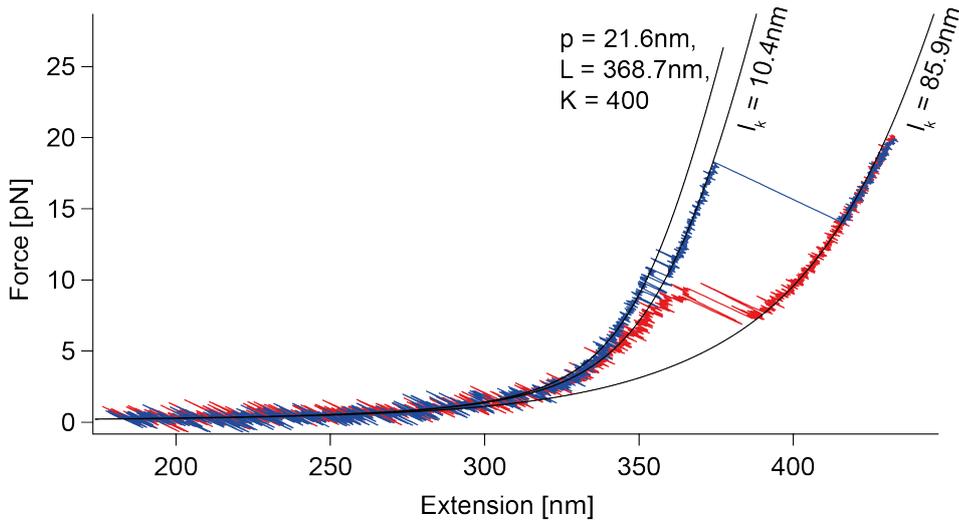


Figure 22: Stretch-relax cycle of the GR-LBD at 500nm/s with WLC fits.

WLC-fits such as the one above in Fig.22 are useful in many ways. Several quantities of interest can be extracted. Those include the persistence length of DNA, the overall length of the dumbbell construct, the elastic stretch modulus, the contour length of peptide chain unfolding events as well as the overall length of the unfolded peptide chain of the protein. The fits shown above rendered a DNA persistence length of  $p = 21.6$  nm, an overall length of the dumbbell construct of  $L = 368.7$  nm, and an elastic stretch modulus of  $K = 400$ , while the first unfolding step of the GR-LBD has an unfolding contour length of  $l_{\text{contour}} = 10.4$  nm and the overall length of the entire GR-LBD's peptide chain is determined to be  $l_{\text{contour total}} = 85.9$  nm.

Therefore, WLC fits can serve as a sanity check of the entire assay by comparing the values extracted from the fits to the expected literature values.

In addition, the values from WLC fits are mandatory to transform the measured forces into the equivalent contour lengths of the molecule.

## 4.2 Hidden Markov Models

A Hidden Markov Model (HMM) provides a mathematical tool to infer the true hidden state of a system from a measured observable that is affected by noise.

HMMs were developed in the 1960s and 1970s. At that time, they were mostly used for signal processing, in particular speech recognition [114].

Later they were applied to all kinds of single-molecule data [115] [116].

The application of HMMs is based on the assumption that an observable undergoes a memory-less continuous time Markov process. "Hidden" in this case means that the experimenter has no way to directly and with absolute certainty determine the state of the system at any given moment.

Single-molecule experiments typically record the change of some kind of observable over time. In single-molecule fluorescence experiments, it could be the photon count. In single-molecule force spectroscopy, as in this work, it is a time series of force or extension values.

If the molecule of interest can be in different states, this is reflected in the measured observable. In an ideal experiment without noise, the observable would directly and unambiguously indicate the state of the molecule. However, experimental reality

usually brings about instrumental and thermal noise, which render the interpretation of the data more challenging.

The experiment renders a time series of measured values of some observable. The challenge is to extract the most likely time series of hidden molecular states that underly and produce the measured time series of observable values. That is to say, the result of the HMM analysis will be the most likely series of snapshots of the actual continuous-time molecular dynamics for each point in time at which a data point of the observable was recorded.

Fig.23 illustrates the situation with a tangible example: The molecule can be in one of two states, folded (blue) or unfolded (red) (Fig.23A). On a discrete time scale, the molecule can either remain in the same state (horizontal arrows) or undergo a transition to the other state (diagonal arrows) between two time points. Each of these events happens with a certain transition probability. In the given example,  $P_{FF}$  is the probability that the molecule remains in the folded state,  $P_{UU}$  the probability that the molecule remains in the unfolded state,  $P_{FU}$  that the molecule undergoes a transition from the folded to the unfolded state and  $P_{UF}$  the probability that the molecule goes from the unfolded to the folded state. These probabilities are usually summarized in a transition matrix  $T_{ij}$ .

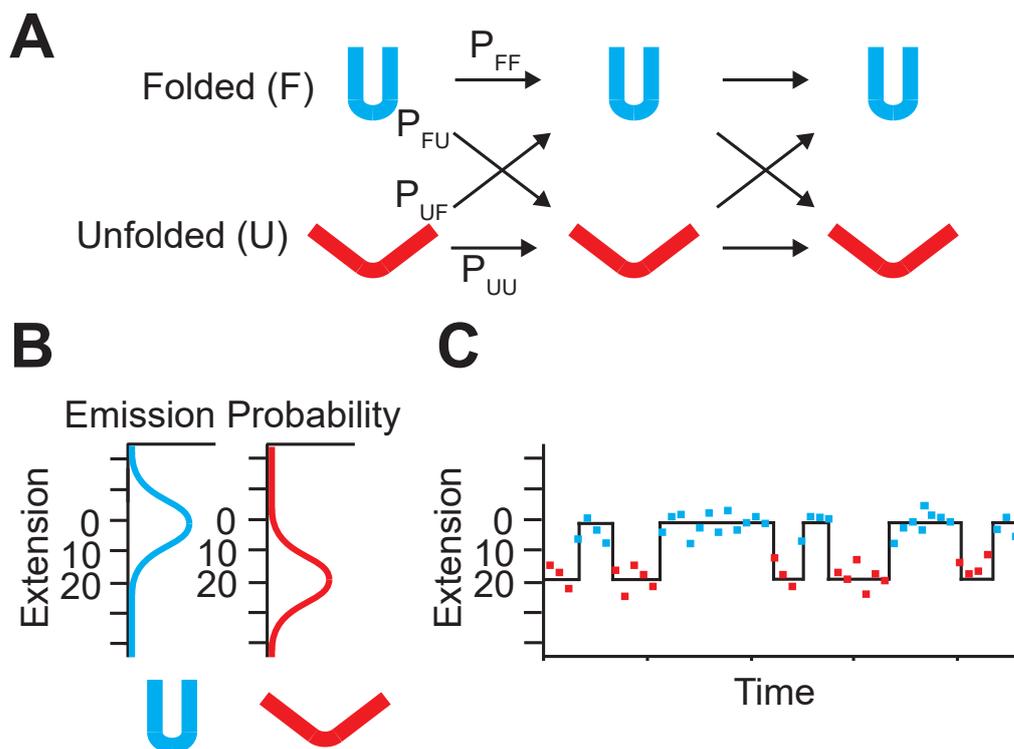


Figure 23: **Illustration of a Markov process with hidden molecular states producing a time series of data points with certain emission probabilities:** **A)** In this scenario, the molecule can be in two states, either folded (blue) or unfolded (red). Between discrete points in time, the molecule can either remain in its previous state or undergo a transition to the other state. **B)** Each state of the molecule produces measurement values with certain emission probabilities. Therefore, the measured value is not a definitive indicator of the "hidden" true state of the molecule. **C)** An example trace of data points assigned to the different states of the molecule.

In an ideal experiment the measured extension would be an unambiguous indicator of the state of the molecule, e.g. 0 nm whenever the molecule is folded and e.g. 20 nm when it is unfolded. However, as a consequence of the noise in the data, each state has a certain "emission probability"  $E_i(x)$  (Fig.23B), which is often, at least for initialization, assumed to be an approximately Gaussian distribution around the true value. If the extension were measured to be e.g. 10 nm at a certain instant, it would be unclear whether the molecule is in the folded or unfolded state at that moment, since both the folded as well as the unfolded state produce such an extension with a certain finite probability. Also, in more complicated scenarios, the assignment between states and observable might be degenerate in the sense that more than one state could project on the same value of the observable.

Taken together, the transition matrix  $T_{ij}$  and the emission probabilities  $E_i(x)$  are called the "model parameters". Fig.23C gives a conceivable example trace, illustrating the challenge to assign each data point to a certain state of the molecule.

Using algorithms based on statistics, such as the Viterbi algorithm [117] [118], the Forward-Backward algorithm [119] [114] and the Baum-Welch algorithm [120], the model then infers the most likely series of true hidden states of the molecule from the trajectory of data points of the observable at each point in time.

### 4.3 Extracting Equilibrium Free Energies from Passive Mode Traces

After a successful HMM analysis of a passive-mode trace, each data point is assigned a particular molecular state. One can then calculate the population probability of each state by summing up the dwell times in each state and dividing by the total measurement time. Let  $p_j$  be the probability for the molecule to be in state  $j$ . The population probabilities of two states  $i$  and  $j$  are linked to the free energy difference between these two states via the Boltzmann relation:

$$\frac{p_j(F_j)}{p_i(F_i)} = \exp\left(-\frac{\Delta G_{ij}(F_i, F_j)}{k_B T}\right)$$

with  $p_i(F_i)$  the probability of the system to be in state  $i$  at the force  $F_i$  and  $\Delta G_{ij} = G_j(F_j) - G_i(F_i)$  the free energy difference between state  $i$  and state  $j$ .

The free energy  $G_i(F_i)$  associated with a certain state  $i$  under force is the sum of the energy stored in the folded protein conformation  $G_i^0$ , which we want to extract, and the force-dependent energy stored in the system consisting of beads, DNA and unfolded peptide:

$$G_i(F_i) = G_i^0 + G_i^{beads}(F_i) + G_i^{DNA}(F_i) + G_i^{unfolded\ protein}(F_i)$$

$G_i^{beads}$  is the energy of the bead displacement in the harmonic trap potential:

$$G_i^{bead} = \frac{1}{2}x(F_i)F_i$$

$G^{DNA}(F)$  can be calculated by integration over the eWLC curve of DNA:

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

$G^{un\ folded\ protein}(F)$  can be calculated by intergration over the WLC curve of protein:

$$G^{un\ folded\ protein}(F) = \int_0^{x_{WLC}(F)} F_{WLC}(F)(x') dx'$$

As a result, by using the Boltzmann relation, the probability to find the system in state  $i$  is given by:

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

#### 4.4 Transition Rates

Once an HMM has attributed a certain state to each data point, it is possible to extract the transition rates between the different states. There are multiple ways to go about this. The most robust and straight-forward way works as follows: First, the off-rate from a certain state is determined by calculating the average dwell time in this state and then taking its reciprocal:

$$k_i = \frac{1}{\langle \tau_i \rangle}$$

with  $k_i$  the off-rate from state  $i$  and  $\langle \tau_i \rangle$  the mean of all dwell times  $\tau_i$  in state  $i$ .

In order to arrive at the specific transition rates from state  $i$  to all the individual other states in the network, this overall off-rate has to be split up into the contributing pathway components.

Let  $k_{ij}$  be the transition rate from state  $i$  to state  $j$ . Summing over the transition rates to all possible states  $j$  must again result in the total off-rate from state  $i$ , i.e.  $k_i$ :

$$k_i = \sum_{j \neq i} k_{ij}$$

Let  $N_{ij}$  be the number of transitions from state  $i$  to state  $j$  detected by the HMM. Then:

$$\frac{k_{ij}}{k_{ik}} = \frac{N_{ij}}{N_{ik}}$$

Solving for  $k_{ij}$  gives:

$$k_{ij} = k_{ik} \cdot \frac{N_{ij}}{N_{ik}}$$

Inserting this into the equation for  $k_i$  results in:

$$k_i = \sum_{j \neq i} k_{ik} \cdot \frac{N_{ij}}{N_{ik}}$$

Pulling the special case of  $j = k$  out of the sum results in:

$$\begin{aligned} k_i &= k_{ik} \cdot \frac{N_{ik}}{N_{ik}} + \sum_{\substack{j \neq i \\ j \neq k}} k_{ik} \cdot \frac{N_{ij}}{N_{ik}} \\ &= k_{ik} \cdot \left[ 1 + \sum_{\substack{j \neq i \\ j \neq k}} \frac{N_{ij}}{N_{ik}} \right] \end{aligned}$$

Solving for the transition rate from state  $i$  to state  $k$  results in:

$$k_{ik} = \frac{k_i}{\left[ 1 + \sum_{\substack{j \neq i \\ j \neq k}} \frac{N_{ij}}{N_{ik}} \right]}$$

#### 4.5 Single-Exponential Lifetimes of States

An underlying Poisson process yields a single-exponential distribution of lifetimes in a certain state. This situation is familiar from nuclear decay for example, but also applies to the lifetime distribution of molecular states.

Assume a molecule in state 1, which can undergo a transition to state 2 with a certain off-rate  $k_{12}$ . The probability for the molecule to remain in state 1 after a time  $t$  is then given by:

$$p_1(t) = e^{-k_{off} \cdot t}$$

The probability that a transition to state 2 has occurred after a time  $t$  is then:

$$p_{12}(t) = 1 - p_1(t) = 1 - e^{-k_{off} \cdot t}$$

Fig.24 shows a trace of the aforementioned N-terminal lid flipping along with the lifetime distributions of the lid-closed state (purple) and lid-open state (dark blue). The lifetimes perfectly follow the single-exponential fits (black).

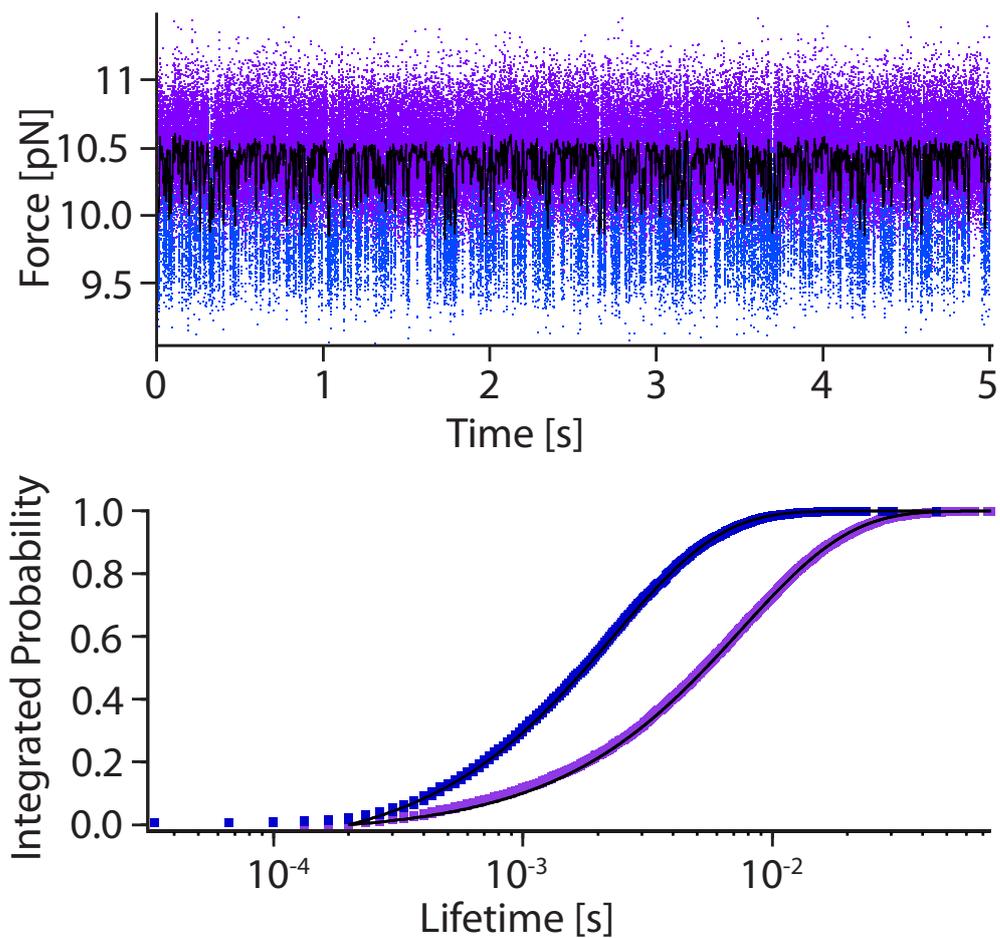


Figure 24: Single-exponential lifetime distributions of flipping closed and flipping open state.

# Results

## 5 GR-LBD Basics

The GR-LBD’s ligand binding dynamics as well as its protein folding pathway and free energy landscape have been published by Suren and Mößner in 2018 [44].

In order to facilitate understanding of the upcoming advanced chaperone experiments presented in this work, a basic introduction to the most important early results dating from the collaboration between Suren and Mößner is given here. For a more detailed discussion, refer to [44].

### 5.0.1 Introduction to Passive-Mode Experiments with GR-LBD

Fig.25A shows a typical passive-mode trace of GR-LBD. The trace was recorded at a force bias of  $\sim 10$  pN in the natively folded and ligand bound state (purple, holo state). The DEX concentration was  $200 \mu\text{M}$ . At  $t = 0$ , the GR-LBD starts in the purple holo state. At the given force bias, the N-terminal secondary structure element of holo GR-LBD, consisting of the first 33 amino acids, opens and closes in rapid equilibrium, giving rise to fast “flipping” transition between the purple (closed) and dark blue (open) states (Fig.25B). Section 6 will clarify how exactly we identified and located this structural element at the N-terminus.

DEX dissociation exclusively occurred from the flipping-open conformation of GR-LBD. Conversely, the flipping-closed conformation could only be populated when DEX was bound. This secondary structure element therefore acts as a “lid” for ligand binding and dissociation (and will be referred to as such for the rest of this thesis). When DEX dissociates from the lid-open conformation (dark blue), this results in the light-blue DEX-unbound states (Fig.25C). In 25D, a schematic illustration of the flipping and the DEX dissociation is given.

The following section will elaborate how we identified this light-blue state as the DEX-unbound state. As soon as the GR-LBD rebinds DEX, it resumes the flipping transitions (purple/dark blue) indicative of the native ligand bound state (Fig.25C). Starting from the light-blue apo GR-LBD state, occasional partial force-induced unfoldings of the GR-LBD can occur, resulting in the red state.

### 5.0.2 Variation of DEX Concentration in Passive-Mode Experiments

In order to show that the light-blue state in Fig.25 is in fact the DEX-unbound state, we varied the DEX concentration. Fig.26 shows two traces, one at  $5 \mu\text{M}$  and the other at  $200 \mu\text{M}$ :

The dwell times in the light-blue state clearly depend on DEX concentration. At low DEX concentrations, the dwell times in this light-blue state become more extended, since after DEX dissociation, it takes a longer time for a new DEX to bind. At high DEX concentrations, these light-blue phases become very short, since DEX can rebind almost immediately after every dissociation. The time-scale is the same in both traces and the difference in dwell-times of the light-blue state is striking.

Since the dwell-time in the light-blue state scales inversely with DEX concentration, this state must be the DEX-unbound state. The DEX binding rates of  $k_{\text{bind}} = 0.046 \text{ s}^{-1} \mu\text{M}^{-1}$  and  $k_{\text{bind}} = 0.035 \text{ s}^{-1} \mu\text{M}^{-1}$  extracted from these two traces are

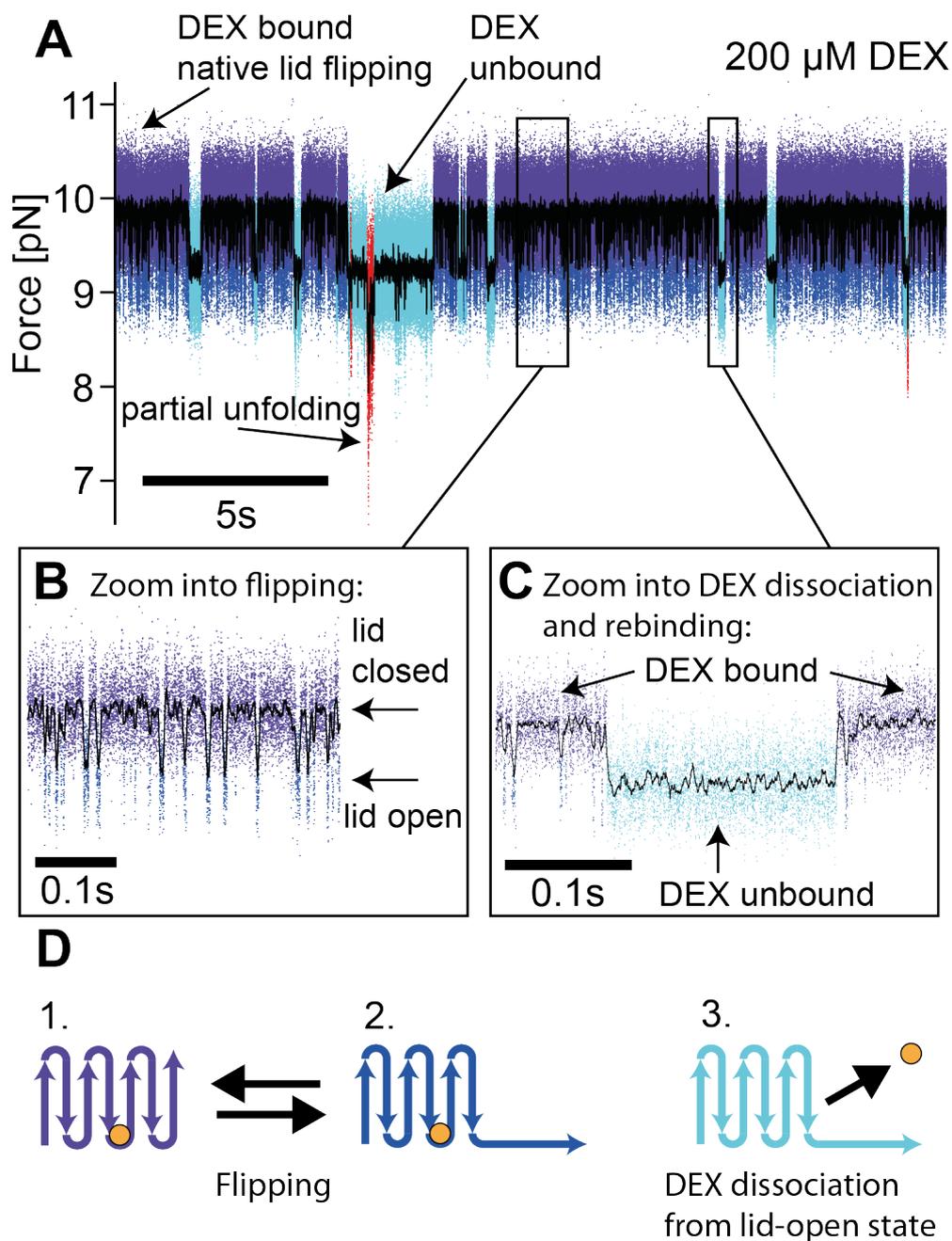


Figure 25: **A**) An illustrative passive-mode trace of GR-LBD exhibiting ligand bound flipping of the "lid" (fast transitions between purple and dark-blue), DEX dissociation (transition from dark blue to light blue), DEX-unbound phases (light blue), DEX rebinding (transition from light blue back to dark blue) as well as partial unfoldings (red). **B**) Zoom into the fast flipping transitions of the lid. Lid-closed state is colored in purple, the lid-open state in dark blue. **C**) Zoom into a DEX dissociation and rebinding event. **D**) Schematic illustration of the flipping (1. and 2., rapid opening and closing of the lid) and DEX dissociation from the lid-open state (3.).

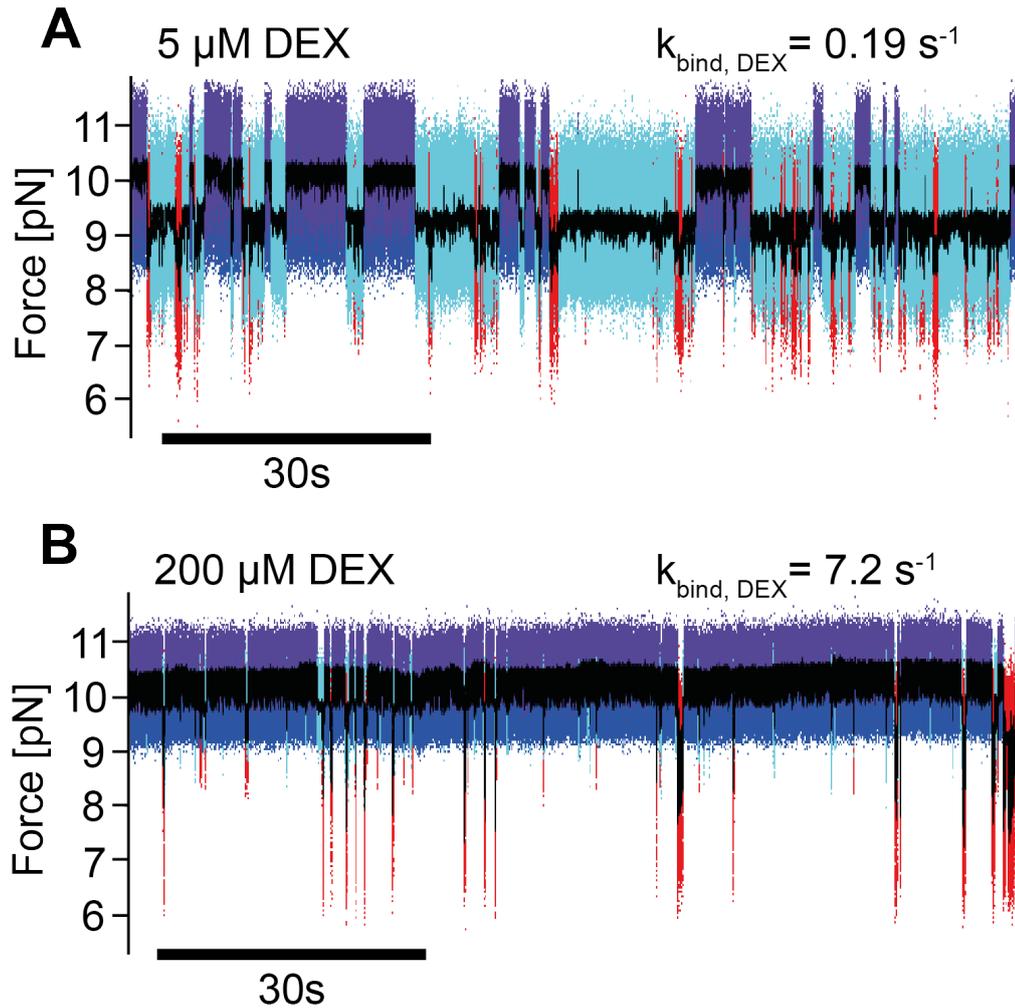


Figure 26: **A)** Passive-mode trace of GR-LBD in the presence of a low DEX concentration of 5  $\mu\text{M}$ . Light-blue states are populated for long dwell times and the average on-rate of DEX, as determined from this trace, lies at  $k_{\text{bind}} = 0.19 \text{ s}^{-1}$  (i.e.  $k_{\text{bind}} = 0.036 \text{ s}^{-1} \mu\text{M}^{-1}$ ) **B)** Passive-mode trace of GR-LBD in the presence of a high DEX concentration of 200  $\mu\text{M}$ . Light-blue states are populated for short dwell times and the average on-rate of DEX, as determined from this trace, lies at  $k_{\text{bind}} = 7.2 \text{ s}^{-1}$  (i.e.  $k_{\text{bind}} = 0.036 \text{ s}^{-1} \mu\text{M}^{-1}$ )

in very good agreement with the more extensive analysis conducted by Suren and Mößmer in 2018 [44], when we found  $k_{\text{bind}} = 0.033 \text{ s}^{-1} \mu\text{M}^{-1}$ .

The DEX dissociation rate should not be affected by the DEX concentration in solution. Averaging the DEX dissociation rate from the lid-open state of the two traces shown renders  $\tilde{k}_{\text{diss}} = 3.6 \text{ s}^{-1}$ , which is also in excellent agreement with the previously determined  $k_{\text{diss}} = 3.0 \text{ s}^{-1}$  [44].

### 5.0.3 Variation of Force in Passive-Mode Experiments

Increasing the force bias only slightly from  $\sim 10.0$  pN to  $\sim 11.2$  pN in the native holo state has tremendous effects on the GR-LBD, as depicted in Fig.27.

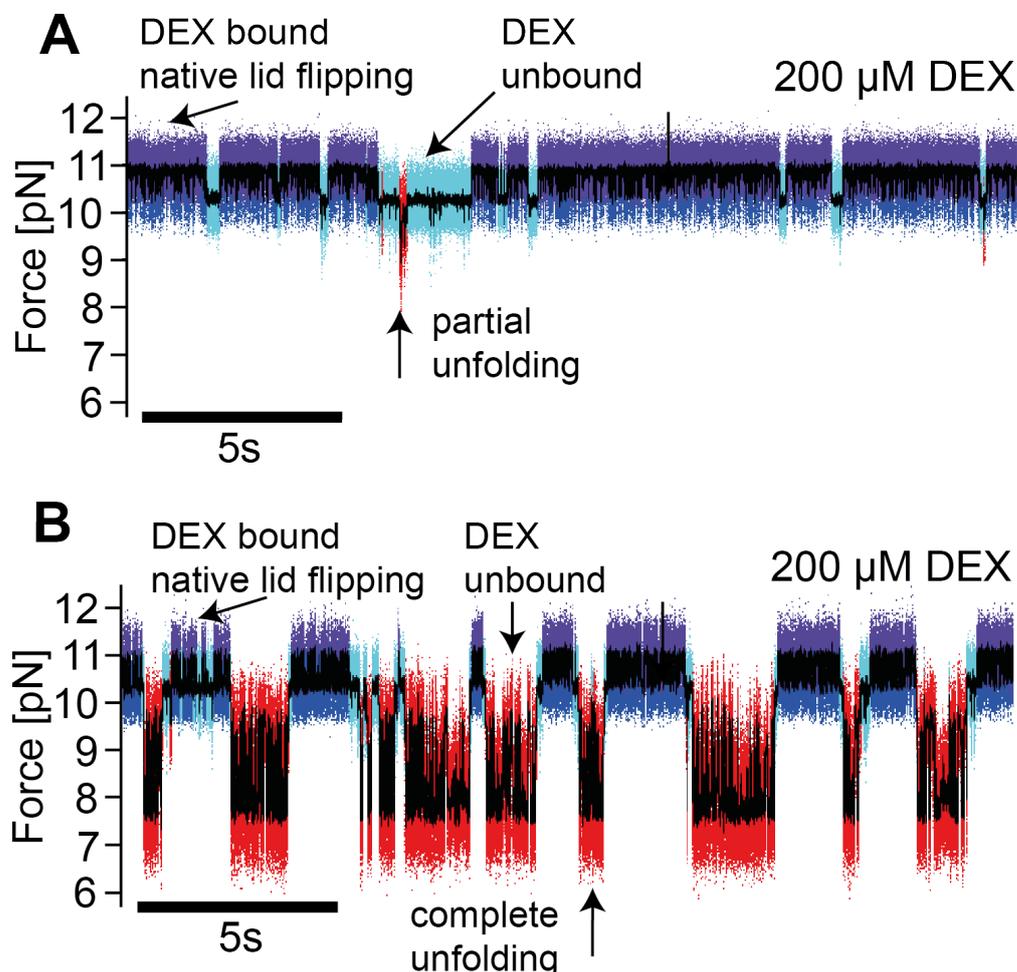


Figure 27: **A)** Passive-mode trace at  $\sim 10.0$  pN in the holo state. The typical GR-LBD signature involving flipping, DEX dissociation, DEX rebinding and occasional partial unfoldings is observed. **B)** Passive-mode trace at  $\sim 11.2$  pN in the holo state. Already a small increase in the force bias leads to frequent and often complete unfoldings (red) of the GR-LBD, exhibiting multiple folding intermediates on the way. The lid-open conformation of the flipping (dark blue) is populated more often than at the lower force.

The ratio between open and closed conformation of the lid in its flipping transitions is force-dependent, with the open conformation being populated more with increasing force. This is best observed by comparing the flipping at  $\sim 10.0$  pN and  $\sim 11.2$  pN directly, as shown in Figs. 28A and 28B. Also, under a higher force bias, unfolding of the ligand-unbound GR-LBD occurs more frequently and unfoldings proceed further, often reaching the completely unfolded state. Several unfolding intermediates are populated on the way [44]. Remarkably, the GR-LBD can refold within less than 0.1s from the completely unfolded state to the native apo state even against this high force bias. A detailed analysis of the folding intermediates as well as the free

energy landscape has been conducted in [44].

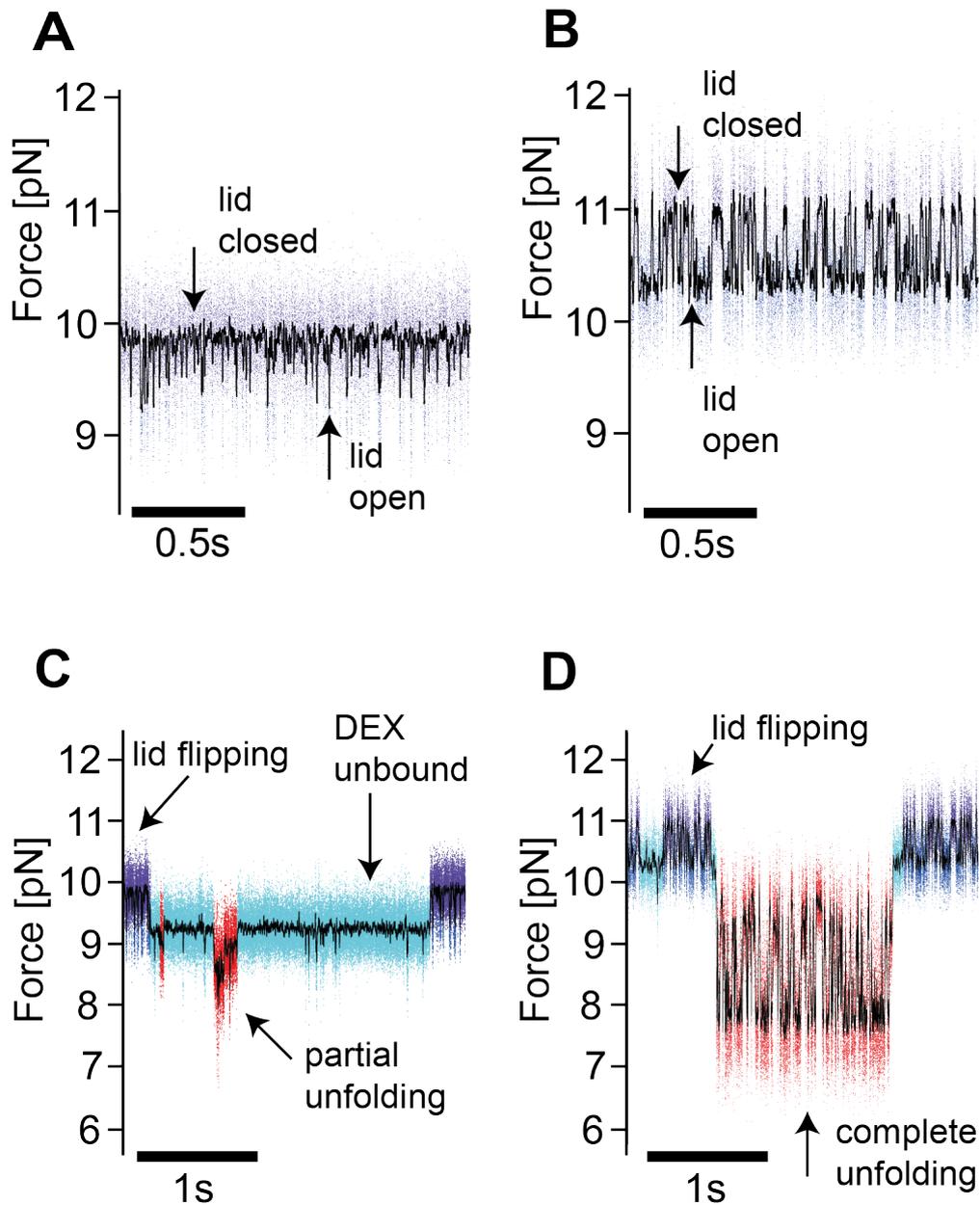


Figure 28: **A)** At a comparatively low force bias of  $\sim 10.0$  pN in the holo state, the flipping lid is mostly in the closed conformation and rarely in the open conformation. **B)** At a slightly higher force bias of  $\sim 11.2$  pN in the holo state, the flipping lid spends more time in the open conformation. **C)** At  $\sim 10.0$  pN in the holo state, the GR-LBD shows only very occasional and partial unfoldings. **D)** At  $\sim 11.2$  pN in the holo state, the GR-LBD frequently unfolds completely, populating multiple unfolding intermediates on the way.

#### 5.0.4 Introduction to Stretch-Relax Experiments with GR-LBD

A typical stretch-relax fingerprint of GR-LBD is given in Fig.29.

Up until  $\sim 7.5$  pN, the force-extension behavior of the dumbbell-assay is governed by

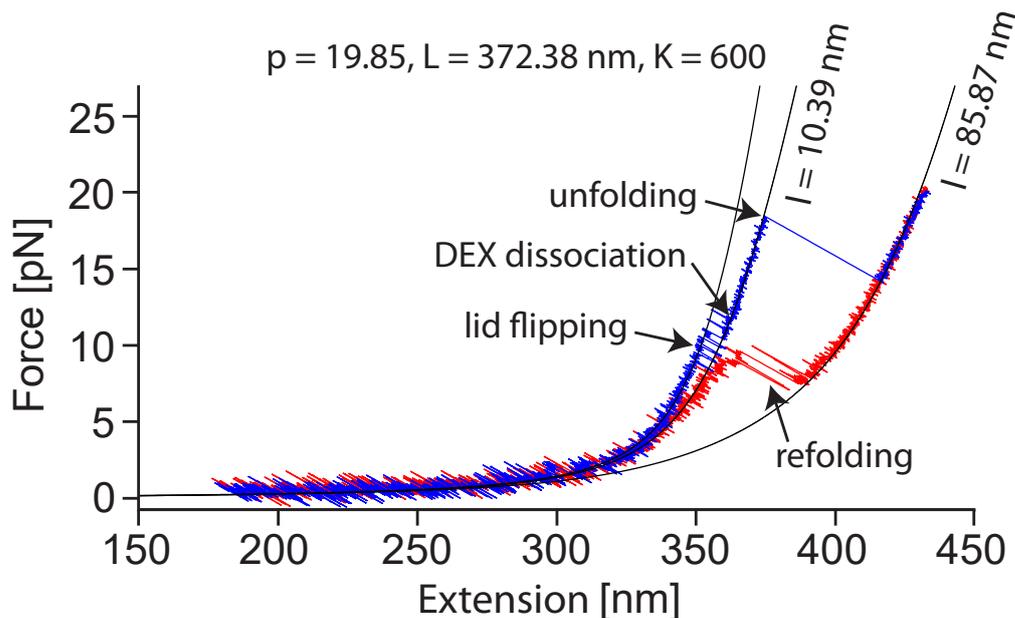


Figure 29: A typical stretch-relax cycle of GR-LBD. Using our knowledge from the passive-mode traces in the previous section, the parts of the trace can clearly be related to the processes of lid flipping, DEX dissociation and complete unfolding.

the stretching of the DNA handles, which act as an entropic spring (cf. section 4.1). At around  $\sim 7.5$  pN, the fast unfolding and refolding "flipping" transitions (already familiar as the purple/dark blue transitions from the passive-mode traces in the previous section) appear, indicating the opening and closing of the lid. Using WLC fits (cf. section 4.1), the unfolding contour length of the flipping can be determined as  $\sim 11$  nm. The end of the flipping (in Fig.29 at  $\sim 11.8$  pN) marks the dissociation of DEX (light-blue phase in passive-mode traces), which makes a refolding of the "lid" impossible until the rebinding of DEX. Eventually, the folded core of the GR-LBD unfolds, typically at forces between 15 pN and 30 pN.

Fig.30 shows an overlay of a DEX-bound (faded blue) and a DEX-unbound (green) stretching trace (the relaxation part is omitted here for clarity). Notably, in the green DEX-unbound trace the flipping is missing and the unfolding of the protein starts with an already unfolded lid.

The unfolding patterns in Fig.29 and 30 constitute a "fingerprint" of GR-LBD. In optical trapping experiments, recognition of this signature fingerprint of GR-LBD allows to identify a correctly tethered molecule at the beginning of each experiment. Fig.31 shows how reproducible these unfolding patterns are, with 10 consecutive stretch-relax cycles. 4 of the unfoldings occur with DEX-bound to GR-LBD (dark blue) and 6 of them occur in the DEX-unbound state (green). Refolding is always depicted in red.

During the unfolding of the core both in DEX-bound and DEX-unbound unfoldings, a short-lived intermediate at  $\sim 37$  nm contour length is populated. However, it only becomes clearly visible in stretch-relax cycles if the core happens to unfold at a relatively low force. Fig.32 shows an overlay of multiple bound and unbound pulls, all exhibiting the 37 nm intermediate.

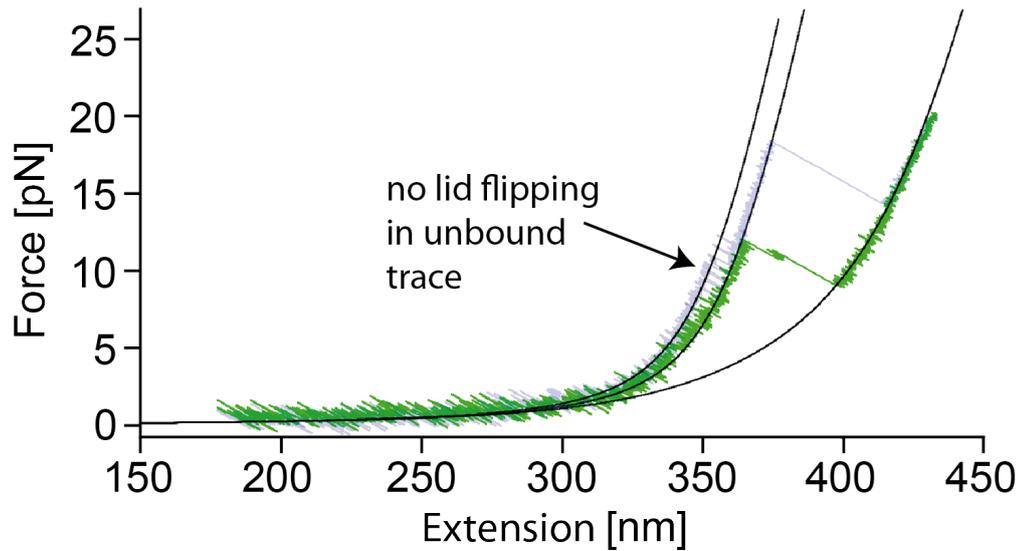


Figure 30: Overlay of a DEX-bound unfolding (faded blue) exhibiting lid flipping, and a DEX-unbound unfolding (green), which lacks the lid flipping and where the complete unfolding occurs at a much lower force than in the DEX-bound trace. Refolding traces are omitted here for clarity.

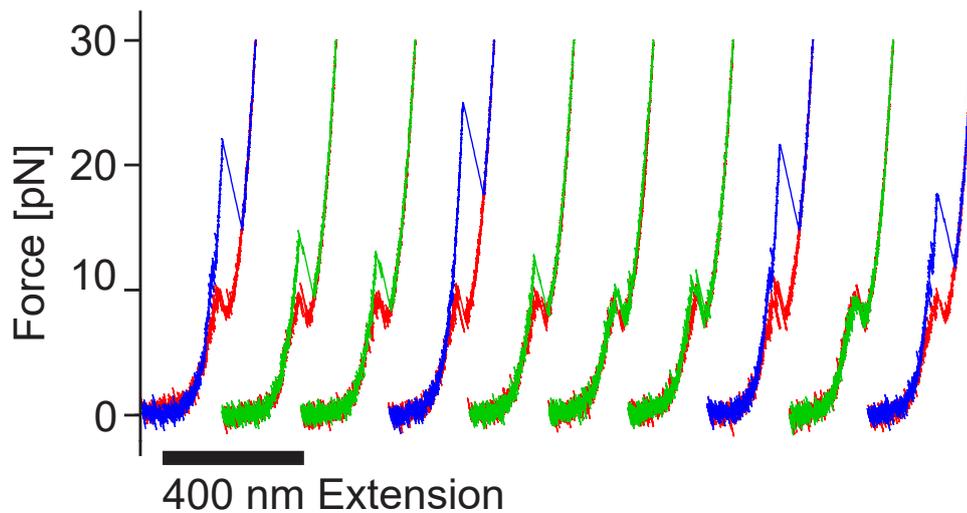


Figure 31: 10 consecutive stretch-relax cycles with DEX-bound unfolding traces (blue), DEX-unbound unfolding traces (green) and refolding traces (red).

The flipping of the lid acts as a convenient identifier or "force-extension fingerprint" of GR-LBD. Even without further unfolding of the molecule, the flipping transitions at  $\sim 10$  pN already identify the molecule for a trained experimenter's eye. As a result, it is sufficient to perform stretch-relax cycles that end at the flipping (as in Fig.33) and proceed no further in order to have a strong indication that one is starting the experiment with a natively folded holo GR-LBD. This is particularly advantageous in experiments of GR-LBD in combination with chaperones, where DEX dissociation or partial unfolding of GR-LBD mark the kickoff for chaperones to attack GR-LBD. In these experiments, it is therefore desirable to identify GR-LBD without unfolding it at the beginning of the experiment.

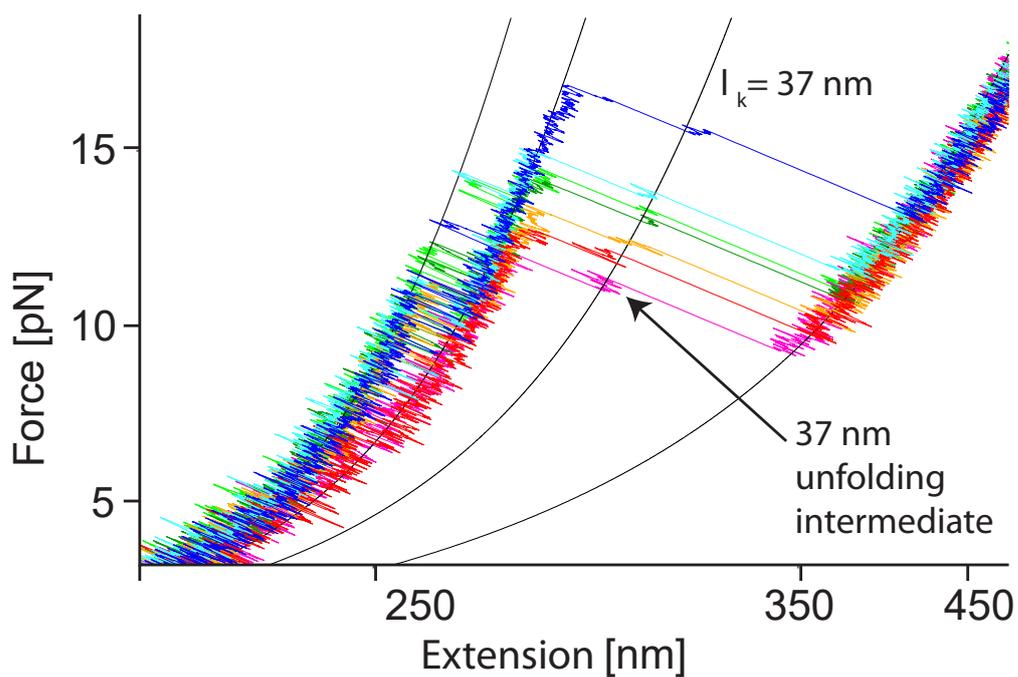


Figure 32: Overlay of multiple unfolding traces, all exhibiting the short-lived 37 nm unfolding intermediate.

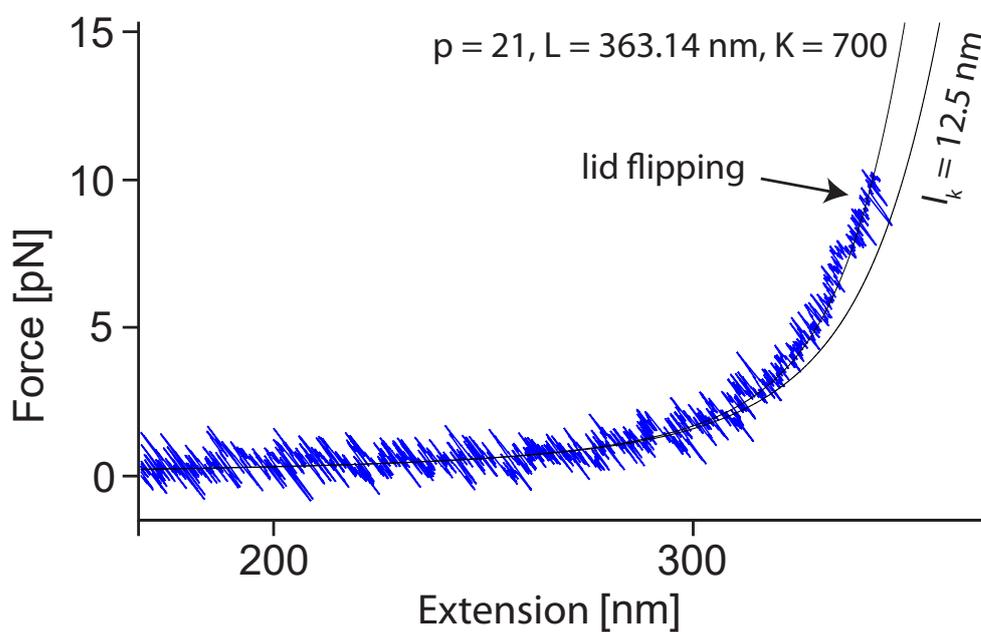


Figure 33: Just one flipping transition is enough to identify the GR-LBD during optical trapping experiments. No further unfolding is needed.

Identifying the GR-LBD at the beginning of an experiment without unfolding it will turn out to be essential in unfolding experiments with the Hsp70/Hsp40 chaperone system.

## 6 Structural Identification of the "Lid" Governing Hormone Binding

As introduced in section 5.0.1, the GR-LBD shows a characteristic flipping in the force range of  $\sim 7.5\text{--}10\text{ pN}$  with an unfolding contour length of  $\sim 11\text{ nm}$ . Some secondary structure element of the GR-LBD must be unfolding and refolding with high kinetics in this force range. Since this flipping constitutes the first unfolding of the GR-LBD and the molecule is tethered at its termini, it likely has its origin in the unfolding and refolding of a terminal secondary structure of GR-LBD. However, judging simply from the data presented so far, it is unclear whether it is the N-terminal or C-terminal secondary structure element that is flipping.

The unfolding contour length of  $\sim 11\text{ nm}$  corresponds to an estimated length of the flipping part of 33 amino acids, assuming an average length of an amino acid in a peptide chain of  $0.365\text{ nm}$  [121]. Close examination of the terminal parts of GR-LBD in its crystal structure showed that both the first N-terminal as well as the first C-terminal secondary structure element consist of about 33 aa. Fig.34 shows the respective parts, with the N-terminus in Fig.34 A and the C-terminus in Fig.34 B.

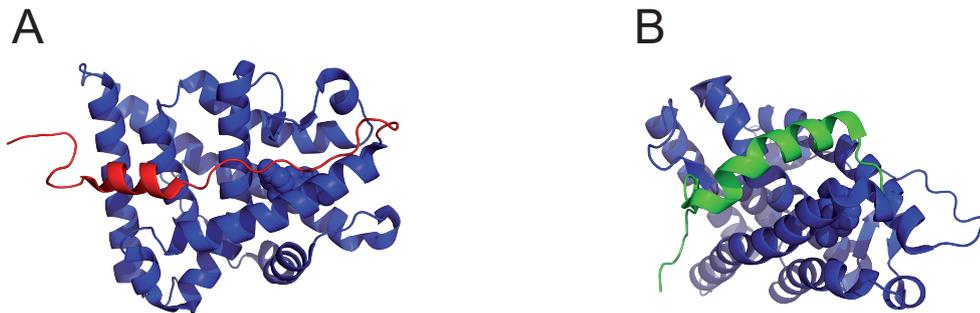


Figure 34: Both the N-terminal (**A**) as well as the C-terminal (**B**) secondary structure element are about 11 nm in length.

In order to find out which part of the molecule is responsible for the flipping, several experimental approaches were tried. In the end, three of them succeeded, proving without a doubt that the flipping part must be the N-terminal secondary structure element. Since these experimental approaches might prove useful for future experiments and are elucidating on their own, they will be discussed in the following three sections.

### 6.1 Tethering between an Internal and a Terminal Cysteine

The first experimental approach was to create two different constructs of GR-LBD that allowed to unfold only the N-terminal half or only the C-terminal half of the protein individually. The expectation was that only one of the halves would exhibit a flipping during its unfolding, which would suggest that this half contained the sec-

ondary structure we were looking for. Of course, unfolding the protein from different attachment points might, in principle, lead to unexpected changes in the unfolding pathway.

To this end a surface cysteine of GR-LBD that is located roughly in the middle of its amino acid sequence (position 638, cf. section 1.2) and which is usually genetically removed to avoid unspecific oligo binding, was used to tether the molecule between the N-terminus and the internal cysteine in one construct and between the C-terminus and the internal cysteine in another construct. The two constructs were purified by Daniel Rutz and are shown in Fig.35 with their respective tethering points.

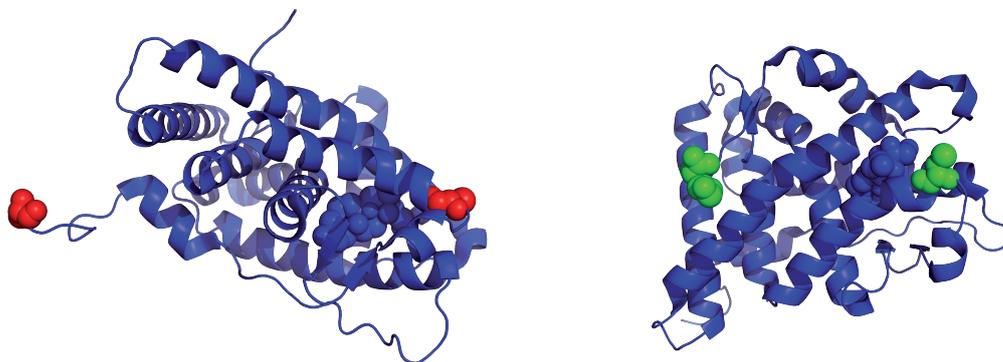


Figure 35: **A)** Tethering the molecule between N-terminal and internal cysteine. **B)** Tethering the molecule between C-terminal and internal cysteine.

The experiment was successful in so far as one half of GR-LBD, namely the N-terminal half, showed a flipping, while the C-terminal half did not. Fig.36 compares the unfolding traces of the two halves.

The flipping now occurred at a much higher force of  $>20$  pN (before  $\sim 10$  pN) and exhibited a lower unfolding contour length of  $\sim 6$  nm (before:  $\sim 11$  nm). However, these differences are entirely plausible: As long as the total energy of the flipping of the native lid and the flipping of the elongated lid is conserved, everything is still theoretically sound. The observed flipping in the internal cysteine construct occurred at about twice the force of the usual flipping ( $\sim 20$  pN vs.  $\sim 10$  pN), but also exhibited about half the usual unfolding contour length ( $\sim 6$  nm vs.  $\sim 11$  nm). This experiment gave a first and very strong hint that it could be the N-terminal secondary structure that constitutes the "lid" element, which governs the ligand binding behavior of GR-LBD.

However, there still remained some doubts whether this experiment was the definitive answer to the question of the origin of the lid flipping. Strangely enough, the N-terminal secondary structure does not form part of the ligand binding pocket, and it is not evident how it should have such a crucial impact on hormone binding. Furthermore, it was still conceivable that, when tethering the GR-LBD between its N-terminus and the internal cysteine, some other secondary structure element in the N-terminal half of GR-LBD also caused a different flipping, distinct from the flipping we were investigating.

Therefore, two further experiments were conducted in parallel, in order to answer

the question of the origin of the flipping beyond any doubt and in a lot more detail. They are discussed in the following two sections.

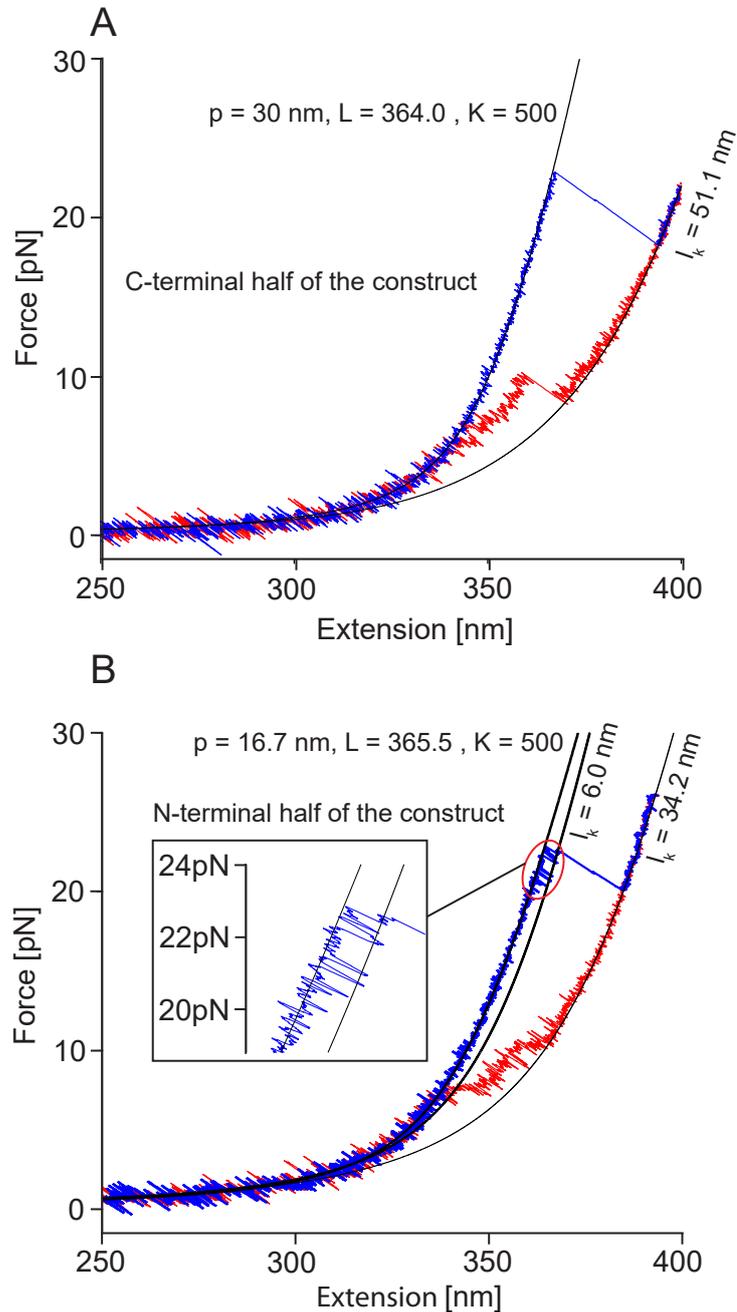


Figure 36: **A**) The GR-LBD was tethered between the C-terminus and the internal cysteine at amino acid position 638. No flipping could be detected. The contour length of the completely unfolded construct was 51.1 nm. **B**) The GR-LBD was tethered between the N-terminus and the internal cysteine at amino acid position 638. At a high force  $>20$  pN, a flipping with an unfolding contour length of  $\sim 6$  nm was detected. The contour length of the completely unfolded construct was 34.2 nm. Together, the complete contour lengths of the two constructs add up to the usual 85 nm of the entire GR-LBD construct.

## 6.2 Competition Assay between Lid and Peptide Chain

The second experimental approach to determine which terminus the flipping originated from was a competition assay. A copy of the N-terminal secondary structure element of the GR-LBD was ordered from Biomers [122] as a 33 amino acids peptide chain. In what follows, this peptide chain will be called "N-peptide" (shown in Fig.37 A) and is to be contrasted from the GR-LBD's actual "N-terminus" (shown in Fig.37 B), which is connected to the folded core of the protein.

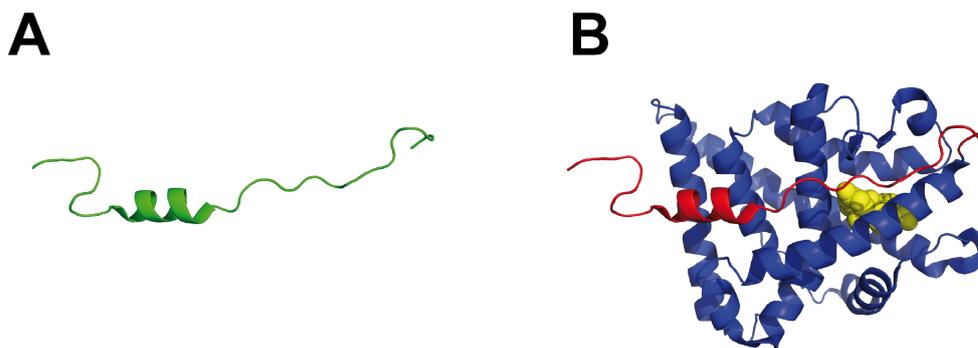


Figure 37: **A)** The "N-peptide" in green. **B)** The GR-LBD's "N-terminus" is highlighted in red, DEX in yellow.

This N-peptide was added in solution to the optical trapping experiments with tethered full-length GR-LBD (Fig.38(1)). If the N-terminus is what causes the flipping, then whenever the GR-LBD is in the flipping-open DEX-bound conformation (Fig.38(2)), the N-peptide and the GR-LBD's N-terminus will compete for the same binding contacts on the folded core of the protein. Therefore, two scenarios are possible, illustrated in Fig.38(3a) and (3b):

**Fig.38(3a):** The N-peptide manages to temporarily occupy the binding contacts on the folded core that are usually covered by the N-terminus. When the N-peptide occupies these binding contacts, the GR-LBD's actual N-terminus cannot attach to the folded core of the protein ("the lid cannot close"). Since the N-peptide is floating freely in solution and is not subjected to force in our experiments, its binding should therefore result in unusually long flipping-open phases.

**Fig.38(3b):** The N-terminus of GR-LBD closes again, which is the standard transition from the flipping-open to the flipping-closed state (dark-blue to purple).

This expectation was perfectly met by experiment. When looking at the flipping of GR-LBD in the presence of 10  $\mu\text{M}$  N-peptide and 20  $\mu\text{M}$  DEX, a mixture of both scenarios (Fig.383a and Fig.383b) was observed. Fig.39A shows the resulting flipping trace, and Fig.39B shows a zoom into it. The flipping-closed state is depicted in purple, the flipping-open state, which is the dark-blue in all other figures, is colored green here for better contrast. The state that occurs, when the N-peptide blocks the binding site on the folded core of the protein, is colored in red. Note that the red dwell-times are markedly longer than the green dwell-times, but too short to be DEX unbound phases at the given DEX concentration of 20  $\mu\text{M}$ .

HMM analysis renders an estimate of the on-rate of the N-peptide to the flipping-

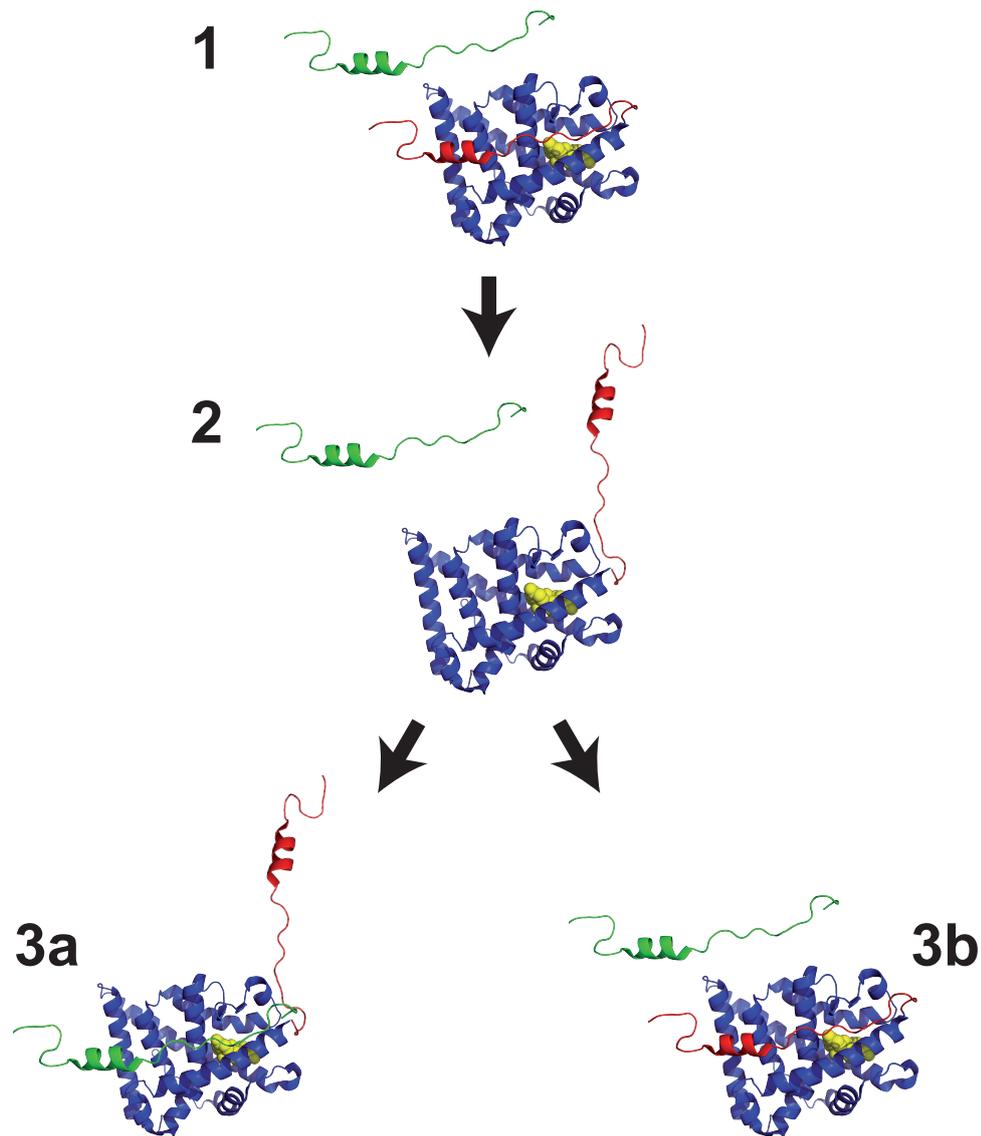


Figure 38: **1)** The N-peptide is added in solution to the GR-LBD optical trapping experiment. **2)** The GR-LBD's N-terminus unfolds, leading to the flipping-open state. Now the N-peptide and the N-terminus will compete for the same binding region on the folded core of the protein. **3a)** In the first scenario, the N-peptide manages to occupy the binding region that is usually covered by the N-terminus. This will lead to unusually long flipping-open phases in our experiments. **3b)** In the second scenario, the N-terminus will bind to the folded core as usual, leading to the normal flipping-closed state.

open state of  $k_{\text{on N-peptide}} \approx 100\text{s}^{-1}$  and an off-rate of  $k_{\text{off N-peptide}} \approx 30\text{s}^{-1}$ . Note that the on-rate of the N-peptide is about 30 times higher than the DEX off-rate ( $k_{\text{off DEX}} \approx 3\text{s}^{-1}$  [44]), which means that the majority of red phases in Fig.39 are not the DEX-unbound state, but the N-peptide blocked state. The usual flipping rates between the closed and open state (purple and green in Fig.39) remained unaffected by the presence of the peptide, as Fig.40 shows. Increasing the force bias lead to a higher probability of the flipping-open state, which

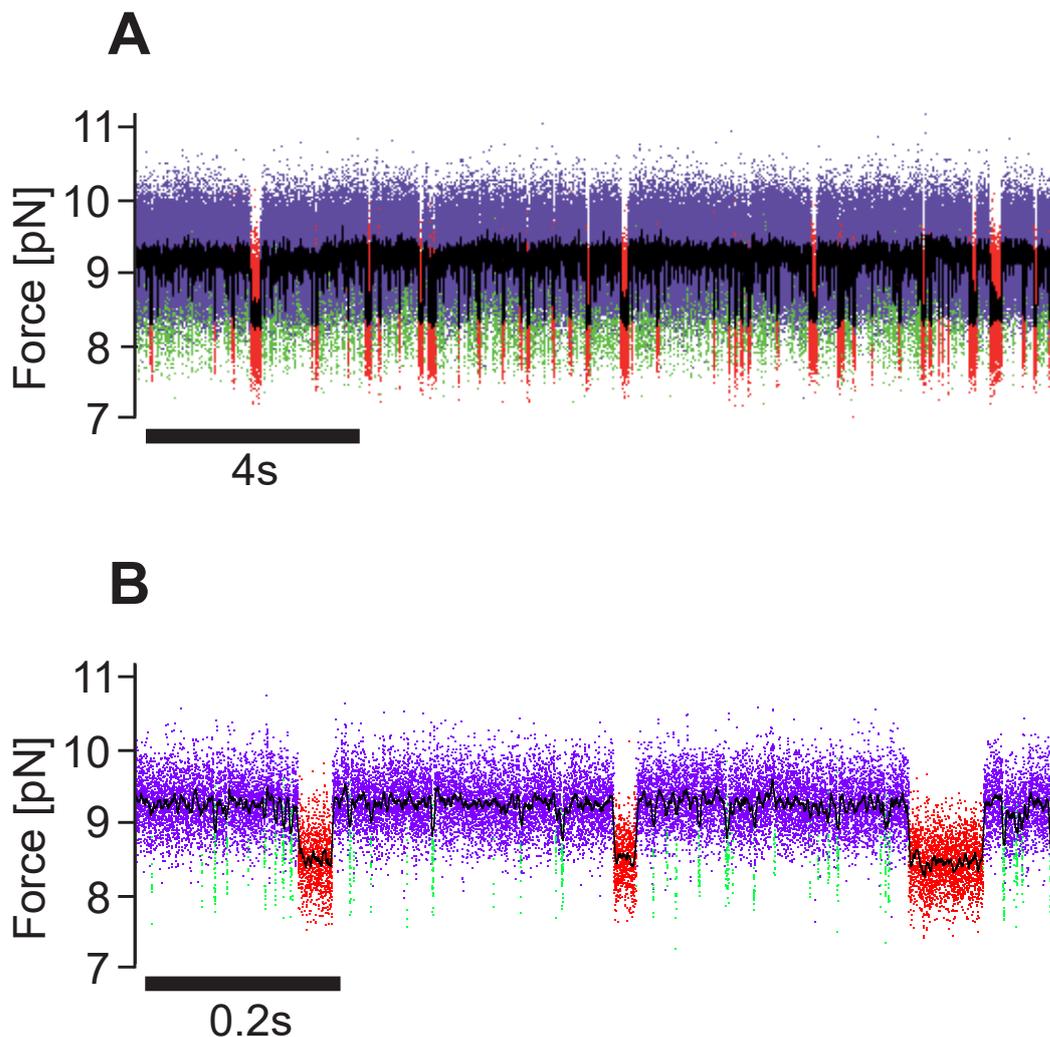


Figure 39: **A)** Flipping in the presence of the N-peptide in solution. The usual flipping-closed (purple) and flipping-open (green) state still occur, but an additional "N-peptide-blocked" lid-open state shows up (red) resulting in occasional unusually long flipping-open dwell-times. **B)** Zoom into A), illustrating the contrast between the usual short-lived flipping-open states (green) and the unusually long-lived N-peptide-blocked states (red). The DEX concentration was  $20\ \mu\text{M}$ , which means that the red phases cannot be DEX unbound phases (at least not predominantly), since the dwell-times in the red state are too short for DEX rebinding at this low concentration.

accordingly also to a proportionally higher frequency of the N-peptide blocked state. As a control experiment, another peptide corresponding to the 25 C-terminal amino acids was added at concentrations up to  $100\ \mu\text{M}$ , which did not affect the flipping kinetics.

The competition assay described in this section offers a simple, fast and elegant method to identify structural elements of a protein in unfolding events, with a wide range of possible applications.

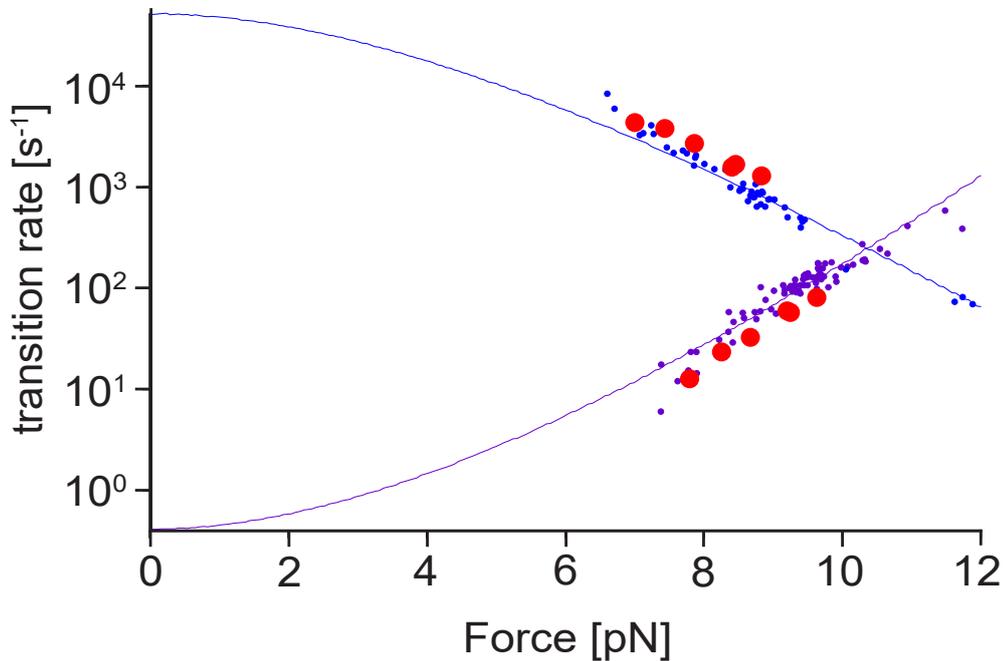


Figure 40: The force-dependent flipping rates between open and closed state are the same in the presence of peptide (red data points) and in the absence of it (purple and dark blue data points, adapted from [44]).

### 6.3 Linker Construct

The most straightforward way to show that the N-terminal secondary structure element is the flipping part of GR-LBD was to create a "linker construct", where an 11 amino acids loop (GGSGGSGGSGG) was inserted into the N-terminal secondary structure element, between the serine at amino acid position 551 and the valine at amino acid position 552 (Fig.41A). Assuming that the N-terminal secondary structure is the flipping element, this should result in an increased unfolding contour length of the flipping. The total  $\Delta G$  of the transition is assumed to remain roughly the same, as no new contact points between the N-terminal secondary structure and folded core were created, although the insertion of the loop might inhibit some of the previous contact points.

Naturally, we expected the contour length of the transition to increase due to the loop insertion, while the force at which the flipping occurs was expected to decrease accordingly.

Experiments perfectly confirmed this assumption, as the comparison between the flippings of the normal F602S construct and the linker construct in Fig.41B shows. The flipping of the linker construct exhibits a longer unfolding contour length and appears at a lower force than the usual flipping of the F602S construct.

Looking at the lid-flipping of the linker construct, the lid-open state (dark blue) shows the usual kinetics (cf. Fig.42), while the lid-closed state (purple) unexpectedly seems to exhibit a double-exponential dwell-time distribution. This can only have

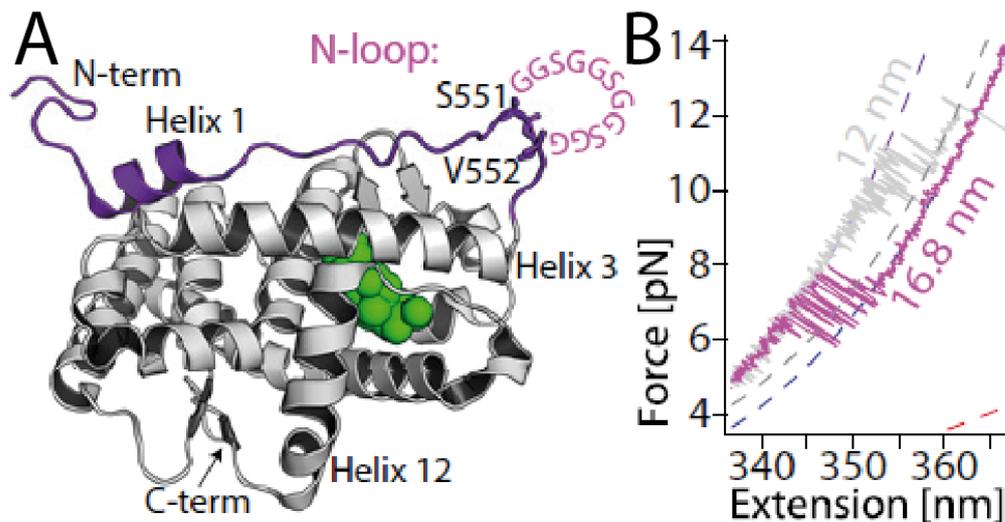


Figure 41: **The longer flipping contour length in the GR-LBD linker construct proves that the lid must be the N-terminal secondary structure element:** **A)** Schematic representation of the linker construct, with an inserted loop of 11 amino acids (GGSGGSGGSGG) in the N-terminal secondary structure. **B)** The flipping of the linker construct (grey) shows a longer unfolding contour length and appears at a lower force than the usual flipping of the F602S construct (pink). Traces were recorded at a stretching speed of  $50 \text{ nm s}^{-1}$ . Figure reprinted from [44] with permission. Copyright 2018 by NAS.

one cause: the lid-closed state identified by HMM analysis actually consists of two different states with different off-rates.

The question is now how the extension of the N-terminal secondary structure could result in the possibility of this structure to assume two different closed conformations with equal contour length, but different off-rates. Looking at Fig.42, it is evident that the "new" dwell times are on the shorter end of the spectrum, i.e. inserting the loop gave rise to previously non-existent short dwell times in the flipping-closed conformation. One possible explanation takes into account the possibility of a proline switch. The N-terminal secondary structure of GR-LBD contains 5 prolines. One of these prolines is directly adjacent to the end of the inserted linker, at position 553 (the linker is inserted between position 551 and 552). Also, even in the native flipping without linker, about 5% of the time, the flipping shows unusually fast kinetics for  $\sim 2 \text{ s}$ , and we hypothesized that this is due to a proline switch at position 553. However, whenever this proline switch occurred, it affected the dwell-time in the flipping-open conformation, and did not seem to affect the dwell-times in the flipping-closed conformation.

Proline is a secondary amino acid, which means it does not contain the  $\text{NH}_2$  amine group. Its side chain is connected to the protein backbone twice, forming a five-membered ring. This cyclic structure of its side chain gives proline an exceptionally high conformational rigidity compared to other amino acids. As a result, proline is highly prevalent in thermophilic organisms. Furthermore, prolines are often found in turns, at the beginning of alpha-helices and in the edge-strands of beta sheets.

All amino acids can assume a *trans* and a *cis* conformation. *Trans* refers to the conformation where the  $\text{C}\alpha$  carbon atoms of two amino acids connected by a peptide

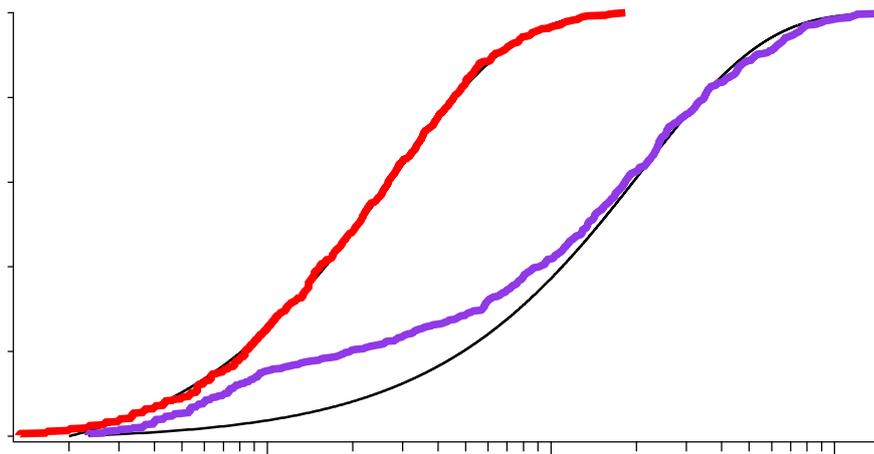


Figure 42: **The linker construct shows double-exponential dwell-time distribution of the flipping-closed state:** **A)** Flipping of the linker construct with the flipping-closed state in purple and the flipping-open state in red. **B)** Zoom into A). **C)** The dwell-time distributions of the flipping-closed and flipping-open state. While the flipping-open state shows identical kinetics as in the GR-LBD construct without linker, the flipping-closed conformation seems to show a double-exponential dwell-time distribution.

bond are on opposite sides of the backbone of the peptide chain, and *Cis* refers to the opposite case, where the  $C\alpha$  carbon atoms are on the same side of the backbone of the peptide chain. *Trans* is by far the most common conformation. 99.9% of all amino acids in proteins assume this conformation under unstrained conditions. The unique exception is proline, which due to its special structure is found in the *cis*-conformation unusually often, about 3-10% of the time. It is conceivable that the two different conformations of the proline can lead to two different flipping-closed states, which possess different off-rates.

Another possible explanation takes into account the increased mobility and flexibility of the N-terminal secondary structure after insertion of the loop. It is possible that the N-terminal structure can find alternate binding modes to the folded core of the protein, leading to two different flipping-closed states.

## 7 Improvement of the Biochemical Assay of GR-LBD Experiments

### 7.1 DTT Abolishes the "Ligand Binding Incompetent State" of GR-LBD

In early experiments by Suren and Mößner, the GR-LBD frequently assumed a non-native ligand-unbound and quite stable conformation, which was incompetent of binding DEX. Accordingly, this state was named "binding incompetent state" (BIC). Several examples of this BIC both in passive-mode as well as stretch-relax cycles are given in Fig.43.

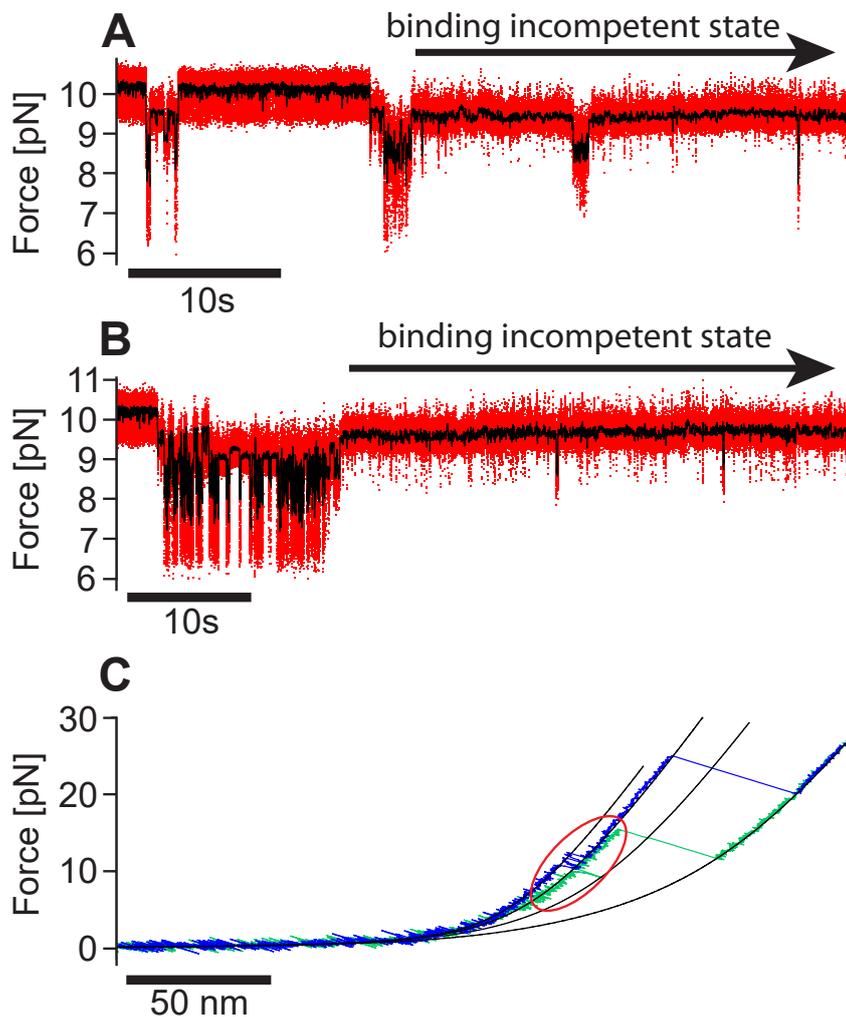


Figure 43: The "binding incompetent state" (BIC) of GR-LBD: **A)** The two passive-mode traces at the top show typical occurrences of the BIC after partial unfoldings of GR-LBD. The contour length of this state is indistinguishable from the contour length for the DEX unbound state. Once caught in this BIC, the GR-LBD usually never recovers from it on its own. **B)** The BIC in a stretch-relax cycle in green, with a native unfolding in blue for comparison. Again, with respect to its contour length, the BIC is indistinguishable from the DEX unbound state.

With respect to its unfolded contour length, this state was identical to the natively folded DEX-unbound state. However, it could never rebind ligand, even after long waiting periods. The BIC severely hindered experiments since the GR-LBD typically was "trapped" in this conformation indefinitely. The only way to give the GR-LBD a chance to fold back to the native conformation was to completely unfold it by force and relax it afterwards. This sometimes resulted in the recovery of the native conformation, but it also meant a severe interruption of the previous experiment. Since this BIC exclusively appeared after at least partial force-induced unfolding of GR-LBD, one possible explanation for its occurrence was the formation of intramolecular non-native interactions such as disulfide bridges between cysteines. Furthermore, the fact that GR-LBD was typically trapped in this non-native state indefinitely implied that it was likely to originate from a strong non-native bond. Disulfide bridges count among the strongest intramolecular bonds, with a typical bond dissociation energy of  $251 \text{ kJ mol}^{-1}$ . Disulfide bridges are formed by the oxidation of the thiol-groups (-SH) of two cysteines (Fig.44).

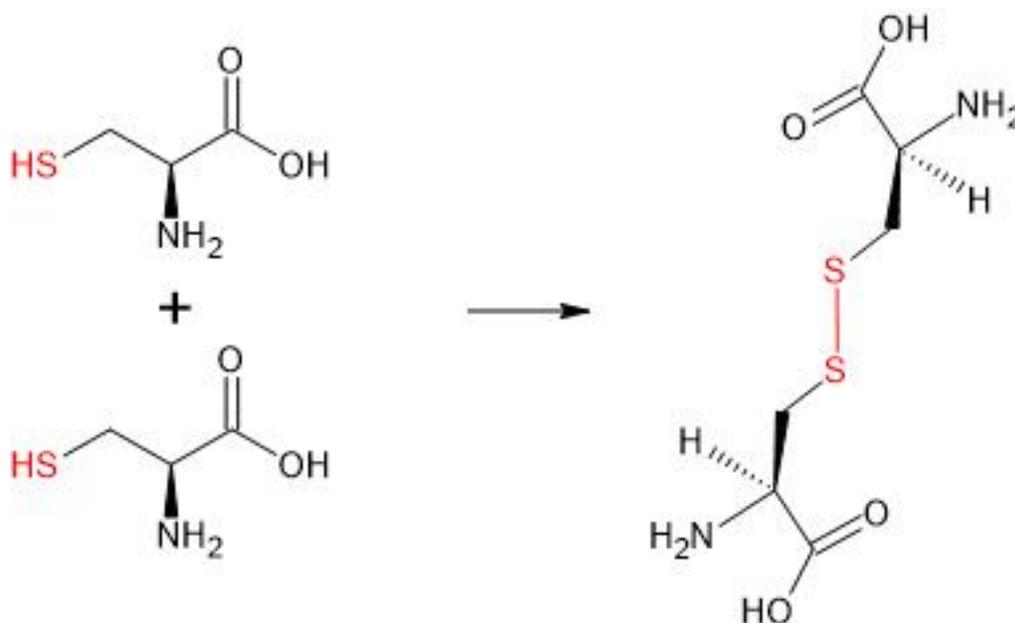


Figure 44: Disulfide-bridge formation between two cysteines.

The sequence of GR-LBD (F602S, cf. section 1.2) contains 4 internal cysteines at amino acids 622, 643, 665 and 736 (neglecting the surface cysteine at amino acid 638, which was genetically removed, cf. section 1.2). Fig.45 shows their positions in the crystal structure. Given their close proximity in the folded structure, it is conceivable that these cysteines could form non-native disulfide bridges.

To test whether disulfide bonds were the origin of the binding-incompetent state, we added the reducing agent Dithiothreitol (DTT) to the assay. DTT is able to reduce and thereby break disulfide bridges. The reaction proceeds in two sequential thiol-disulfide exchanges as shown in Fig.46. Note that the reaction typically does not stop after the first step, when an intermolecular disulfide bridge has been established between the DTT and the previously reduced thiol group. Rather, it continues with the second step, where the two thiol groups of the DTT close the intramolecular ring.

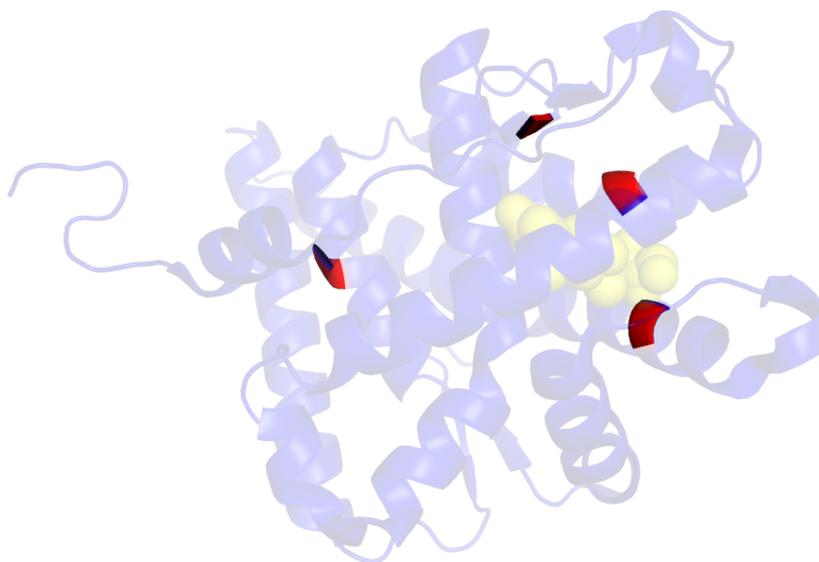


Figure 45: Internal cysteines in GR-LBD are highlighted in red. Bound DEX is depicted in yellow.

The reaction results in a reduced disulfide bridge and an oxidized DTT molecule. Therefore, the DTT does not remain bound to the cysteine.

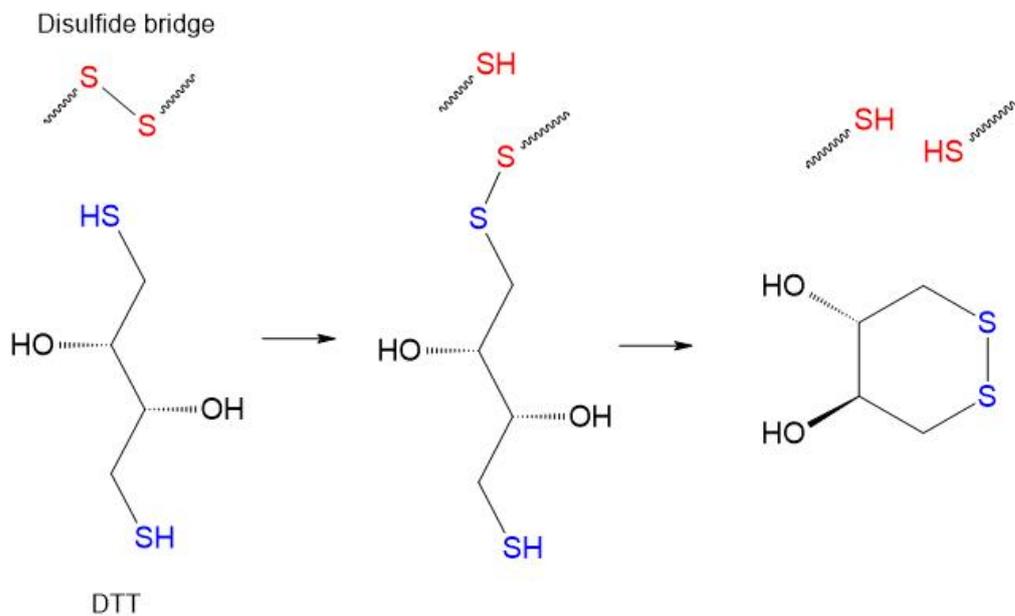


Figure 46: Reduction of disulfide bond by DTT in a two-step reaction.

This approach proved to be successful and the binding-incompetent state never occurred again in the presence of 1mM DTT, which greatly simplified all further experiments.

## 7.2 Optimization of the Oxygen Scavenging System

### 7.2.1 Misfolding Tendency of GR-LBD

In early single-molecule force spectroscopy experiments with GR-LBD, a severe and recurring problem was its general tendency to misfold upon partial or complete force-induced unfolding and to quickly enter a variety of non-native misfolded conformations. Fig.47 shows three exemplary traces, where the GR-LBD quickly misfolds after having been partially or completely unfolded.

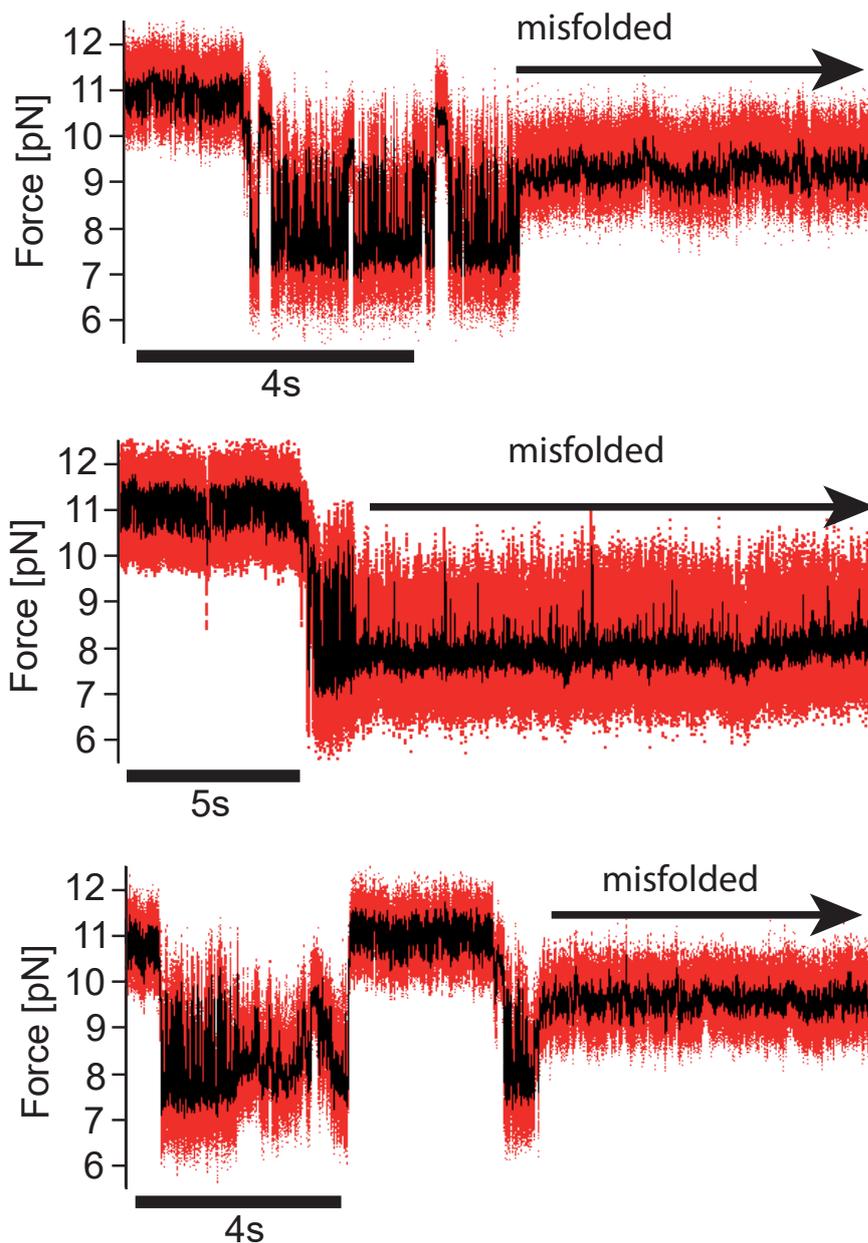


Figure 47: Three representative traces showing a misfolding of GR-LBD after partial or complete unfolding. The left end of the arrow indicates when misfolding happened, and typically, the GR-LBD was trapped in the misfolded state indefinitely.

These misfoldings happened in virtually every experiment and the instances of such traces were in the hundreds. Recording long traces or observing force-induced unfolding followed by refolding of the molecule were practically impossible under these conditions. If the biochemical assay had not been drastically improved soon after, none of the data shown in this thesis would have been possible to obtain.

One common reason of unwanted biochemical damage to proteins are reactive oxygen species. Optical trapping experiments are particularly prone to encounter this problem, since the incident laser light drastically enhances the production of reactive oxygen species, in particular at the surface of the beads [110].

Thus, an oxygen scavenging system is routinely used in all single-molecule optical trapping samples. The usual choice is the combination of glucose oxidase, catalase and glucose. However, this scavenging system was insufficient to create buffering conditions in which the GR-LBD could thrive. The following sections will present alternative scavenging systems and their respective advantages and disadvantages. Afterwards, a discussion regarding the best choice of scavenging system for the GR-LBD assay follows.

### 7.2.2 Glucose Oxidase/Glucose/Catalase

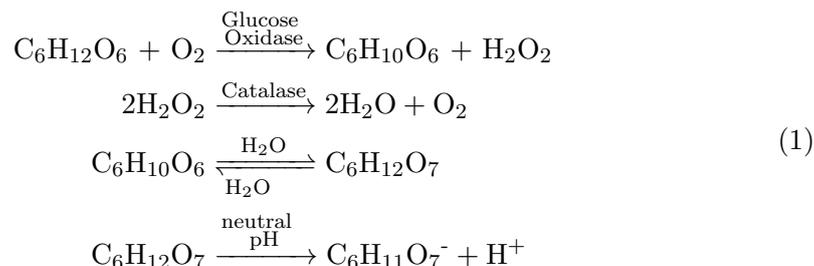
The most commonly used scavenging system in the Rief group consists of glucose oxidase, catalase and glucose (GOC). It deprives the solution of oxygen via the oxidation of glucose, which is catalyzed by glucose oxidase. Catalase is added to decompose the harmful byproduct hydrogen peroxide.

In aqueous solution, glucose ( $C_6H_{12}O_6$ ) exists in two forms, namely 36.4%  $\alpha$ -D-glucopyranose and 63.6%  $\beta$ -D-glucopyranose. Glucose Oxidase specifically binds to  $\beta$ -D-glucopyranose and not to  $\alpha$ -D-glucopyranose. Still, glucose oxidase is able to catalyze the oxidation of all the glucose in solution, since the equilibrium between  $\alpha$ -D-glucopyranose and  $\beta$ -D-glucopyranose is driven towards the side of  $\beta$ -D-glucopyranose as it is consumed.

Glucose oxidase catalyzes the reaction of  $\beta$ -D-glucopyranose to D-glucono- $\delta$ -lactone and hydrogen peroxide ( $H_2O_2$ ). Glucono- $\delta$ -lactone is pH neutral, but hydrolyses in water to gluconic acid. In aqueous solutions at neutral pH, gluconic acid then forms the gluconate ion by losing one  $H^+$ .

Since hydrogen peroxide is a reactive oxygen species, it causes damage to proteins, particularly unfolded peptide chains. Therefore, the enzyme catalase is added to catalyze the decomposition of hydrogen peroxide ( $H_2O_2$ ) into water ( $H_2O$ ) and oxygen ( $O_2$ ):

The equations of the chemical reaction read as follows:



Structural representation of the reaction including terminology:

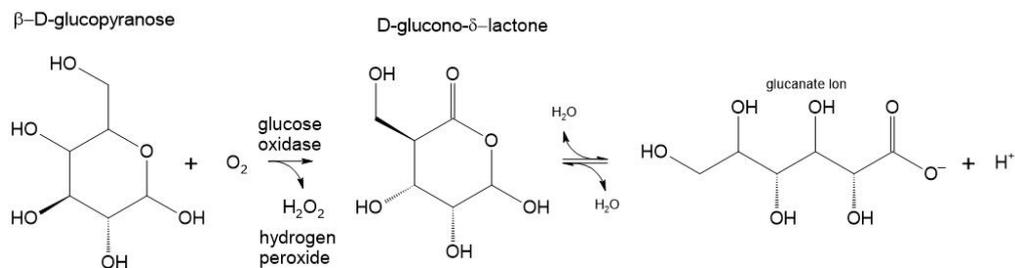


Figure 48: The chemical reaction of the GOC scavenging system.

While in the oxidation reaction (first line in equation 1) one  $H_2O_2$  is produced for every consumed  $O_2$ , in the catalase reaction (second line in equation 1) one  $O_2$  is produced for every two consumed  $H_2O_2$ . Therefore, the net result regarding the scavenging of  $O_2$  is the consumption of one  $O_2$  molecule for every cycle through both of these reactions.

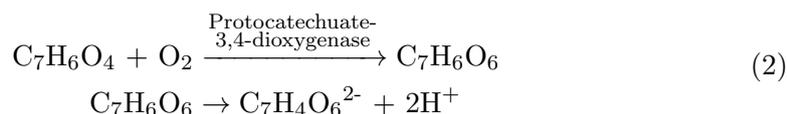
Notably, one of the reaction products is gluconic acid, which reduces pH.

### 7.2.3 Protocatechuate-3,4-dioxygenase/Protocatechuic Acid

The scavenging system consisting of protocatechuate-3,4-dioxygenase and protocatechuic acid (PCD) [123] represents an alternative to the more common GOC. It has been shown to improve the stability of fluorescent dyes in single-molecule experiments as compared to GOC ([124]).

PCD catalyzes the reaction between PCA and oxygen, which produces 3-Carboxy-cis-muonic acid ( $C_7H_6O_6$ ). This then decomposes into two protons and the remaining ( $C_7H_4O_6^{2-}$ ).

The equations of the chemical reaction read as follows:



Structural representation of the PCD/PCA reaction including terminology:

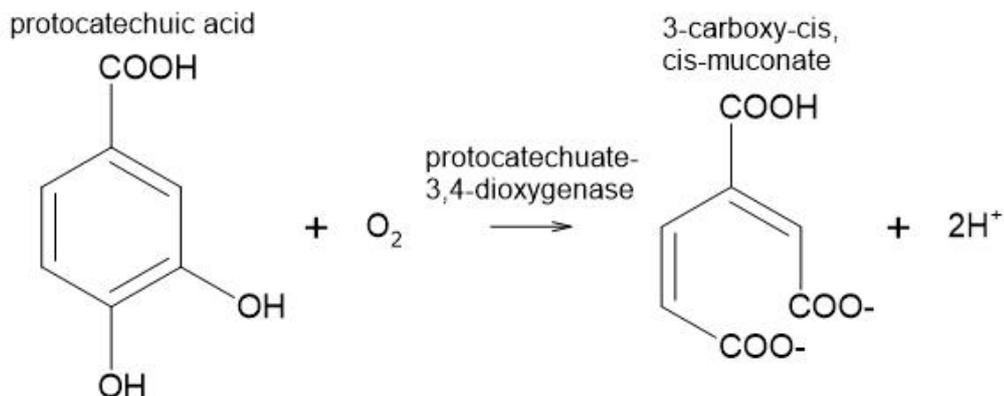


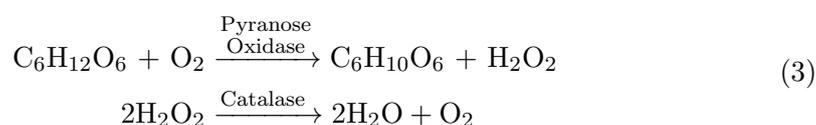
Figure 49: The chemical reaction of the PCD/PCA scavenging system.

### 7.2.4 Pyranose Oxidase/Glucose/Catalase

A further scavenging system consists of pyranose oxidase, catalase and glucose (POC). As in the case of GOC, it works via the oxidation of glucose, which in this case is catalyzed by pyranose oxidase, producing hydrogen peroxide and 2-dehydro-D-glucose ( $C_6H_{10}O_6$ ). It is important to note that 2-dehydro-D-glucose ( $C_6H_{10}O_6$ , the product of oxidation by pyranose oxidase) and D-glucono-1,5-lactone ( $C_6H_{10}O_6$ , the product of oxidation by glucose oxidase) are identical in their chemical formula, but differ in their structure and chemical properties, as discussed in the following section (7.2.5).

Again, in a second reaction, catalase catalyzes the decomposition of hydrogen peroxide ( $H_2O_2$ ) into water ( $H_2O$ ) and oxygen ( $O_2$ ).

The equations of the chemical reaction read as follows:



Structural representation of the reaction including terminology:

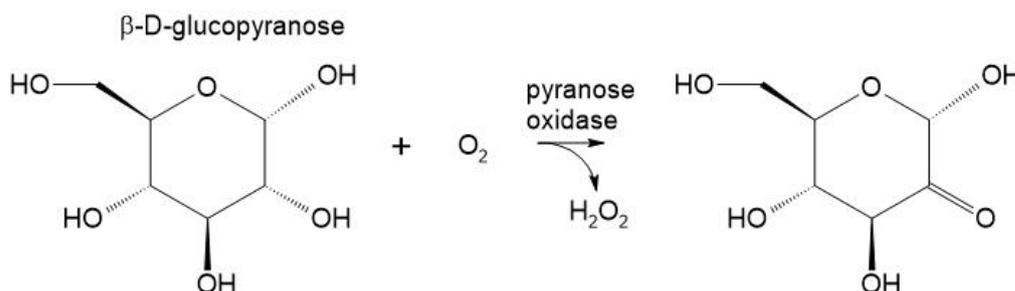


Figure 50: The chemical reaction of the POC scavenging system.

### 7.2.5 Scavenging Systems Discussion

The standard oxygen scavenging system used in single-molecule optical trapping experiments is GOC. For most molecules, this scavenging system is adequate and shows good results. When Suren started working with the GR-LBD, he also used GOC as the scavenging system. In these early experiments, the GR-LBD showed a markedly high tendency to misfold (cf. section 7.2.1). Recording long traces or observing complete force-induced unfoldings followed by refolding of the molecule were practically impossible under these conditions.

In search of a modification of the assay that would lead to a more reliable and stable behavior of the GR-LBD in optical trapping experiments, Suren tried the PCD scavenging system (cf. section 7.2.3) [123]. This oxygen scavenging system had been shown to improve the stability of fluorescent dyes in single-molecule experiments, as compared to the GOC scavenging system. When applied to the GR-LBD assay, it led to mixed results, with some measurements working better than with the GOC scavenging systems, while others went equally badly.

Eventually, the actual reason for the particular biochemical fragility of the GR-LBD in our optical trapping assay was discovered as part of this study, collecting evidence

from multiple different clues.

The first hint came from the competition assay discussed in section 6.2, which required the dilution of the "lid peptide" in 3% ammonia water due to its high hydrophobicity. Since ammonia is strongly basic and thereby increases the pH of the sample, we balanced the pH using hydrochloric acid. In the following experiments, the GR-LBD suddenly performed unusually well, with measurements lasting over an hour (for the first time ever) and involving hundreds of ligand dissociations and rebindings as well as partial and complete unfoldings. This was the first clue that the pH-value of the sample and strong buffering conditions might be the key to a more reliable behavior of the GR-LBD.

A further hint pointing into the same direction was hidden in our protocols on the different dilutions of scavenging systems we tried. The PCD/PCA scavenging system never convincingly worked until we began diluting the PCA in 10-fold TAE buffer, even though the measurement was conducted in HEPES buffer at 40 mM HEPES. Since the added PCA in solution usually made up about 1/7 of the entire sample, this increased the buffering strength of the solution by more than a factor of 2. In effect, two different kinds of buffers, HEPES and TAE, were mixed in these experiments. We first ascribed the improved measurement quality to the fact that the acidic PCA works better when diluted in a strongly buffered solution. But the crucial side-effect was that, along the way, the buffering capacity of the entire sample was increased.

All of this suggested that the pH value of the sample played a crucial role for the behavior of the GR-LBD. Both scavenging systems used in the GR-LBD project until then, namely GOC and PCD/PCA, shared a decisive disadvantage: The chemical reactions involved, while fulfilling their purpose of depriving the solution of oxygen, both have an acidic end product. This progressively and irreversibly decreases the pH value in the sample. The POC scavenging system (cf. section 7.2.4) on the other hand does not result in an acidic end product, thereby leaving the pH of the sample unaffected. Fig.51 gives an overview over the reactions of the three different scavenging systems.

Using either GOC or PCD, the pH of the sample can drop drastically already within the first minutes of experiment, while it remains constant when using POC. Fig.52 illustrates this effect for all three scavenging systems, in each case starting at pH values of 7.0, 7.5 and 8.0.

In early experiments with GR-LBD using GOC, the buffer was a HEPES buffer at pH 7.2. Due to the experimental preparations necessary for an optical trapping experiment, the earliest point in time when a molecule can possibly be measured is about 5 minutes after the addition of the scavenging system to the sample. 15-30 minutes is a more realistic timespan. According to the graph in Fig.52, the pH of the sample under our experimental conditions must always have dropped below 6.5 within the first 10 minutes. Therefore, the GR-LBD molecules measured with GOC and PCD must have been measured at markedly low pH values.

After this realization, the assay was modified in several ways. Instead of GOC or PCD/PCA, the POC scavenging system was used from then on. On top of that, the pH value at which the sample was measured was increased from 7.2 to 8.0. The previously used HEPES buffer has a  $p_K = 7.5$  and works best over the pH range between 6.8 and 8.2. For a pH value of 8.0, Tris buffer is better suited, with a  $p_K = 8.07$  and a buffering range between 7.1 and 9.1. Therefore, the measurement buffer

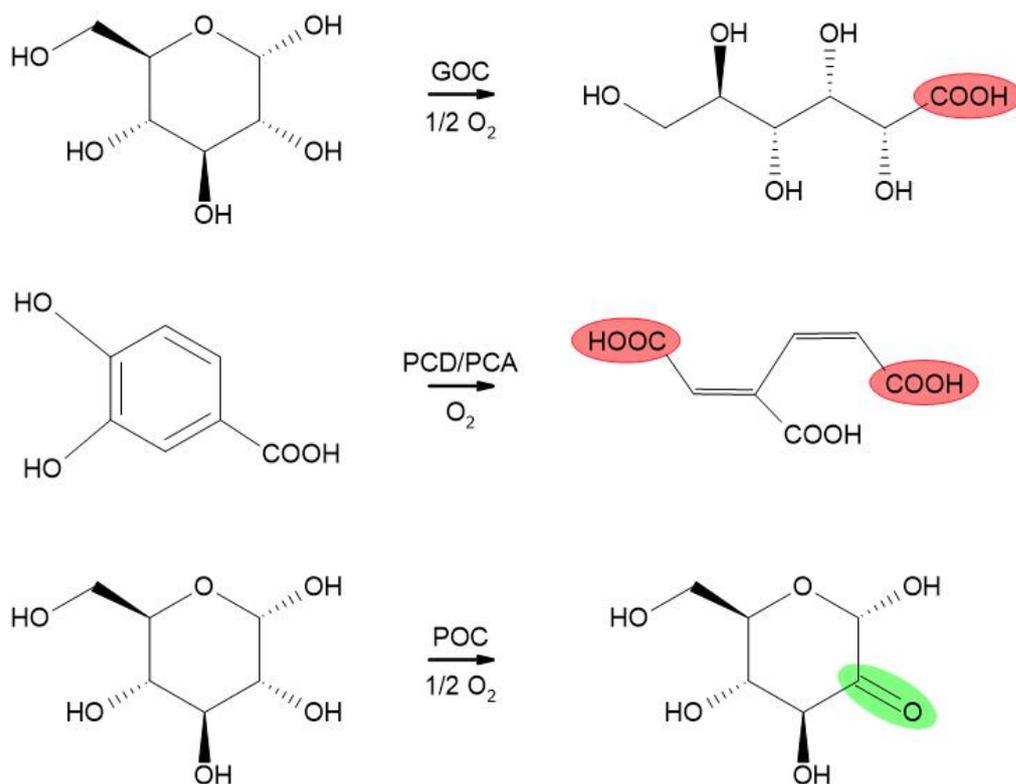


Figure 51: While the GOC and PCD scavenging systems both have acidic end products (red), the POC scavenging system does not (green). Note that the end products of PCD and POC are composed of identical sets of atoms, but have different structures.

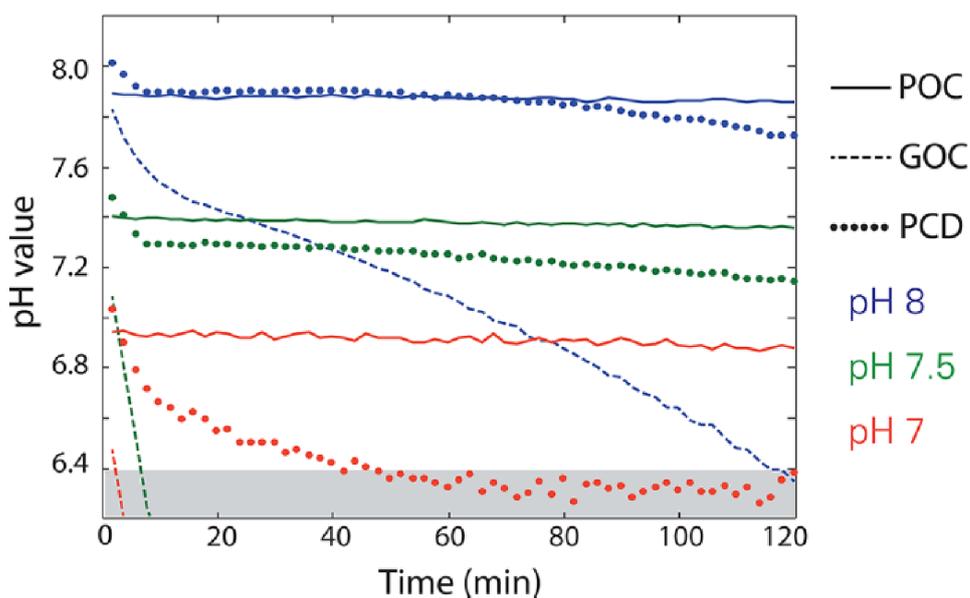


Figure 52: Both GOC and PCD lead to a fast drop in pH of the sample, regardless of the initial pH value. Only POC leaves the pH value of the sample unaffected. Figure reprinted from [125] with permission. Copyright 2012 by ACS Nano.

was changed from HEPES to Tris, and the pH was from then on set to 8.0. Since this pH of 8.0 is higher than the pH in the intracellular environment (pH 7.2), a literature search was performed, listing publications that worked with GR-LBD and used Tris buffers at pH 8.0 or higher ([126] [127] [128]). The GR-LBD did not show any adverse effects in these studies with comparable buffering conditions.

**Buffer before switching scavenging system:**

40mM HEPES  
 150 mM NaCl  
 pH 7.5  
 5 mM PCA  
 0.6 units PCD/ml  
 NaOH as needed to adjust the pH value

**Buffer after switching scavenging system:**

50 mM Tris  
 150 mM NaCl  
 pH 8.0  
 8 units pyranose oxidase/ml (stock has 400 units/ml, dilute 1:50 in sample)  
 33% Glucose diluted 1:50 in sample  
 17000 units catalase/ml (main stock has 85000 units/ml, dilute 1:50 in sample)

These modifications of the assay marked the breakthrough in experiment quality for the GR-LBD project. Fig.53 shows the strikingly stable unfolding and refolding behavior of the GR-LBD even during extremely long dwell times in the completely unfolded state: From then on, the assay worked consistently well for the last three years and up to this day, never relapsing into its initial hickups.

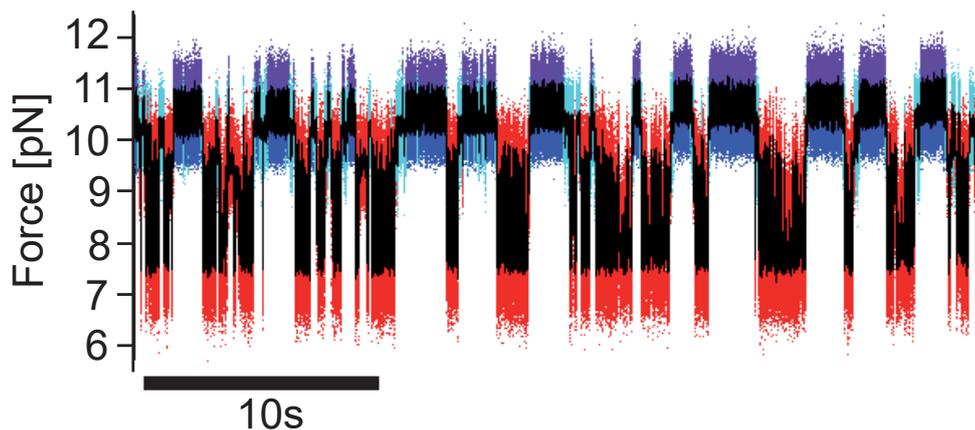


Figure 53: Using the POC scavenging system, the GR-LBD proved to be remarkably reliable concerning its refolding capacity even after long dwell times in the partially or completely unfolded state. The trace shown exhibits at least 15 complete unfoldings and refoldings, including DEX rebinding, over the course of 40s.

The natural follow up question is, why the pH value plays such an important role for the behavior of the GR-LBD.

One conceivable explanation considers the iso-electric point (pI) of a protein, i.e.

the pH value at which the net charge of a molecule is zero. When the pH value lies under the pI, the protein's net charge is positive, while at pH above the pI, its net charge is negative, both leading to repellant forces between the charged molecules. Proteins exhibit the least solubility at their pI, where there are no repellant forces.

Three different algorithms were used to calculate the GR-LBD's pI: The isoelectric point calculator from [129] renders  $pI_{GR-LBD} = 6.25$ . Another algorithm from [130] gives  $pI_{GR-LBD} = 6.58$ . A third algorithm from [131] yields  $pI_{GR-LBD} = 6.62$ .

It is conceivable that the pH drop due to the scavenging system was sufficient to reach a pH value of  $\sim 6.5$  (cf. Fig.52), which would result in zero net charge of GR-LBD and lead to a particularly poor solubility. This might then increase the already intrinsically high tendency of GR-LBD to aggregate. In a single-molecule assay, this might lead to unwanted intermolecular interactions between a tethered GR-LBD and GR-LBDs floating in solution, in particular if the GR-LBD is in an unfolded state. However, there is a study suggesting that the optimum pH value for stability as well as activity of a molecule do not correlate with the isoelectric point [132].

Another possible explanation might be that too low pH values affect the interaction between side-chains in the secondary and tertiary structure of the molecule, particularly salt bridges and hydrogen bonds. The pH-value is a logarithmic measure of the concentration of free hydrogen ions ( $H^+$ ). Lower pH means a higher concentration of  $H^+$ . Changes in pH affect the attractions between side chain groups, since these interactions rely, among others, on salt bridges and hydrogen bonds. Both salt bridges and hydrogen bonds are sensitive to changes in pH. In salt bridges, the attraction between ionic side chain can disappear when, due to a too low (or too high) pH, one of the ionic groups becomes neutral (too low pH turns the negative ionic group neutral, too high pH turns the positive ionic group neutral). Hydrogen bonds can be described as an electrostatic dipole-dipole interaction, which can also be disturbed by low pH. Since the interactions between the side chains determine the shape of the molecule, extreme pH values can denature the protein. How sensitive a molecule is to pH depends on its respective secondary and tertiary structure. Apparently, the GR-LBD is particularly sensitive to pH values already slightly lower than 7.2.

## 8 Interaction of the GR-LBD with Hsp70/40

### 8.1 Hsp70/40 Induces Complete and Stepwise Unfolding of GR-LBD

Hsp70/40 shows a drastic effect on the GR-LBD in optical trapping experiments. At chaperone concentrations of 10  $\mu\text{M}$  Hsp70, 2  $\mu\text{M}$  Ydj1, 5 mM  $\text{MgCl}_2$  and 5 mM ATP (which is in the physiological range [133] [134]), a complete, stepwise and irreversible unfolding of the DEX-unbound GR-LBD occurs (Fig.54).

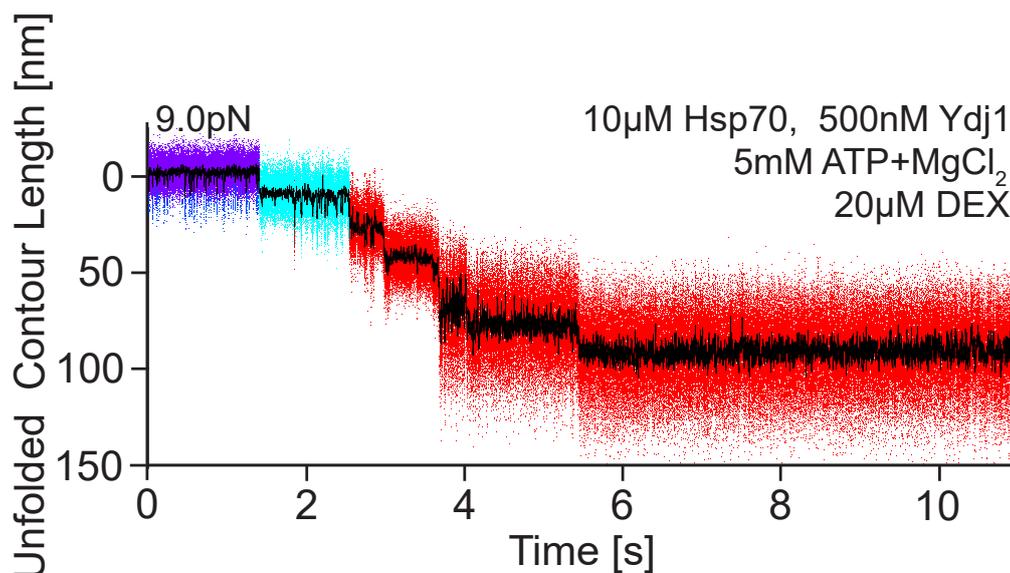


Figure 54: Hsp70/40 unfolds the DEX-unbound GR-LBD in a stepwise manner within  $\sim 3$  s. Purple/dark blue transitions: lid flipping, light-blue: DEX-unbound state, red: chaperone-induced unfolding intermediates.

At  $t = 0$ , the GR-LBD starts in the folded holo state (purple/dark blue flipping transitions). After the first ligand dissociation at  $\sim 2.5$  s (transition to light-blue state), chaperone unfolding sets in and the GR-LBD unfolds from the apo state to the entirely unfolded peptide chain via five sequential unfolding intermediates (red phase with steps). The entire unfolding occurs over the course of only  $\sim 3$  s. Afterwards, the GR-LBD remains completely unfolded for the last  $\sim 5$  s of the trace.

Note that the force bias in the passive-mode trace in Fig.54 lies at  $\sim 9.0$  pN and is, therefore, lower than the force bias in previously shown passive-mode traces (such as in Fig.25, for example, where the force bias lies at  $\sim 10.0$  pN). Spontaneous force-induced unfoldings at these forces were rare and never led to a complete, let alone irreversible, unfolding of GR-LBD. Hence, they can be excluded in Fig.54.

The upcoming sections will elaborate on the details of this chaperone-induced unfolding.

### 8.2 Lid Flipping Remains Unaffected by Chaperones

To test whether Hsp70/40 can bind to the N-terminal lid, the force-dependent lid flipping rates in the presence and in the absence of chaperones were compared. Fig.55 shows the result of this analysis.

Flipping rates remain the same in the presence of chaperones (red for Ydj1, green

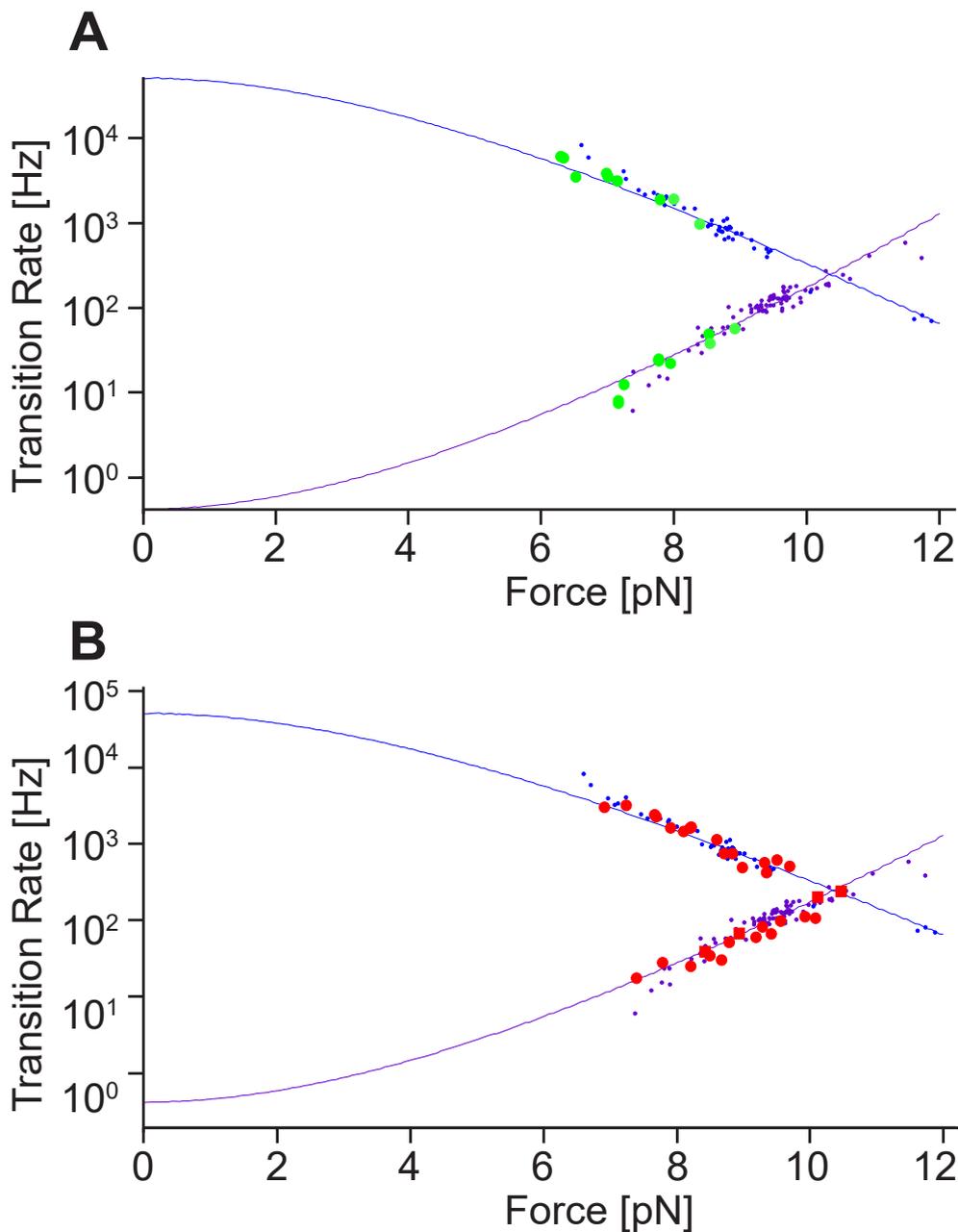


Figure 55: **A)** Flipping rates remain the same in the presence of the full chaperone system ( $10\ \mu\text{M}$  Hsp70,  $2\ \mu\text{M}$  Ydj1,  $5\ \text{mM}$   $\text{MgCl}_2$ ,  $5\ \text{mM}$  ATP, green data points) as in the absence of chaperones (blue and purple data points, adapted from previous analysis by Suren [44]). **B)** Flipping rates remain the same in the presence of  $4\ \mu\text{M}$  Ydj1 (red data points) as in the absence of Ydj1 (blue and purple data points, adapted from previous analysis by Suren [44]).

for full Hsp70/Hsp40 chaperone system) as in the absence of chaperones (purple and blue, adapted from Suren's analysis in [44]). This implies that Hsp70 does not bind to the N-terminal lid itself, nor to the part of the folded core that is in direct contact with the lid, since both scenarios would lead to unusually long lid-open phases or different lid kinetics in general.

### 8.3 Kinetics of the Chaperone-Induced GR-LBD Unfolding

To investigate the chaperone-induced unfolding kinetics of GR-LBD, the Hsp70 and Hsp40 concentrations were varied systematically.

#### 8.3.1 Variation of Hsp70 Concentration while Holding Hsp40 Concentration Constant

Fig.56 shows Hsp70/40 induced unfoldings of GR-LBD at three different Hsp70 concentrations. The Ydj1 concentration was 2  $\mu$ M in all cases.

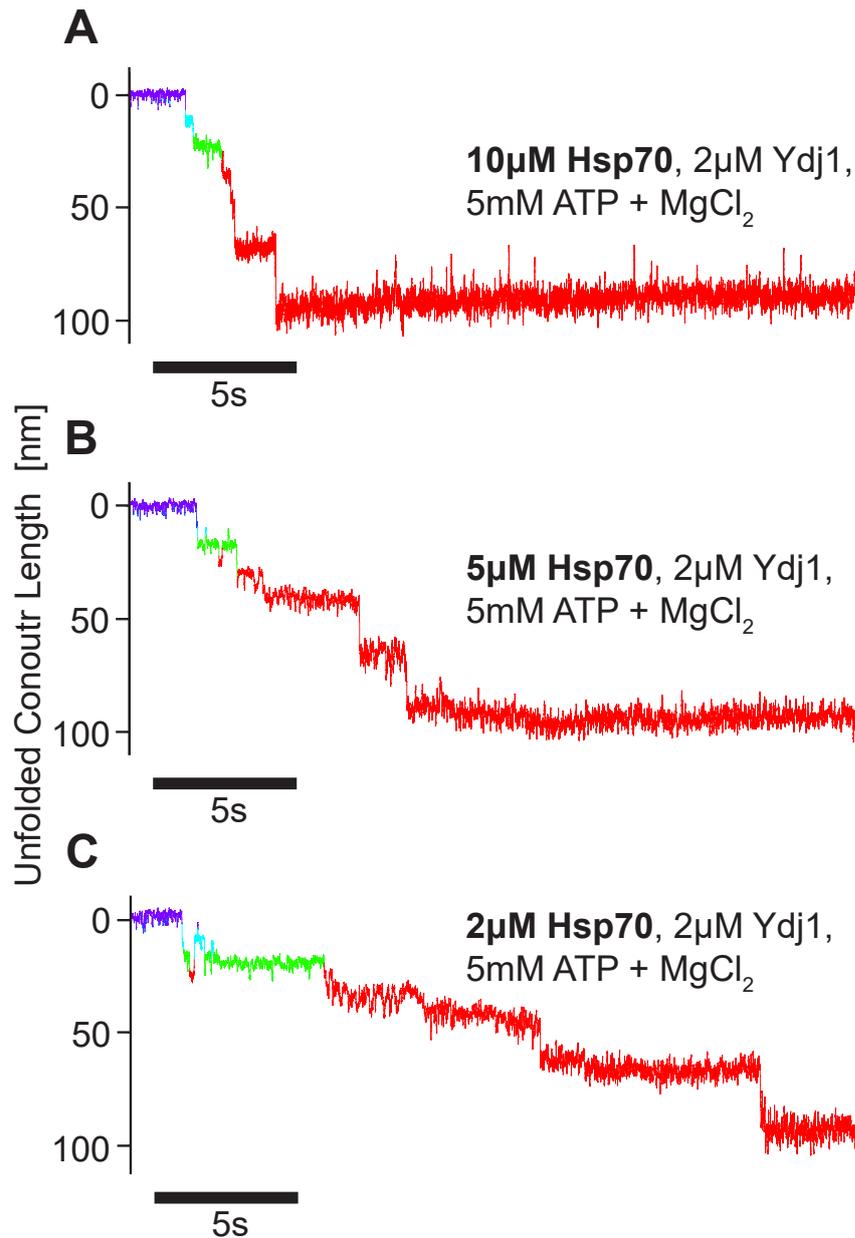


Figure 56: The dwell-times on unfolding intermediates scale inversely with Hsp70 concentration. **A)** 10  $\mu$ M **Hsp70**, 2  $\mu$ M Ydj1, 5 mM MgCl<sub>2</sub>, 5 mM ATP. **B)** 5  $\mu$ M **Hsp70**, 2  $\mu$ M Ydj1, 5 mM MgCl<sub>2</sub>, 5 mM ATP. **C)** 2  $\mu$ M **Hsp70**, 2  $\mu$ M Ydj1, 5 mM MgCl<sub>2</sub>, 5 mM ATP. The green state is a Ydj1-bound state explained in section 75.

As Fig.56 illustrates, both the entire unfolding time (from the light-blue DEX-unbound state to the completely unfolded peptide chain) as well as the dwell times on the individual unfolding intermediates scale inversely with Hsp70 concentration. The higher the Hsp70 concentration, the shorter both the entire unfolding time as well as the intermediate dwell times become, and vice versa.

The analysis of  $\sim 200$  dwell times on unfolding intermediates at Hsp70 concentrations ranging from 500 nM to 30  $\mu$ M Hsp70 resulted in the graph given in Fig.57:

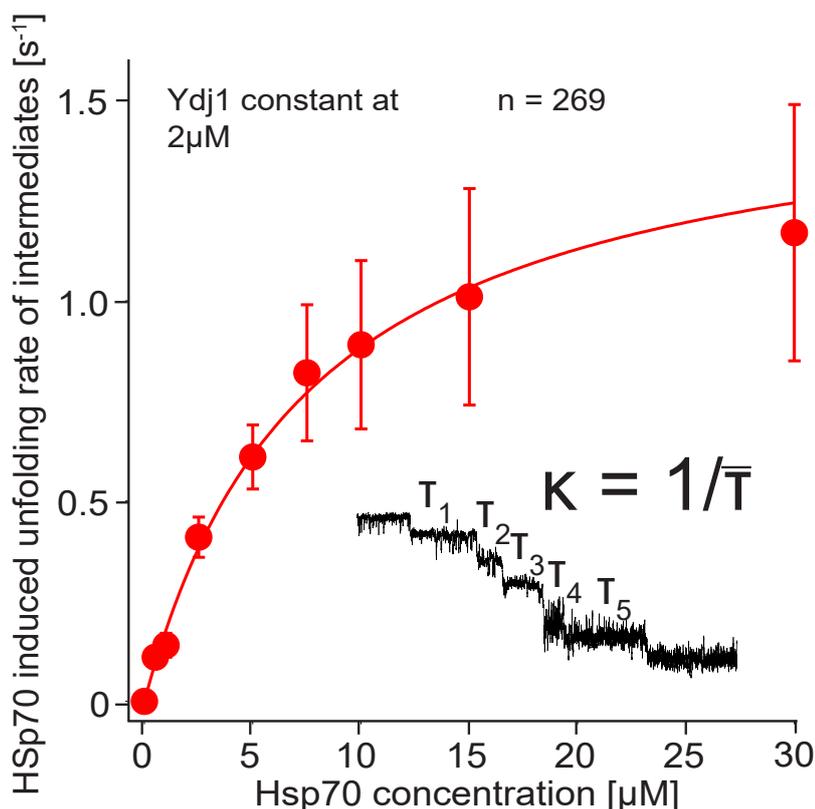


Figure 57: Unfolding rates (inverse of average dwell time per step) vs. Hsp70 concentration at a constant Ydj1 concentration of 2  $\mu$ M.

At low Hsp70 concentration, we observe an approximately linear dependence between the unfolding rate from one intermediate to the next and the Hsp70 concentration, while towards high concentrations, the curve saturates.

### 8.3.2 Variation of Hsp40 Concentration while Holding Hsp70 Concentration Constant

For the Hsp40 variation experiments, we chose a truncated version of Hsp40 consisting only of the J-Domain and the G/F-rich region of Ydj1 (JD). We had found that high concentrations of full length Hsp40 (both Ydj1 and Hdj2) interfered destructively with the experiments because of its binding to the substrate (for details, see section 10). Using the JD, we could eliminate the competition for binding sites between Hsp70 and Hsp40 and exclusively study the effect of the JD on Hsp70's ATP hydrolysis rate.

With JD as Hsp40, the Hsp70/40 induced unfolding of GR-LBD still worked. Fig.58

shows Hsp70/40 induced unfoldings of GR-LBD at three different JD concentrations. The Hsp70 concentration was kept constant at 10  $\mu$ M in all cases. The fact that  $\sim$ 50-fold higher concentrations of JD than of Ydj1 were needed to achieve equal unfolding rates has been reported before [38] [135] and will be discussed in section 10.

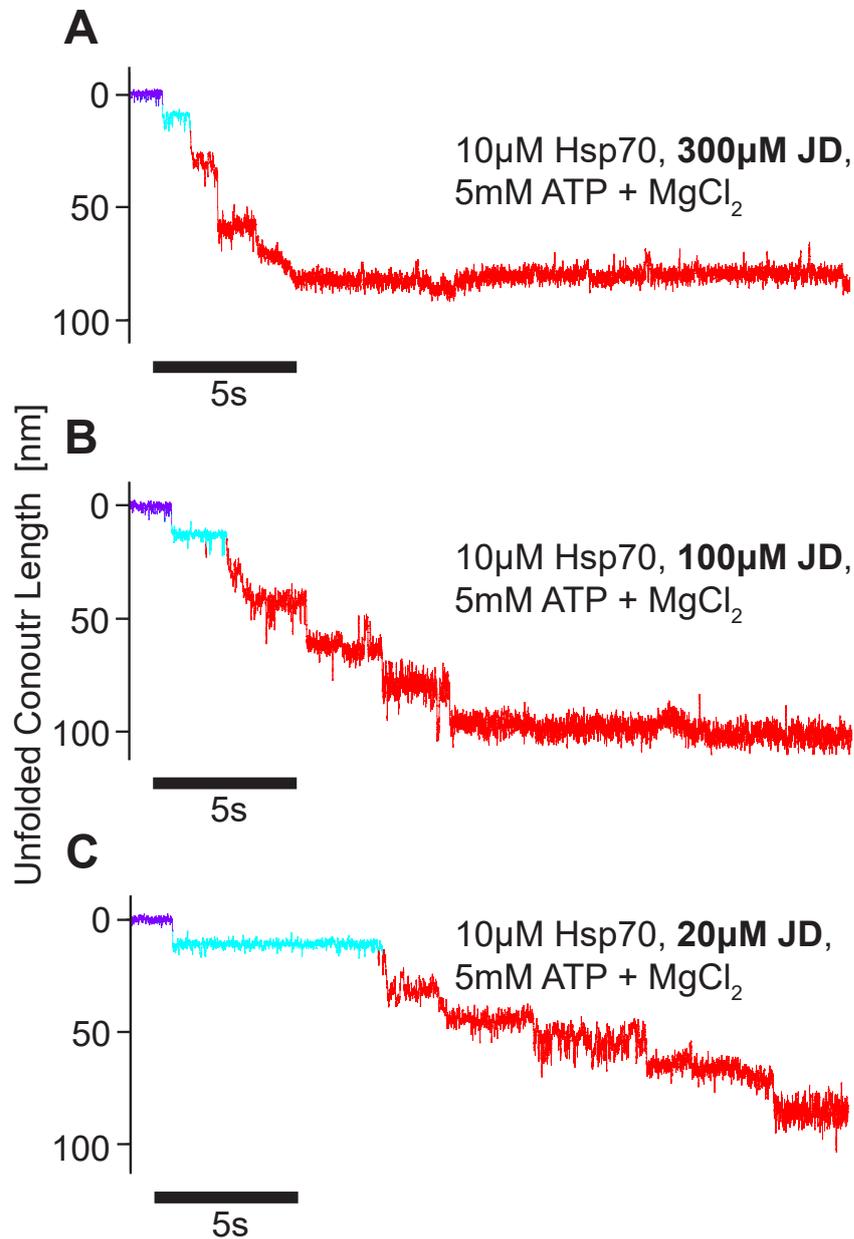


Figure 58: Hsp70/40 induced complete unfolding of GR-LBD at varying JD concentration, holding Hsp70 concentration constant at 10  $\mu$ M. Dwell-times on unfolding intermediates scale inversely with JD concentration. **A)** 10  $\mu$ M Hsp70, 300  $\mu$ M JD, 10 mM MgCl<sub>2</sub>, 50 mM ATP. **B)** 10  $\mu$ M Hsp70, 100  $\mu$ M JD, 5 mM MgCl<sub>2</sub>, 5 mM ATP. **C)** 10  $\mu$ M Hsp70, 20  $\mu$ M JD, 5 mM MgCl<sub>2</sub>, 5 mM ATP.

Again, both the entire unfolding time (from the light-blue DEX-unbound state to the completely unfolded peptide chain) as well as the dwell times in the individual unfolding intermediates scaled inversely with Hsp40 concentration.

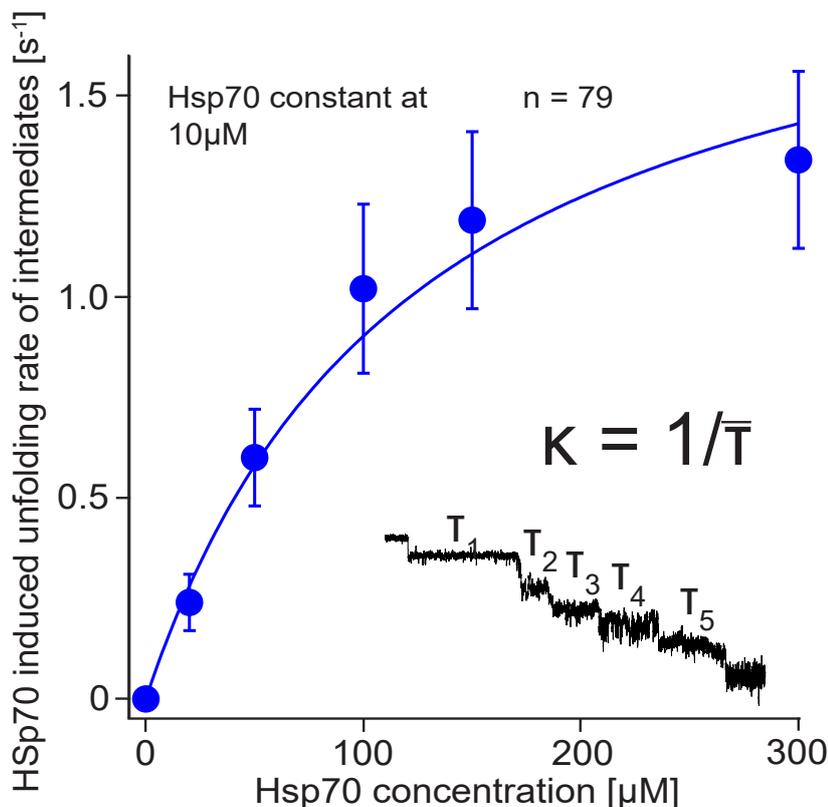


Figure 59: Unfolding rates (inverse of average dwell time per step) vs. J-Domain concentration at a constant Hsp70 concentration of 10 μM.

#### 8.4 Discussion of Chaperone-Induced Unfolding Kinetics

Several conclusions can be drawn from the concentration variation experiments with Hsp70 and Hsp40:

When varying the Hsp70 concentration and holding the Hsp40 concentration constant, the unfolding rate between intermediates scaled with Hsp70 concentration. This implies that each unfolding step requires the binding of at least one Hsp70. At non-saturating Hsp70 concentrations, the dependence of the unfolding rate between individual unfolding intermediates on the Hsp70 concentration is approximately linear. This suggests that it is one Hsp70 binding event per unfolding step. At high Hsp70 concentrations, the curve in Fig.57 saturates, because the ATP hydrolysis rate at the given constant Hsp40 concentration becomes limiting.

The maximum number of observed Hsp70/40 induced unfolding steps reproducibly lay at 5 (cf. Figs. 54, 56, 58). If one accepts the assumption that it is one Hsp70 binding event per unfolding step, this suggests that the maximum number of Hsp70s involved in the unfolding of GR-LBD is 5. It is possible that there are, in fact, always 5 Hsp70s involved and the individual steps simply do not always show up clearly in our optical trapping experiments. The variability in the quality of our traces is likely related to Hsp70 binding to different sites from N and C-terminus as well

as unfolding steps driven by mechanical force and backward steps through successful refolding attempts of GR-LBD. Running the LIMBO algorithm [136] designed to predict Hsp70 binding sites over the sequence of GR-LBD results in 5 high-likelihood binding sites (cf. section 8.9).

The dwell time on each individual unfolding intermediate also scaled inversely with Hsp40 concentration. For these experiments, we used the JD construct, which consists of only the J-Domain of yeast Hsp40. Since JD is lacking any kind of substrate binding domain, its only possibility to affect the assay is by stimulating Hsp70's hydrolysis rate. If the JD concentration affects the dwell time on each individual unfolding intermediate, this suggests that each unfolding transition from one intermediate to the next must involve ATP hydrolysis by Hsp70, stimulated by Hsp40. Again, at non-saturating Hsp40 concentrations, the dependence of the unfolding rate between intermediates on the Hsp40 concentration appeared to be linear, while the curve saturates at high Hsp40 concentrations. In this case, probably both the ATP hydrolysis rate as well as the binding rate of Hsp70 become limiting, since Hsp70 concentration is held constant.

Taking into account the dependence of unfolding rate on both Hsp70 and Hsp40 concentration and the asymptotic saturation behavior in both graphs (Fig.57 and Fig.59) at high concentrations, a simple model of the process was derived. It assumes that each transition from one unfolding intermediate to the next requires two processes:

1. Binding of one Hsp70
2. ATP hydrolysis of Hsp70 stimulated by Hsp40

Mathematically, this results in the following relation:

$$k_{unfold}(C_{Hsp70}, C_{Hsp40}) = \frac{1}{\frac{1}{k_{on,Hsp70} \cdot C_{Hsp70}} + \frac{1}{k_{on,Hsp40} \cdot C_{Hsp40}} + \frac{1}{k_{max,hydrolysis}}} \quad (4)$$

With:  $k_{unfold}$ : unfolding rate from one intermediate to the next

$C_{Hsp70}$ : Hsp70 concentration

$C_{Hsp40}$ : Hsp40 concentration

$k_{on,Hsp70}$ : Hsp70 binding rate to GR-LBD

$k_{on,Hsp40}$ : Hsp40 on-rate to Hsp70, in order to stimulate ATP hydrolysis. In the case of Ydj1, this could be intertwined with an on-rate to GR-LBD

$k_{max,hydrolysis}$ : fastest theoretically possible hydrolysis rate. A potential delay between ATP hydrolysis and unfolding transition is lumped into this term as well.

The fits to the graphs in Fig.57 and Fig.59 are global fits of the above model to both data sets. They yielded the following parameters:

$$k_{on,Hsp70} = 0.20 \pm 0.03 s^{-1} \mu M^{-1};$$

$$k_{on,Ydj1} = 0.78 \pm 0.20 s^{-1} \mu M^{-1}$$

$$k_{on,JD} = 0.016 \pm 0.004 s^{-1} \mu M^{-1}$$

$$k_{max,hydrolysis} > 1000 s^{-1}$$

## 8.5 Hsp70/40 Inhibits Refolding of GR-LBD

After ATP hydrolysis, Hsp70's SBD has closed down on substrate. In the absence of nucleotide exchange factors, ADP dissociation rate from this state is slow, around  $0.004\text{--}0.035\text{ s}^{-1}$  [137]. Only after ADP dissociation and ATP binding the SBD opens and releases its substrate again. This way, Hsp70 is able to hold its substrate in an unfolded state for extended periods of time, preventing it from (mis-)folding. Hence, Hsp70 is often referred to as a "holdase".

For the experiments with GR-LBD, this implies that after GR-LBD has been completely unfolded by Hsp70, it is decorated with multiple Hsp70s, which remain bound to it for a certain timespan. If the concentration of Hsp70 is high enough, whenever an Hsp70 is dissociating from the GR-LBD's peptide chain, a new one can take its place. The refolding of GR-LBD is inhibited by the bound Hsp70s. This becomes apparent both in stretch-relax cycles as well as in passive-mode experiments, as the following sections will elaborate.

### 8.5.1 Hsp70/40 Inhibiting Refolding of GR-LBD in Stretch-Relax Experiments

Fig.60 shows a comparison of stretch-relax cycles in the absence and in the presence of physiological chaperone concentrations. In the trace in Fig.60A, ten consecutive stretch-relax cycles of the same GR-LBD molecule exhibit a mixture of ligand-bound unfoldings (high unfolding force peaks, cycles 1, 4, 8 and 10) and ligand-unbound unfoldings (low unfolding force peaks, 2, 3, 5, 6, 7 and 9). Note that, in every cycle, the GR-LBD folds back completely to the apo lid-open state.

Upon addition of Hsp70/40 (Fig.60B), the situation changed drastically. Now, only the first pull shows the signature unfolding trace of holo GR-LBD (cycle 1, blue trace). Already in the first relaxation trace (cycle 1, red trace) the molecule cannot refold to the apo lid-open state, indicating chaperones promptly interfered with folding. After five further stretch-relax cycles, no part of the molecule is able to fold anymore. The stretch-relax cycles now correspond to the worm-like chain traces of the completely unfolded peptide chain of GR-LBD. These experiments show that, after the protein was unfolded using mechanical force, the chaperones quickly bind to the unfolded peptide chain, efficiently blocking refolding of GR-LBD.

### 8.5.2 Hsp70/40 Inhibiting Refolding of GR-LBD in Passive-Mode Experiments

The inhibition of the GR-LBD's refolding by Hsp70/40 also becomes apparent in passive-mode experiments. At Hsp70 concentrations  $>5\text{ }\mu\text{M}$  and Ydj1 concentrations  $>1\text{ }\mu\text{M}$ , the unfolding is usually irreversible, i.e. the GR-LBD remains unfolded indefinitely and can never refold. Long traces of  $>400\text{ s}$  (Fig.61) show that, after the initial stepwise unfolding, the GR-LBD remains unfolded. Low frequency fluctuations are due to drift. Occasionally, some part of the GR-LBD seems to be able to refold temporarily, before being unfolded by Hsp70 anew. However, GR-LBD never reaches the DEX-unbound state again.

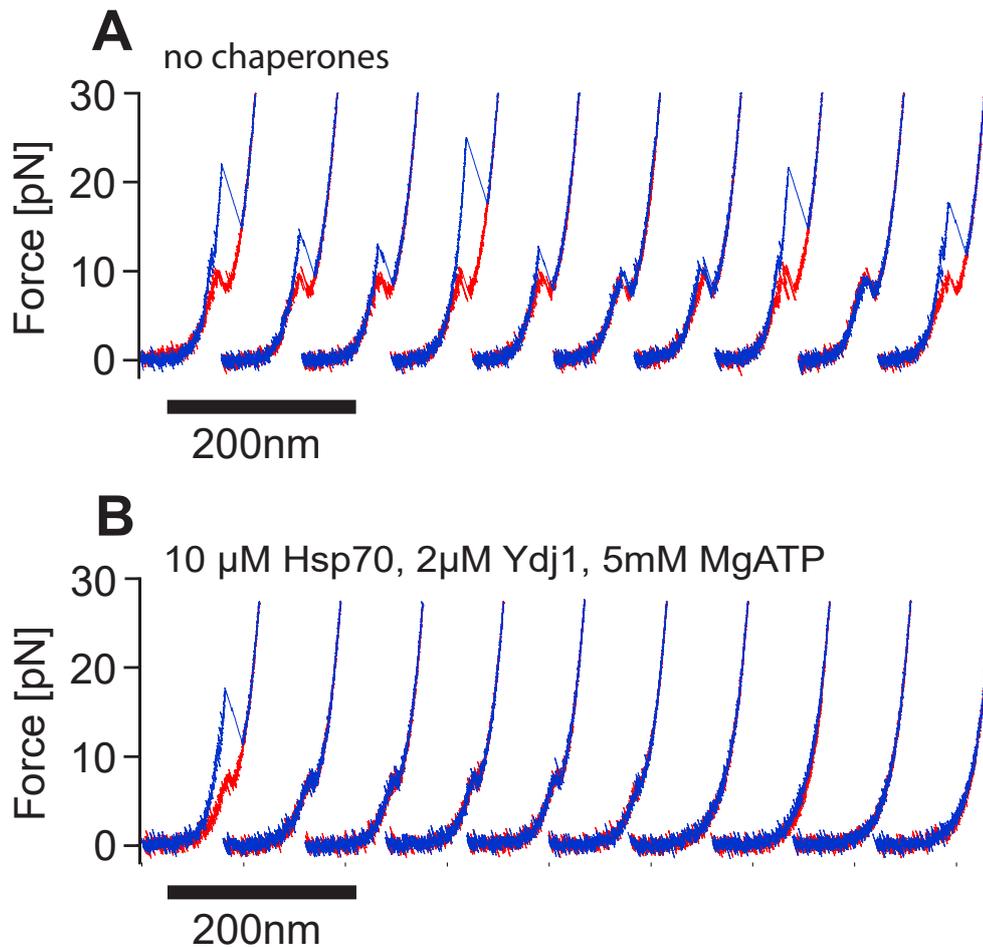


Figure 60: **A**) Stretch-relax cycles of GR-LBD in the absence of chaperones, with DEX-bound (cycles 1, 4, 8 and 10) as well as DEX-unbound (cycles 2, 3, 5, 6, 7 and 9) unfoldings. **B**) Stretch-relax cycles of GR-LBD in the presence of chaperones. Only the first cycle exhibits a native DEX-bound unfolding, while all the following cycles are held in the first partially (cycles 2, 3, 4, 5 and 6) and then completely unfolded state (cycles 7, 8, 9 and 10) by the Hsp70/40 chaperone system. The GR-LBD remains completely unfolded indefinitely in such pulls.

## 8.6 BAG1 Counteracts the Unfolding by Hsp70/40

Nucleotide exchange factors such as BAG1 accelerate the release of ADP from Hsp70 in its closed conformation. Once ADP has dissociated, a new ATP can bind to Hsp70, inducing the opening of Hsp70's substrate binding domain. As a result, the presence of BAG1 also indirectly increases the off-rate of Hsp70 from substrate.

This function of the nucleotide exchange factor BAG1 could be clearly observed in our optical trapping experiments. In Fig.62, the addition of BAG1 to Hsp70/40 showed a striking effect. The initial unfolding proceeds equally fast as, for example, in Fig.61, but is immediately followed by refolding attempts of GR-LBD. It refolds several times to the light-blue ligand unbound state and at  $\sim 250$  s (arrow) even rebinds ligand in spite of the presence of high concentrations of chaperones. This never happened in the absence of BAG1.

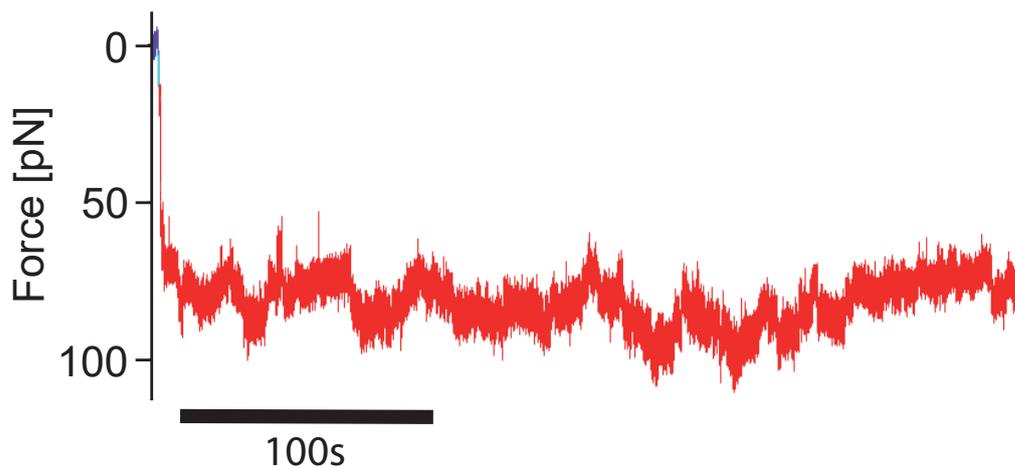


Figure 61: After an initial Hsp70/40 induced unfolding the GR-LBD remains unfolded indefinitely. This explains the fact that, in the presence of high chaperone concentrations, the GR-LBD is unable to rebind DEX [43]. The low frequency fluctuations are due to drift.

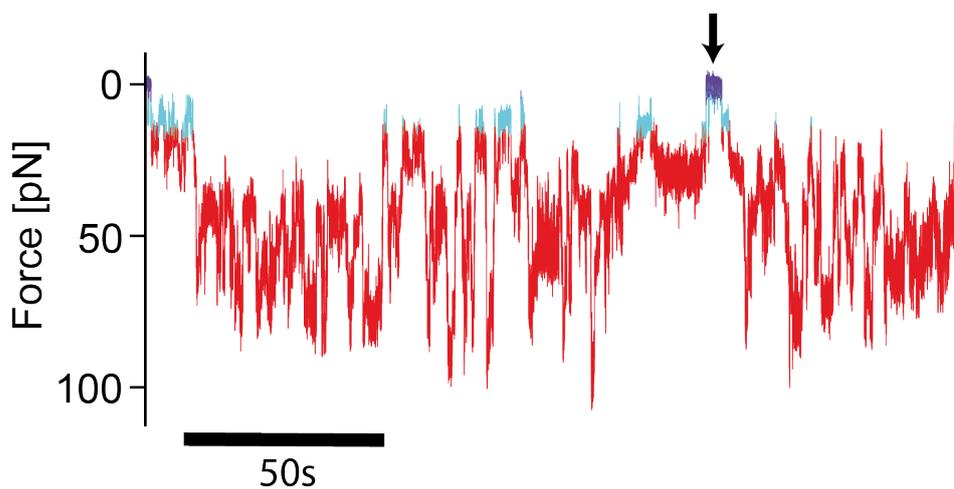


Figure 62: The nucleotide exchange factor BAG1 increases the Hsp70 dissociation rate from substrate and thereby allows the GR-LBD to refold even in the presence of high chaperone concentrations of 10  $\mu\text{M}$  Hsp70, 2  $\mu\text{M}$  Ydj1, 5 mM ATP and 5 mM  $\text{MgCl}_2$ .

Note that the only difference in experimental conditions between Fig.61 and 62 is the addition of BAG1.

## 8.7 GR-LBD Can Refold at Low Hsp70/40 Concentrations

If the chaperone concentrations are chosen low enough (roughly  $C(\text{Hsp70}) < 2 \mu\text{M}$  and  $C(\text{Ydj1}) < 1 \mu\text{M}$ ), the GR-LBD can recover from chaperone induced unfolding, i.e. can refold completely and bind ligand even without the assistance of BAG1. Fig.63 shows a trace where GR-LBD is completely unfolded by chaperones around 100s and after many refolding and unfolding steps manages to refold completely several times against the action of the Hsp70/40 chaperone system. Over the course

of hundreds of seconds and throughout multiple complete unfoldings and refoldings, the GR-LBD's unfolding and refolding is under strict control of Hsp70/40. In particular, the GR-LBD never appears to be in a misfolded state. In [44], we found misfolded conformations of GR-LBD that, so far, could not be observed in combination with chaperones. Physiologically, this might be a manifestation of Hsp70 acting as a holdase, preventing the GR-LBD from misfolding while at the same time preparing it for the subsequent interaction with HOP, Hsp90 and p23.

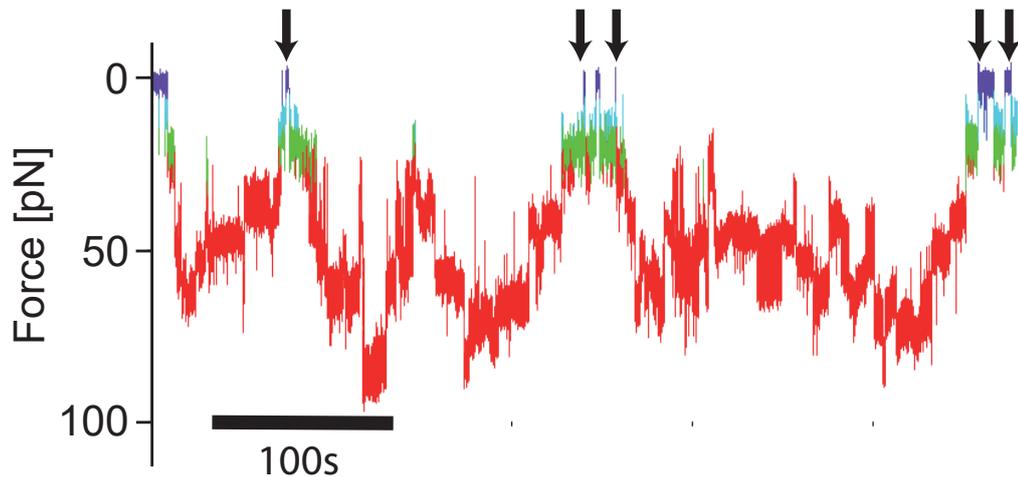


Figure 63: At sufficiently low chaperone concentrations, the GR-LBD can refold even against the action of Hsp70/40. Note that the stepwise unfolding and refolding of GR-LBD is entirely under the control of the chaperones for several hundreds of seconds. Over the entire trace, the GR-LBD appears to never enter the misfolds analysed in [44]. The trace shown was recorded at 1  $\mu$ M Hsp70, 1  $\mu$ M Ydj1, 5 mM ATP and 5 mM  $MgCl_2$ .

### 8.8 Hsp70/40 Unfolding Works in the Absence of an External Force Bias

The complete chaperone-induced unfoldings we observed in our optical trap experiments always occurred under a force bias, raising the question whether Hsp70/40 can drive such complete unfolding also in the absence of a mechanical load. To test this, the following experiment was conducted:

First, it was investigated whether chaperones can attack the DEX-bound GR-LBD. To this end, a sample of GR-LBD was incubated at 10  $\mu$ M Hsp70, 2  $\mu$ M Ydj1, 5 mM ATP and 5 mM  $MgCl_2$ . At these chaperone concentrations, a ligand-unbound GR-LBD would begin to unfold within  $\sim 1$  s in our experiments under force bias. However, in a solution containing DEX, the GR-LBD is in its DEX-bound holo state almost the entire time, and chaperones never attacked the DEX-bound GR-LBD in our experiments. Fig.8.8A shows the resulting stretch-relax cycles after 2 hours of incubation at these chaperone concentrations. Notably, the first stretching trace always exhibits a DEX-bound unfolding of GR-LBD, with the lid flipping (encircled in the first stretch-relax cycle in Fig.64) as an unmistakable sign indicative of its holo state. This was evidence that chaperones do not attack the DEX-bound state, neither under force bias, nor at zero force.

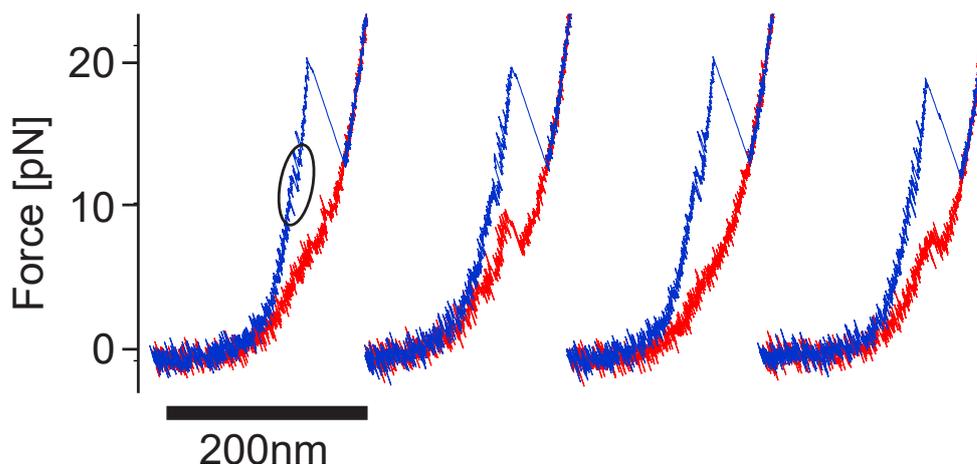


Figure 64: The first stretch-relax cycles of four different molecules after incubation with chaperones in the presence of DEX are shown. They all exhibit the flipping indicative of the DEX bound natively folded conformation. This means that Hsp70/40 did not unfold the DEX-bound GR-LBD even after 2 hours of incubation at 10  $\mu\text{M}$  Hsp70, 2  $\mu\text{M}$  Ydj1, 5 mM ATP, 5 mM  $\text{MgCl}_2$ ,  $\sim 1 \mu\text{M}$  DEX and zero force.

Clearly, the GR-LBD needed to be brought to the DEX-unbound apo state in order for chaperones to bind to it and unfold it. Because of the GR-LBD's tendency to aggregate in the absence of hormone, after purification it is always stored in the holo conformation in a buffer containing 50  $\mu\text{M}$  DEX. In our previous study [44], we had measured the affinity between GR-LBD and DEX to be on the order of  $\sim 1 \text{ nM}$ . In order to reach the apo state of GR-LBD, multiple buffer exchanges had to be conducted, reducing the DEX concentration to below 0.1 nM. This sample was then left on ice for over 16h. In [44], we had extrapolated the DEX off-rate at zero force to be  $k_{diss}(0pN) = (3 \pm 2) \cdot 10^{-5} \text{ s}^{-1}$ . Therefore, a long waiting time of 16 hours (57600 seconds) at a very low DEX concentration  $< 0.1 \text{ nM}$  was likely to suffice for a significant fraction of GR-LBD molecules to release DEX and end up in the apo state.

After the 16h incubation, two more buffer exchanges were conducted, removing all DEX in solution that might have dissociated during the incubation.

Subsequent optical trapping experiments with this sample showed that all GR-LBD molecules found were in the DEX-unbound apo state from the very beginning of the experiment. Fig.65 shows the first stretch-relax cycles of four different molecules. All molecules lack the flipping transition indicative of holo GR-LBD (cf. Fig.64).

This sample now contained mostly DEX-unbound GR-LBD molecules and was therefore ready to be incubated with chaperones in the next step. It was incubated with physiological chaperone concentrations (10  $\mu\text{M}$  Hsp70, 2  $\mu\text{M}$  Ydj1, 5 mM ATP and 5 mM  $\text{MgCl}_2$ ) for 40 min. In the following optical trapping experiments, all GR-LBDs ( $n = 10$ ) were completely unfolded already before the first stretching cycle. The first stretch-relax cycles of four different molecules are shown in SI Fig.66. They lack any sign of a folded structure, but merely follow the extension characteristic of an unfolded GR-LBD's polypeptide chain instead.

It should be noted that, in the absence of chaperones, a GR-LBD that was unfolded already before the first stretching cycle, such as the ones in Fig.66, has never been

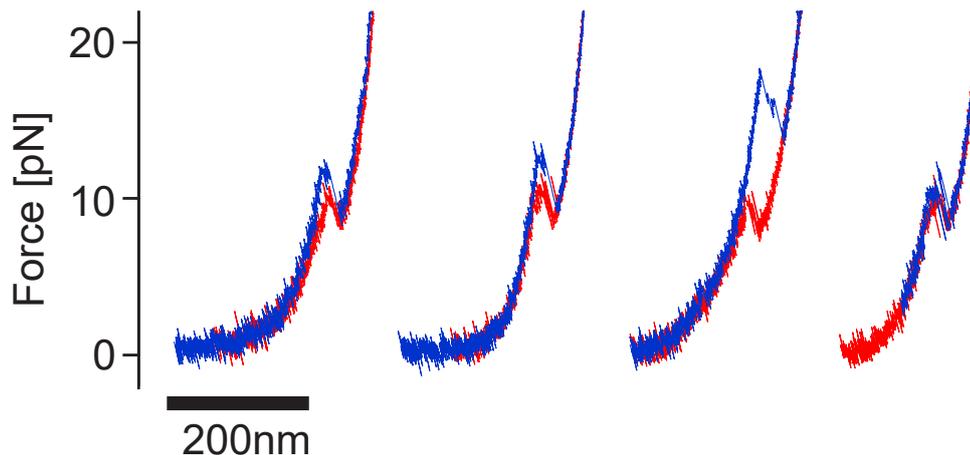


Figure 65: Through multiple buffer exchanges the DEX concentration was reduced to  $<0.1$  nM. After subsequent incubation at this low DEX concentration for 16h, all GR-LBD molecules were DEX-unbound in optical trapping experiments.

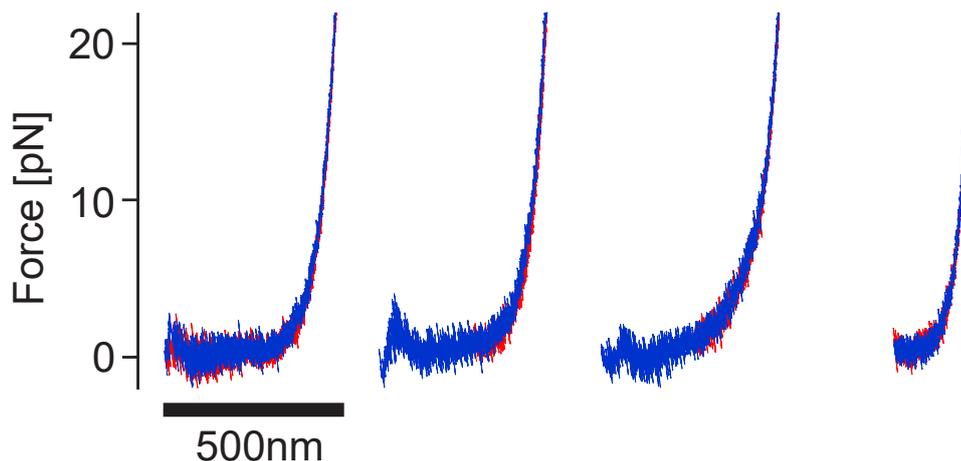


Figure 66: With the GR-LBD in the DEX-unbound apo state initially, Hsp70/40 completely unfolds it at zero force. The first stretch-relax cycles of four different molecules, which were in the apo state before incubation, after 40 min incubation at  $10\mu\text{M}$  Hsp70,  $2\mu\text{M}$  Ydj1,  $5\text{mM}$  ATP and  $5\text{mM}$   $\text{MgCl}_2$  and zero force are shown. All traces correspond to the worm-like chain of a completely unfolded GR-LBD's peptide chain.

observed. This fact increases the significance of the zero-force findings presented above.

### 8.9 Hsp70 Binding Sites in GR-LBD Sequence - Experiment vs. Algorithms

Whenever the Hsp70/40-induced unfolding traces of GR-LBD showed clearly distinguishable steps, the number of unfolding intermediates was reproducibly 5. Fig.67 shows several examples of such traces.

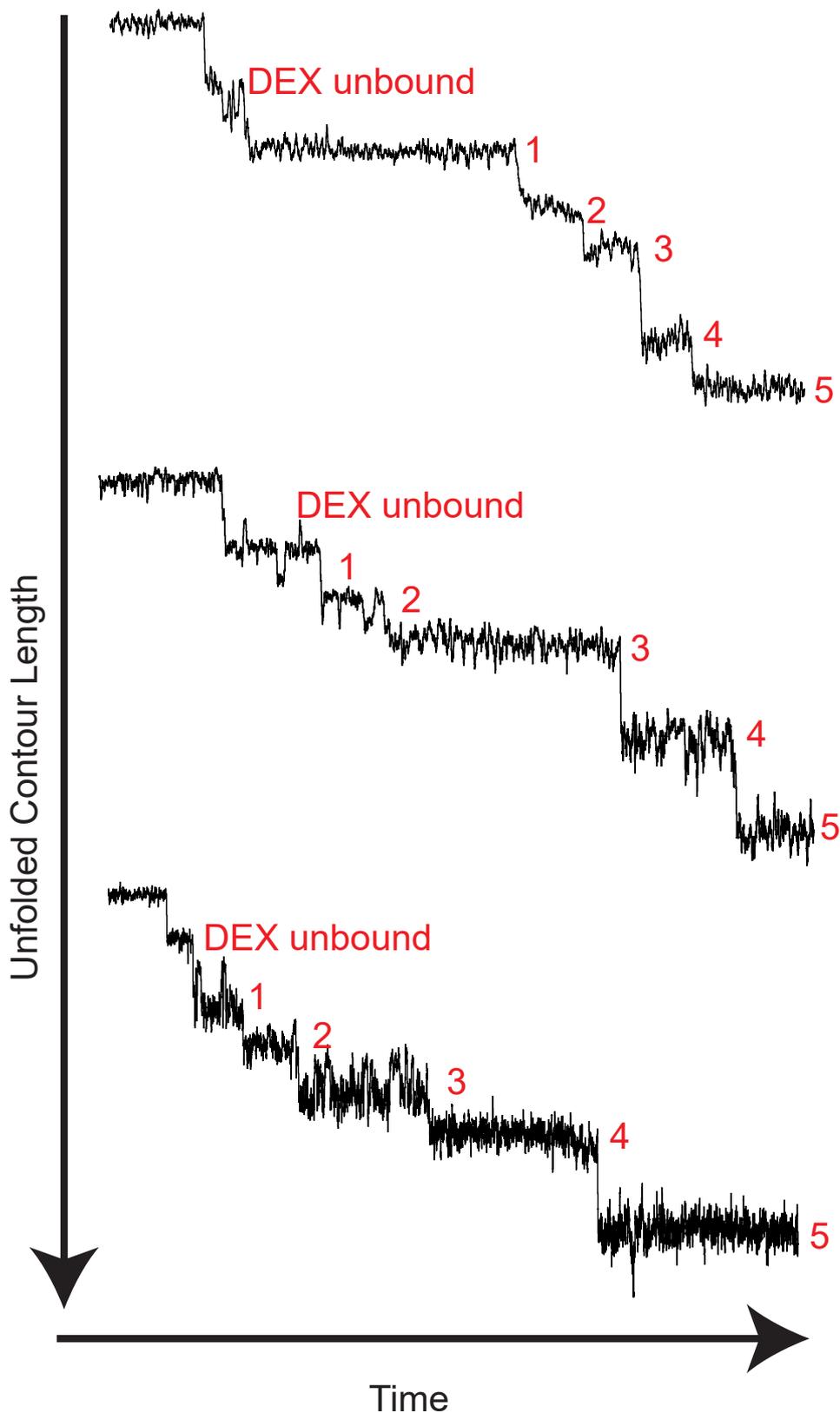


Figure 67: Hsp70/40 unfoldings of GR-LBD, which exhibit clearly visible unfolding steps, always proceed via 5 unfolding intermediates, suggesting the binding of 5 Hsp70s. Note that the first step is always the DEX-unbound state and does not count to the 5 Hsp70 unfolding intermediates that follow.

While the contour length at which these intermediates occur can vary, there is always one reproducible first unfolding intermediate at 32 nm contour length, as Fig.68 shows.

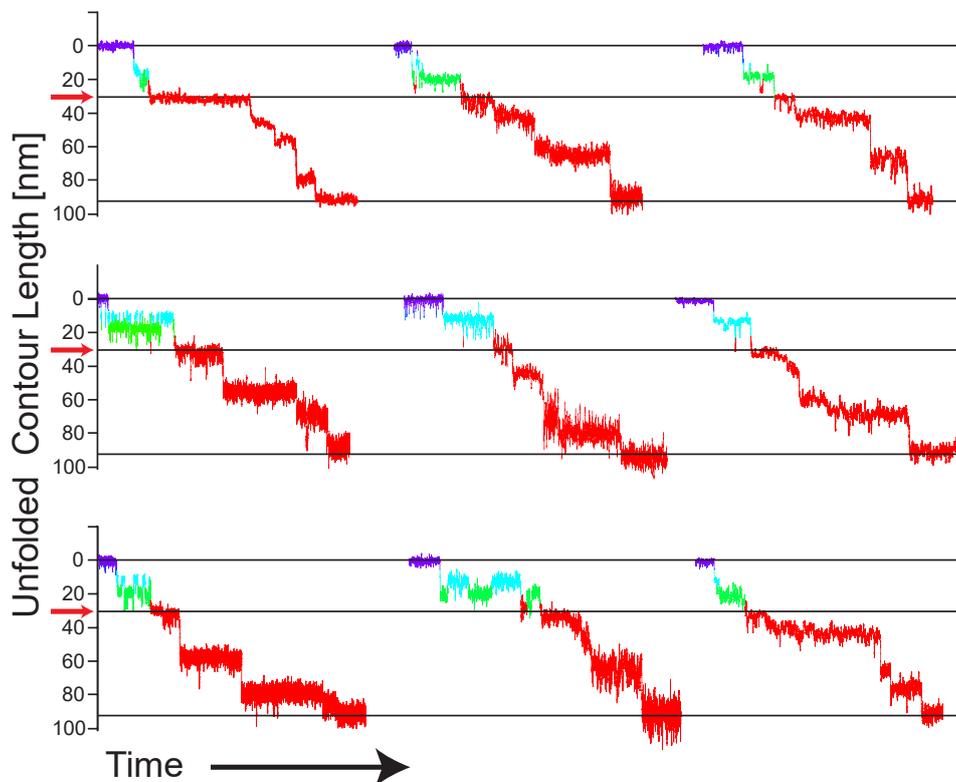


Figure 68: All shown unfoldings exhibit a clear first unfolding intermediate at 32 nm.

There exist several algorithms designed to predict Hsp70 binding sites in given amino acid sequences. Most of them predict binding sites for DnaK, which is the *E.coli* homologue of Hsp70. Human and *E.coli* have similar binding preferences (cf. section 1.6).

In order to test the plausibility of the results shown above, two of these Hsp70 binding site prediction algorithms were applied to the sequence of F602S GR-LBD, namely the LIMBO algorithm [136] and the chaperISM algorithm [138].

### 8.9.1 The LIMBO Algorithm

The LIMBO algorithm by Switchlab [136] is freely available online [139].

When set to the default "best overall" prediction option, applying the LIMBO algorithm to the sequence of GR-LBD (cf. section 1.2, cf. section 12.9.1) yielded the binding site predictions shown in Fig.69.

The LIMBO algorithm detects 6 highly likely Hsp70 binding sites in the sequence of GR-LBD. The first of these binding sites lies within the N-terminal "lid", which in this study was found to govern hormone binding (c.f. section 6). However, Hsp70

LIMBO chaperone binding prediction  
 VIB Switch Laboratory  
 Katholieke Universiteit LeuvenEntry name

Position	Sequence	Score
35-41	WRIMTTL	18.3
83-89	AFALGWR	19.0
144-150	MKTLLLL	11.1
146-152	TLLLLSS	14.1
166-172	EIRMTYI	13.0
248-254	KKLLFHQ	16.4

Figure 69: The high-likelihood Hsp70 binding sites in the GR-LBD sequence as predicted by the LIMBO algorithm, with position in sequence, amino acid sequence, and likelihood score [136] [139].

binding to this lid in our experiments can be safely excluded by looking at the flipping rates in the presence and absence of chaperones, which are identical (cf. section 8.2). In our experiments, chaperones never were able to attack the GR-LBD as long as it was ligand-bound. Chaperone interaction started immediately after DEX dissociation.

Excluding the lid, in the remainder of the GR-LBD sequence, the LIMBO algorithm predicts 5 high-likelihood Hsp70 binding sites, which is in excellent agreement with the results of this work. Importantly, the Hsp70 binding site with the highest binding probability is found at amino acid position 83-89. The end of this binding motif, i.e. the 89th amino acid in the sequence of GR-LBD, is situated at precisely 32.4 nm contour length from the N-terminus [121]. In the presence of chaperones, we consistently detect a long-lived unfolding intermediate at these 32 nm (cf. Fig.68 and section 8.10). This intermediate never showed up in the absence of chaperones. The fact that the highest-likelihood Hsp70 binding site predicted by the algorithm and the novel intermediate in our experiment coincide at the very same contour length is another litmus test confirming our results and the validity of our chaperone assay overall.

The 5 predicted binding sites are evenly spread over the entire sequence of GR-LBD, which is a prerequisite, if multiple sequential Hsp70 bindings are supposed to unfold the protein in a concerted and chronological action. For illustration, a mapping of these binding sites on the sequence of GR-LBD with the predicted highly probable Hsp70 binding sites highlighted in red, is shown:

N-SACK-GR-LBD F602S/C638D-KCL-C (GR-LBD):  
 SACKQLTPTLVSLLEVIEPEVLYAGYDSSVPDSTWRIMTTLNMLGGRQV  
 IAAVKWAKAIPGFRNLHLDDQMTLLQYSWMSLMAFALGWRSYRQSSA  
 NLLCFAPDLIINEQRMTLPDMYDQCKHMLYVSSELHRLQVSYEEYLCM  
 KTTTTLSSVPKDGLKSQELFDEIRMTYIKELGKAIVKREGNSSQNWQRF  
 YQLTKLLDSMHEVVENLLNYCFQTFLDKTMSIEFPEMLAEIITNQIPKY  
 SNGNIKLLLFHQKCL

### 8.9.2 The ChaperISM Algorithm

The ChaperISM algorithm [138] is also freely available, but requires the download of its code from Github [140] and correct implementation using a terminal.

When applied to the sequence of GR-LBD, it resulted in a multitude of possible Hsp70 binding sites, that usually clustered around the same positions as in the prediction of the LIMBO algorithm (the ChaperISM Hsp70 binding site predictions are shown in SI section 12.10).

## 8.10 Intermediates in Chaperone-Induced and Force-Induced Unfoldings

To investigate the mechanism used by Hsp70/40 to unfold GR-LBD, we compared purely force-induced unfolding traces with chaperone-induced unfolding traces of GR-LBD. Figs. 70A and 70B illustrate the striking differences in passive-mode experiments.

Under the influence of a high force of 11.2 pN and in the absence of chaperones (Fig.70A), the GR-LBD undergoes rapid equilibrium transitions between the folded holo state and the completely unfolded state, populating multiple short-lived intermediates on the way. These intermediates and the corresponding force-dependent transition rates have been analysed in detail in our previous studies [44]. The transition rates between these unfolding intermediates were on the order of  $1 \times 10^3 \text{ s}^{-1}$ . Note that refolding from the completely unfolded peptide chain to the natively folded apo state (light-blue) can occur within as little as 0.01s, even against the given high force bias.

In contrast, in the presence of chaperones (Fig.70B), the individual unfolding steps are slow (on the order of seconds) and no additional fast transitions between unfolding intermediates take place. Also, the unfolding intermediates occur at entirely different contour lengths than the unfolding intermediates observed in force-induced unfoldings. This indicates the unfolding as well as the folding steps are under complete control of the chaperones.

Stretch-relax experiments performed with GR-LBD at  $500 \text{ nm s}^{-1}$  in the absence of chaperones and at very high chaperone concentrations yield additional evidence that the chaperones actively control the unfolding process (Fig.71).

A zoom into a set of 5 pulls in the absence of chaperones (Fig.71A) shows that the complete unfolding process from the folded to the unfolded state occurs without any long-lived population of intermediates. This is consistent with the very short intermediate dwell times we find under those conditions in passive mode experiments (Fig.70A). Beyond an occasionally visible very short-lived intermediate at a contour length of 37 nm, no further intermediates can be observed.

By contrast, in ca. 8% of pulling traces (5 out of 61) obtained at very high chaperone concentrations (30  $\mu\text{M}$  Hsp70, 2  $\mu\text{M}$  Ydj1, 5 mM ATP and 5 mM  $\text{MgCl}_2$ ) we observe strikingly long-lived unfolding intermediate states at a contour length of  $\sim 32$  nm, matching our experimental results in passive-mode. Again, this chaperone-induced intermediate differs from the rare intermediate in the absence of chaperones (Fig.71A) both in contour length and dwell time. When one of the components (Hsp70, Hsp40, MgATP) was left out, such intermediates were never observed (0 out of 31). This finding provides evidence, that Hsp70 attacks the folded apo GR-

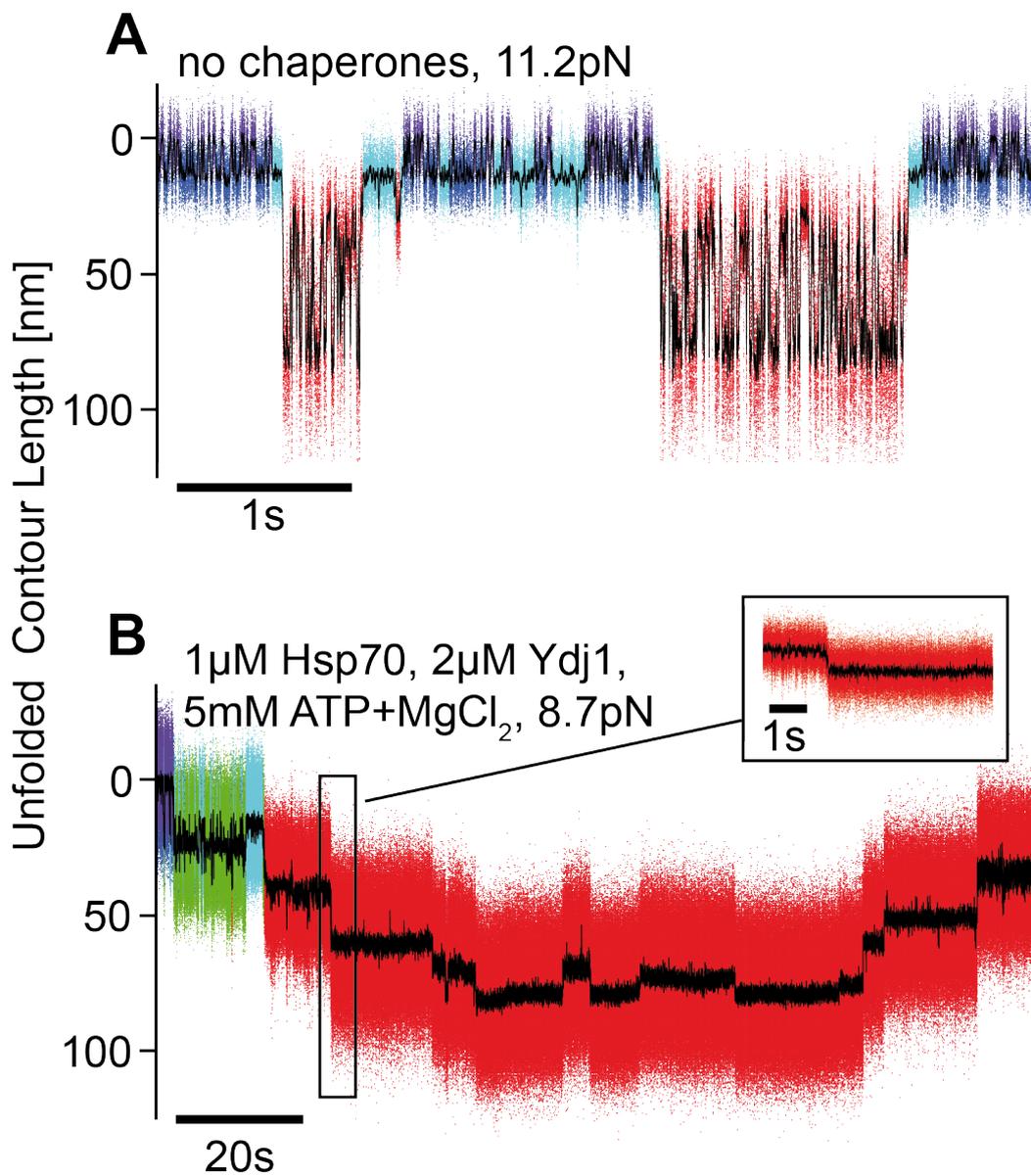


Figure 70: **A)** Under a high force bias of 11.2pN, GR-LBD completely unfolds and refolds in equilibrium, populating at least 4 unfolding intermediates with high kinetics [44]. The unfolding and refolding transition rates between these intermediates lie on the order of  $1 \times 10^3 \text{ s}^{-1}$ . **B)** The unfolding and refolding intermediates in chaperone-induced unfoldings exhibit entirely different dwell-times and contour lengths as compared to the intermediates in force-induced unfoldings (A))

LBD structure and unfolds it to an intermediate state that is stabilized by direct interaction with Hsp70. The chaperone binding site with the highest binding probability (as predicted by the LIMBO algorithm [136]) is situated precisely at  $\sim 32$  nm contour length.

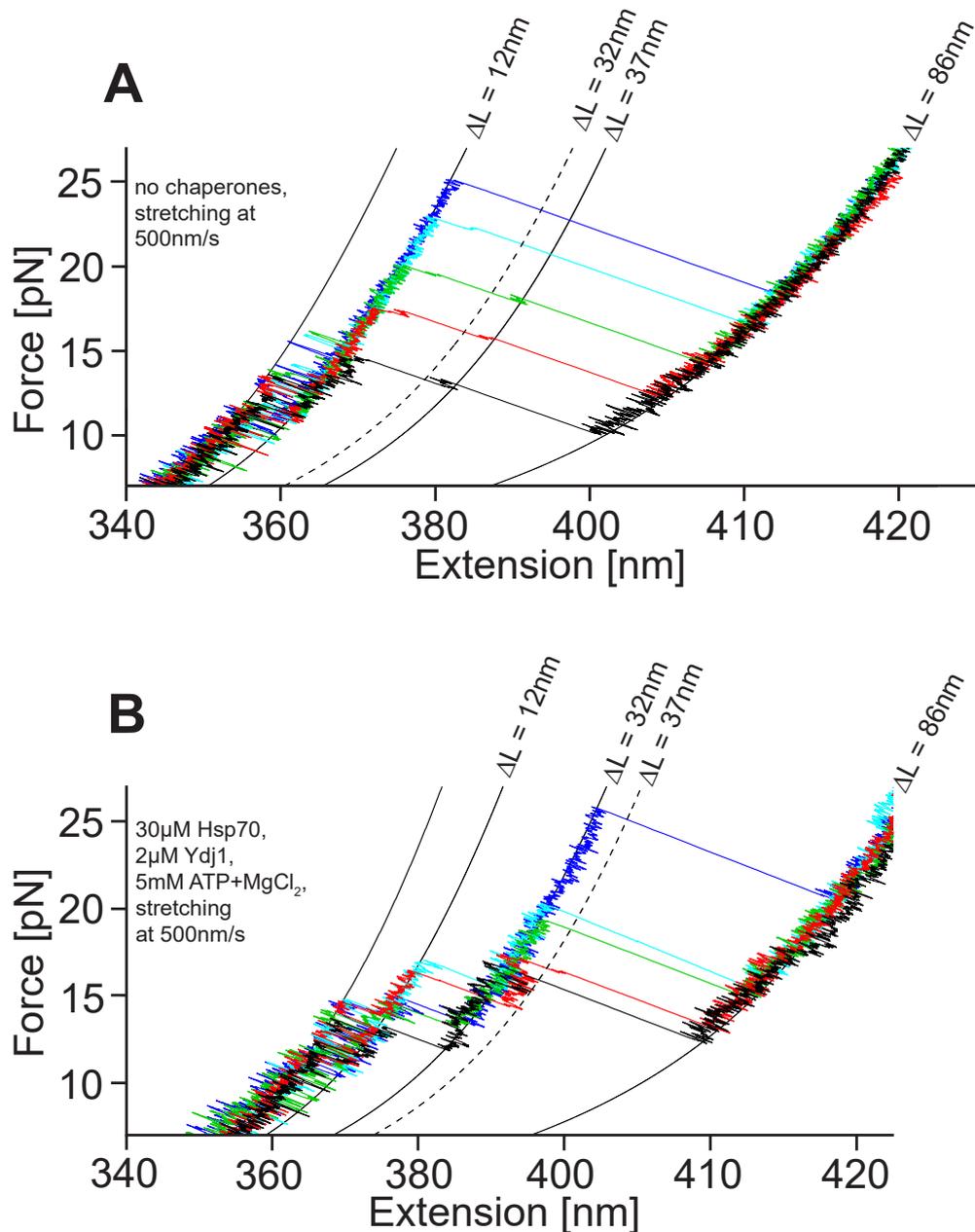


Figure 71: **A**) In stretch-relax cycles of GR-LBD at  $500 \text{ nm s}^{-1}$  in the absence of chaperones, no long-lived intermediates are populated. A very short-lived and rare unfolding intermediate at 37 nm contour length can occasionally be observed, when the core of the protein happens to unfold at a relatively low force. **B**) At extremely high chaperone concentrations of 30  $\mu\text{M}$  Hsp70, 2  $\mu\text{M}$  Ydj1, 5 mM ATP and 5 mM  $\text{MgCl}_2$ , in ca. 8% of pulling traces (5 out of 61) a strikingly long-lived unfolding intermediate at  $\sim 32 \text{ nm}$  can be detected. This intermediate clearly differs from the rare unfolding intermediate in the absence of chaperones (A) both in dwell time and contour length.

# Discussion

## 9 The Hsp70/40 Unfolding Mechanism

Taking into account the entire evidence presented in this work so far, it is possible to derive conclusions concerning the mechanism by which Hsp70/40 performs its chaperoning tasks on the GR-LBD.

### 9.1 Hsp70 as a Holdase

A large body of literature provides evidence that Hsp70 can act as a holdase, i.e. can strongly bind to unfolded peptide chains, thereby inhibiting the aggregation as well as the refolding of the substrate protein [37] [80] [141]. In our experiments, Hsp70 shows its ability to act as a holdase particularly in the stretch-relax cycles in the presence of chaperones (cf. Fig.60). After its mechanical unfolding, the protein chain remains permanently unfolded, indicating that the chain becomes decorated with Hsp70 molecules. The same effect was observed in passive-mode experiments (cf. Fig.61): After initial unfolding, in this case not caused by a force bias, but by the action of Hsp70/40, the GR-LBD is held permanently unfolded for several hundreds of seconds. The fact that the nucleotide exchange factor BAG1 abolishes the chaperone-induced inhibition of GR-LBD refolding in passive-mode experiments further corroborates the holdase hypothesis (cf. Fig.62). Only if the peptide chain of the completely unfolded GR-LBD is previously decorated with multiple Hsp70s can BAG1 have the observed effect.

The fact that Hsp70 must be acting as a holdase also shows itself in the passive-mode traces at low chaperone concentrations (cf. Fig.63). Here, the GR-LBD is unfolded by chaperones in the usual stepwise manner, but also shows refolding in an equally stepwise fashion. Both the unfolding and refolding are under strict control of the chaperones bound to the peptide chain. Whenever an Hsp70 dissociates from the peptide chain, the particular segment of GR-LBD that it occupied before might be able to refold, but only as far as the remaining Hsp70s bound to the chain allow it to fold back. If it were not for the holdase activity of Hsp70, in this trace the GR-LBD would refold rapidly within less than 0.1 s and via its native folding intermediates (cf. Fig.70A).

### 9.2 Hsp70 as an Unfoldase

Hsp70 has also been implicated in the unfolding of natively and non-natively structured proteins, thus exhibiting unfoldase activity [38] [43] [80] [141] [142] [143] [144]. Various models have been suggested for such an unfoldase mechanism, including entropic pulling to induce the breakdown of clathrin cages [144] and amyloid aggregates [38], a ratchet mechanism in the context of the translocation of proteins through membrane pores [145] or a similar mechanism inducing the inactivation of  $\sigma^{32}$  [146].

Most likely, the mechanisms are plentiful and vary from substrate to substrate. At the end of this discussion, we will be in a position to contribute to the debate about the potential unfolding mechanisms of Hsp70/40.

### 9.3 The Hsp70/40-Induced Unfolding of GR-LBD

The results presented in this work provide direct evidence that Hsp70/40 is able to unfold the GR-LBD completely in a sequence of multiple Hsp70/40 binding events and Hsp40 stimulated hydrolysis. A concerted action of many complexes (up to 5 in our experiments) leading to a stepwise unfolding is novel for this class of chaperones and may serve as a more general mechanism for how the energy of multiple ATP hydrolyses can be used to unfold stable proteins.

Obviously, both the applied force bias by the optical traps as well as Hsp70/40 drive the GR-LBD towards the completely unfolded conformation. It is important to note, however, that this Hsp70/40 induced complete unfolding also happens in the absence of force (as shown in section 8.8, cf. Fig.66) and is therefore not just a consequence of the force exerted by the optical traps.

Whether or not the complete GR-LBD is unfolded will depend on the concentrations of chaperones (cf. Figs.61 and 63) as well as co-factors, like BAG1, which are able to shift the equilibrium between folded and unfolded states by stimulating ADP release (cf. Fig.62). Kirschke et al. proposed that Hsp70/40 unfolds GR-LBD only partially, but their HDX-MS experiments were conducted at a molar ratio of GR-LBD and Hsp70 close to 1:1, thus precluding multiple Hsp70 bindings to one GR-LBD molecule [43] [68]. While the physiological GR-LBD/Hsp70/40 ratio may vary in the cell, our scenario is relevant given that cellular Hsp70 concentrations generally exceed GR concentration by far [133] [147] [148].

### 9.4 When Exactly Do Hsp70 Binding and Hydrolysis Occur?

Section 8.3 showed that GR-LBD unfolding by Hsp70/40 involves multiple, often clearly distinguishable, steps (cf. Figs. 54 and 73). The fact that the unfolding rate between intermediates scales with both Hsp70 concentration (cf. Fig.57) and Hsp40 concentration (cf. Fig.59) is evidence that each unfolding step must consist of two events:

1. Hsp70/40 binding
2. subsequent ATP hydrolysis

Accordingly, the dwell times on each unfolding level are the sum of binding times for Hsp70 and Hsp40 as well as the time the very hydrolysis step takes. Our simple kinetic model (Eq.4) formulated this hypothesis mathematically. It fully explains the dependence of the unfolding intermediate dwell times on chaperone concentrations (cf. Figs.56 and 58). It also accounts for the saturation behavior of the graphs in Figs.57 and 59: at very high concentrations of either one chaperone, the binding rate of the other chaperone as well as the highest possible hydrolysis rate  $k_{max,hydrolysis}$  become limiting. The highest unfolding rate per step achieved in our experiments was  $k_{step,unfold} = 1.34s^{-1} \pm 0.22s^{-1}$ , in agreement with previously measured hydrolysis rates [149] [150]. In our model, a potential delay between hydrolysis and unfolding is lumped into  $t_{min,hydrolysis} = \frac{1}{k_{max,hydrolysis}}$ .

For each individual unfolding step, we propose the following chronological sequence of the two events "Hsp70 binding" and "ATP hydrolysis", illustrated in Fig.72: After a new unfolding intermediate has been populated, Hsp70 binds within a cer-

tain amount of time depending on the Hsp70 concentration. For example, at  $10\ \mu\text{M}$  Hsp70, on average, binding occurs within  $\tau_{\text{bind}} \approx 0.5\text{s}$  (cf. SI Table 1). ATP hydrolysis is stimulated by Hsp40 and ends the dwell time in the unfolding intermediate by inducing the transition to the next intermediate at a longer unfolded contour length (e.g. at  $10\ \mu\text{M}$  Hsp70 and  $2\ \mu\text{M}$  Ydj1, the average dwell time on the intermediate is  $\tau_{\text{hydrolysis}} \approx 1.25\text{s}$ , cf. SI Table 1). Reducing the concentration of either one of the chaperones will delay either Hsp70 binding or ATP hydrolysis and thereby elongate the dwell time spent on the unfolding intermediate.

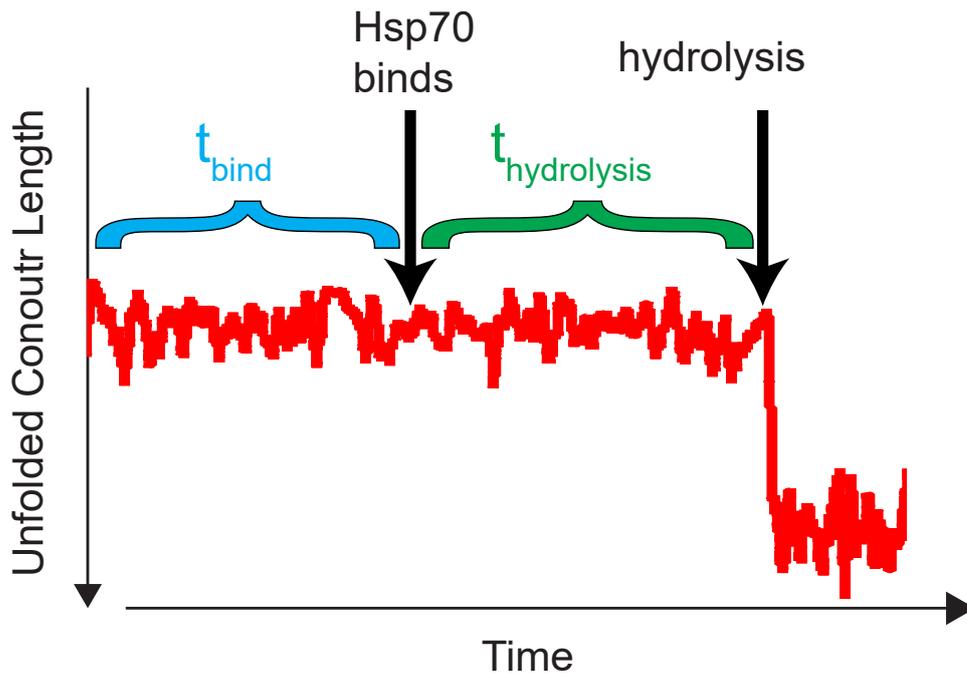


Figure 72: In our model, Hsp70 binds during the dwell time on an unfolding intermediate. Hsp70 binding is invisible in our experiments. The time  $t_{\text{bind}}$  that passes until Hsp70 binding depends on Hsp70 concentration. The time  $t_{\text{hydrolysis}}$  required for ATP hydrolysis depends on the Hsp40 concentration. Upon ATP hydrolysis, the GR-LBD undergoes a transition to the next unfolding intermediate.

Hsp70 binds during the dwell time on an unfolding intermediate. The time that passes until this binding occurs ( $t_{\text{bind}}$ ) depends on the Hsp70 concentration. Afterwards, GR-LBD remains on the unfolding intermediate for as long as ATP hydrolysis requires at the given chaperone concentrations ( $t_{\text{hydrolysis}}$ ). The moment of ATP hydrolysis marks the transition to the next unfolding intermediate, with a potential small delay that it takes for the unfolding mechanism to work. This model can explain the dependence of the unfolding intermediate dwell times on both Hsp70 and Hsp40 concentrations in both fast and slow unfoldings.

The last paragraph explained the chronology of events during an unfolding transition from one chaperone unfolding intermediate to the next. The opposite case of a refolding transition (e.g. the clearly visible upward steps in Fig.63) then corresponds to the dissociation of Hsp70 followed by the refolding of a part of the GR-LBD's peptide chain that had previously been occupied by Hsp70. This process can be accelerated through the addition of the nucleotide exchange factor BAG1 (cf. Fig.62).

To orchestrate the observed stepwise and complete substrate unfolding, Hsp70s must bind to motifs that gradually become exposed as unfolding progresses. The LIMBO algorithm for DnaK binding motif prediction [136] detects five evenly distributed DnaK binding sites in the 224aa sequence of the folded part of apo GR-LBD (cf. section 8.9 and Fig.73C). Sterically, the length of the unfolded GR-LBD polypeptide allows the binding of 5 Hsp70 molecules. In the case of the similar-sized protein rhodanese, binding of up to 7 DnaKs could be modelled onto its unfolded chain [143].

So far, the discussion of the unfolding only concerned the chronology of the "local" events "Hsp70 binding" and "ATP hydrolysis". The next sections will discuss how these events can "globally" lead to the complete unfolding of GR-LBD.

### 9.5 The Overall Model of Hsp70/40-Induced Unfolding of GR-LBD

The previous section discussed what must be happening during each individual unfolding step. For the global unfolding of the entire GR-LBD, we propose the model illustrated in Fig.73. In Fig.73A, the molecular processes are schematically depicted. Fig.73B links the chronology of molecular processes to the events in the observed stepwise unfolding traces (73C) as follows:

#### Figure 73:

##### A) and B)

**1. and 2.:** The GR-LBD starts in the DEX-bound state, exhibiting the familiar lid flipping. In our experiments, Hsp70 never attacks the DEX-bound state, but exclusively the DEX-unbound state. It follows that an Hsp70 binding site must be exposed when DEX dissociates.

**3.:** DEX dissociates from the lid-open state, leaving an Hsp70 binding site exposed.

**4.:** Now an Hsp70, in its ATP bound state with open SBD, attacks the DEX-unbound GR-LBD at the first exposed binding site.

**5.:** Upon ATP hydrolysis, which is stimulated by Hsp40, Hsp70 closes its substrate binding domain, forcing a part of GR-LBD to unfold. How exactly this local Hsp70/40 unfolding might work will be discussed in more detail in the next section.

**6.-8.:** Steps 3. and 4. repeat four more times until GR-LBD is completely unfolded and decorated with at least 5 bound Hsp70s.

##### C)

Illustration of the positions of Hsp70 binding sites in the sequence of GR-LBD as predicted by the LIMBO algorithm [136].

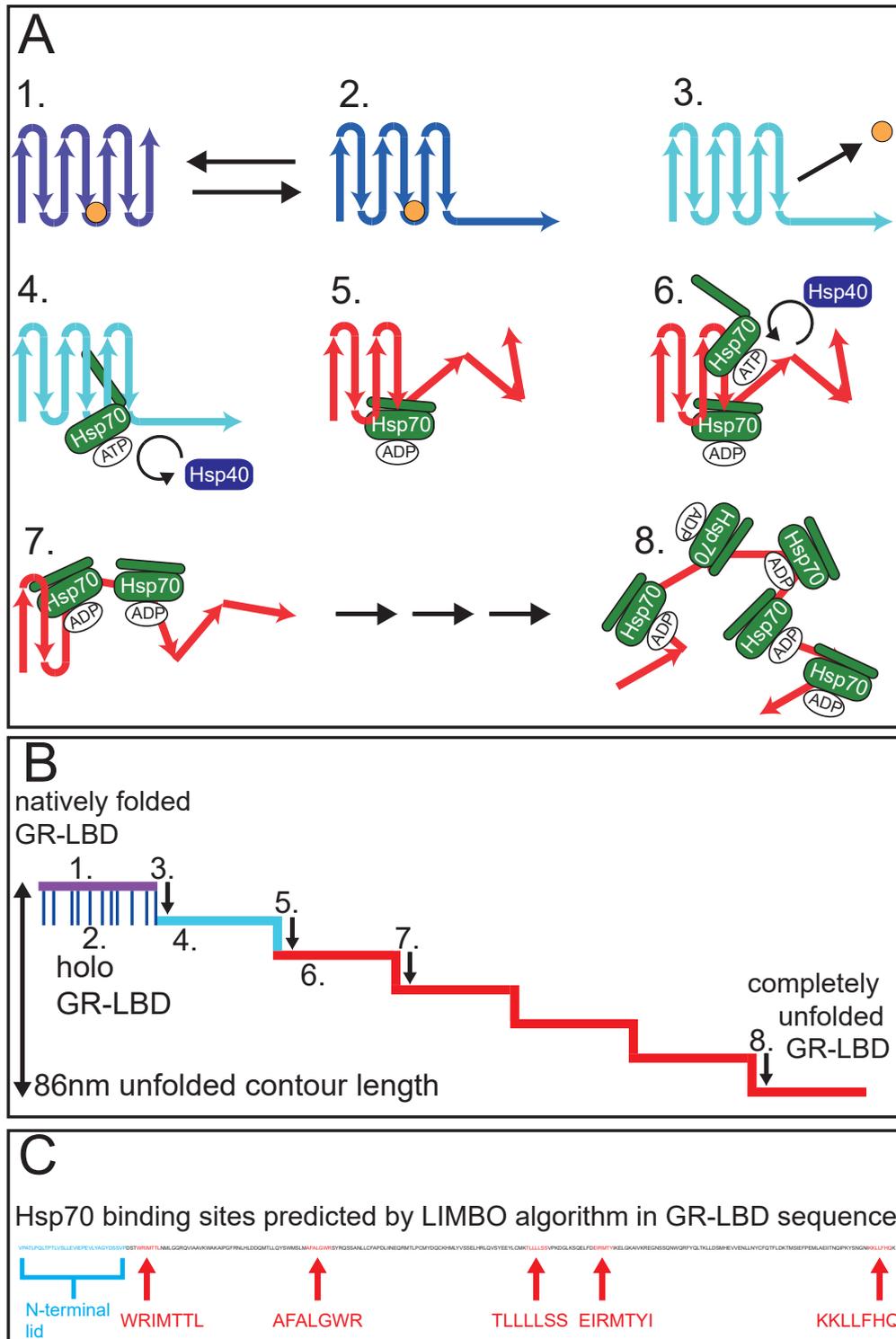


Figure 73: **Model of Hsp70/40 induced unfolding of GR-LBD** The same color code as in all previous passive-mode traces was used to identify the states of GR-LBD: Purple: natively folded, DEX-bound, lid closed; dark-blue: DEX-bound, lid open; light-blue: DEX-unbound, apo structure still folded; red: unfolded by chaperones. The figure caption is given on the previous page.

## 9.6 Potential Unfolding Mechanisms of Hsp70/40

In general, unfoldase mechanisms can be divided into the two categories of passive “ratchet mechanisms” and “active mechanisms” [78], similar to the paradigm juxtaposition of “conformational selection” vs. “induced fit” [151].

In the context of Hsp70/40 unfolding of the GR-LBD, a pure ratchet mechanism would require frequent spontaneous conformational fluctuations of the GR-LBD. Hsp70/40 could then bind to more unfolded conformations (“conformational selection”) and thereby inhibit their refolding. This would rectify the fluctuations into the direction of further unfolded contour lengths and eventually lead to an unfolding of the GR-LBD.

An active unfolding mechanism on the other hand would have to involve the conversion of the hydrolysis-driven closing of Hsp70’s SBD around the GR-LBD’s peptide chain into a force that unfolds the GR-LBD. Afterwards, refolding of the actively unfolded parts of GR-LBD would have to be inhibited.

In the following, we will discuss a “ratchet” as well as two “active” unfolding mechanisms in the light of our experimental results.

### 9.6.1 Ratchet

As stated above, a ratchet mechanism would have to rely on the presence of spontaneous conformational fluctuations of the GR-LBD towards further unfolded states, which Hsp70 could then rectify. In our experiments, we did not observe such fluctuations. Looking at the Hsp70/40 unfolding in Fig.70B for example, the unfolding and refolding steps in the presence of chaperones exhibit long dwell times on the order of  $\sim 5$  s without significant fluctuations present neither short nor long-lived. A pure “ratchet” mechanism is hence unlikely.

One might object that the fluctuations could be too fast for our sampling rate (30kHz) to be recorded, or that the smoothing of our data conceals them. However, even if this were the case, the fluctuations would still only sum up to extremely small amounts of cumulative dwell time in more unfolded conformations, rendering it impossible for Hsp70/40 to bind and conduct ATP hydrolysis within these short time intervals. Analysis of chaperone-induced unfolding traces using a Hidden-Markov-Model rendered an upper limit for the population of such potential (and hence in our assay invisible) transiently unfolded states of 0.5 % at 10  $\mu$ M Hsp70 and 2  $\mu$ M Ydj1. Our experiments involving the variation of chaperone concentration (Figs. 57 and 59) rendered an on-rate of ATP-bound Hsp70 of  $k_{on,Hsp70} = 0.20 \pm 0.03 s^{-1} \mu M^{-1}$ , in agreement with previous studies, which reported a comparable on-rate of ATP-bound Hsp70 to substrate of  $k_{on,Hsp70} = 0.45 \pm 0.03 s^{-1} \mu M^{-1}$  [152] [153] [154]. By contrast, binding of Hsp70 to conformational fluctuations that are only populated  $< 0.5\%$  of the time would require a much higher actual on-rate of ATP-bound Hsp70 to substrate of  $k_{on,Hsp70} > 40 s^{-1} \mu M^{-1}$ .

In addition, we know from the previous discussion that both Hsp70 binding and ATP hydrolysis are obligatory for each unfolding step. Several attempts of binding may be necessary until Hsp70 commits to hydrolysis, rendering a ratchet mechanism even less likely.

All of these arguments lead to the conclusion that Hsp70/40 most probably unfolds GR-LBD through an active unfolding mechanism, not a ratchet.

### 9.6.2 Entropic Pulling

An entropic pulling mechanism [155] would require Hsp70 to bind to an unfolded part of GR-LBD, but extremely close to the still folded remainder of the structure. It would then be entropically more favourable for Hsp70 to be located further away from the remaining folded part of GR-LBD, which would lead to repulsive forces between the Hsp70 bound to the unfolded peptide chain and the folded portion of the intermediate. The consecutive action of 5 Hsp70s binding to evenly distributed binding sites (Fig.73C) might suffice to unfold the 224aa folded remainder of apo GR-LBD. This would, however, require that Hsp70 binding sites and the domain boundaries of each unfolding intermediate match precisely within less than 1 nm. Moreover, given the relative sizes of GR-LBD (30kDa) and Hsp70 (70kDa), entropic pulling seems even more unlikely. With the GR-LBD being less than half the size of Hsp70, neither of the molecules provides a large enough excluded volume, comparable to a membrane or a large aggregate, for an entropic pulling mechanism to work. While we cannot rule out an entropic pulling mechanism entirely, it unlikely constitutes the major mechanism for Hsp70/40 induced GR-LBD unfolding.

### 9.6.3 Direct Interaction between Hsp70 and Folded GR-LBD

As mentioned earlier, the contour lengths and number of the unfolding intermediates induced by Hsp70 can vary substantially in our experiments, but the first intermediate occurring at 32 nm unfolded contour length is very reproducible in passive-mode traces that exhibit clear unfolding steps (cf. Figs. 54, 56, 58, 68).

The same long-lived intermediate at 32 nm also shows up in force-extension traces at very high chaperone concentrations (cf. Fig.71B).

Importantly, this 32 nm intermediate never shows up in the absence of chaperones. In passive-mode unfoldings in the absence of chaperones, we see at least four unfolding intermediates, but none at 32 nm contour length (cf. Fig.70). In stretch-relax cycles in the absence of chaperones we instead only observe a short-lived intermediate at 37 nm (cf. Fig.71A).

Neither a ratchet nor an entropic pulling mechanism can explain the population of the new chaperone-induced intermediate, since for both mechanisms we would expect to find the same intermediates as in force-induced unfoldings in the absence of chaperones (i.e. the short-lived 37 nm intermediate or any of the short-lived unfolding intermediates in Fig.70A). Among the Hsp70 binding sites predicted by the LIMBO algorithm [136] (cf. section 8.9 and Fig.73 C), the binding site with the highest score lies exactly at 32 nm contour length from the N-terminus.

Apparently, once DEX has dissociated from GR-LBD, Hsp70 can bind to this binding site within the folded core of GR-LBD and, upon ATP hydrolysis, unfold the upstream part of the polypeptide towards the N-terminus. Note again, that the 32 nm intermediate only forms in Hsp70-induced unfoldings. Hsp70 likely still forms a direct contact with the folded GR-LBD even after inducing the first unfolding step. Such a direct interaction could explain the longevity of the 32 nm intermediate observed in stretch-relax traces, where this intermediate exists up to forces  $>25$  pN (Fig.71B).

Direct binding of Hsp70 to folded structures and associated stabilization of proteins has been reported before [156] [157]. While in those studies active unfolding by Hsp70/40 could not be observed, such chaperone-bound intermediates might serve as starting points for the active unfolding we observe. We propose a “chewing”

mechanism as a new mode for Hsp70 to act as an unfoldase, where Hsp70 binds to a folded structure and upon Hsp40 mediated ATP hydrolysis forces open a part of the protein attacked. While the results presented in this study give a strong indication for such a mechanism for the first 32 nm intermediate, we believe that, for later unfolding steps, a combination of all three discussed mechanisms may occur. The sequence of events we postulate for Hsp70-induced GR-LBD unfolding is summarized in Fig.73A and 73B: upon lid-opening and subsequent DEX unbinding, the folded portion of GR-LBD exposes the 32 nm binding site. Hsp70 can attack and, after Hsp40 mediated ATP hydrolysis, unfolds the N-terminal part of GR-LBD up to this binding site. This sequence is repeated until unfolding is completed. As outlined above, the later unfolding steps could be a combination of chewing, ratchet, or entropic pulling.

## 10 The Role of Hsp40

The exact role of the Hsp40 co-chaperone in Hsp70/40 interaction with substrate has been an important matter of debate for a while [29] [84]. It is a well-known fact that Hsp40 stimulates Hsp70's ATP hydrolysis via its J-Domain [88] [158]. Furthermore, Hsp40 can interact with substrate and seems to target Hsp70 to intended binding sites, supporting a “tagging” model [159] [160].

The defining feature of all Hsp40s is their J-domain. However, they can differ strongly in their other domains and are therefore subdivided into three classes, namely type I, type II and type III [161]:

**Type I Hsp40:** Type I Hsp40s comprise a J-domain, a glycine-and phenylalanine-rich region, a zinc finger-like region and a C-terminal domain. Examples for Type I Hsp40s are Ydj1 in yeast, Hdj2 in humans and DnaJ in E.coli.

**Type II Hsp40:** Type II Hsp40s comprise a J-domain, a glycine-and phenylalanine-rich region, a glycine/methioine-rich region and two C-terminal domains. In contrast to type I Hsp40, they lack a zinc finger-like region.

**Type III Hsp40:** Type III Hsp40s only comprise a J-domain and otherwise do not have any universal defining characteristics in common.

Most recent studies of the Hsp70/40 and also Hsp90 interaction with GR-LBD used Ydj1 as the Hsp40 co-chaperone [43] [68]. We started with the same approach using Ydj1, but later also conducted control experiments with Hdj2, Hdj1 and JD.

To gain new insights into the role of Hsp40, first each type of Hsp40 investigated was added to GR-LBD alone. Afterwards, each type of Hsp40 was tested on GR-LBD in combination with Hsp70.

### 10.1 Ydj1

As long as the GR-LBD was in the DEX-bound holo state, Ydj1 did not have any measurable effect on it in our experiments.

Fig.74 compares the rates of the N-terminal lid flipping of GR-LBD in the absence of Ydj1 (purple and blue data points, adapted from [44]) and in the presence of a very high Ydj1 concentration of 4  $\mu$ M (red data points). The flipping rates remained unaffected.

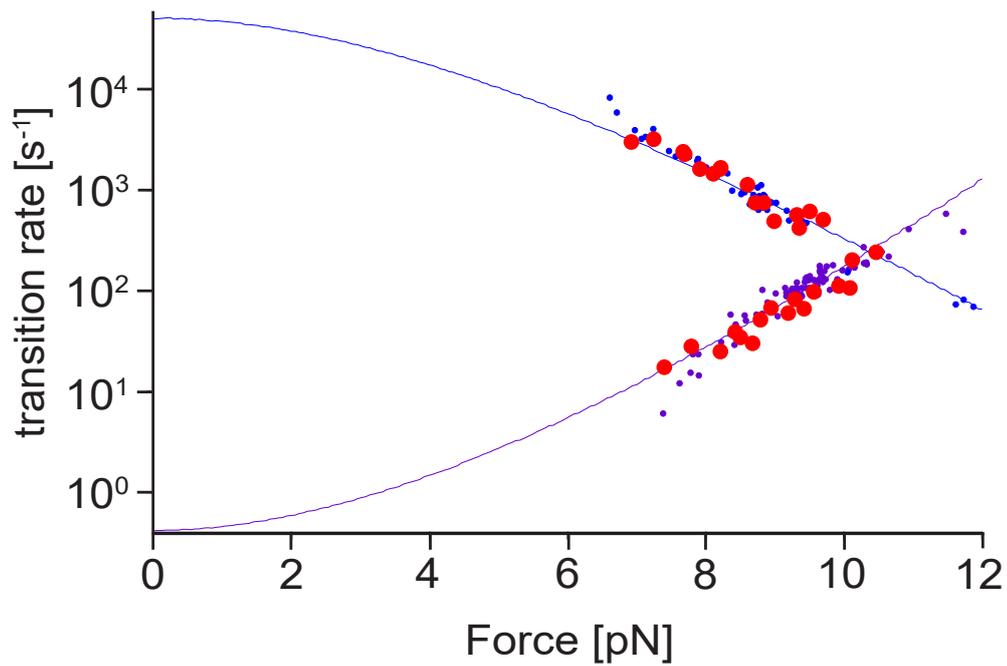


Figure 74: The N-terminal lid flipping rates are the same in the absence of Ydj1 (dark blue and purple data points, adapted from previous analysis by Suren [44]) and in the presence of a high Ydj1 concentration of 4  $\mu\text{M}$  (red data points).

After DEX dissociation, the presence of Ydj1 lead to the population of a new Ydj1-bound state (green) of apo GR-LBD, as illustrated in Fig.75:

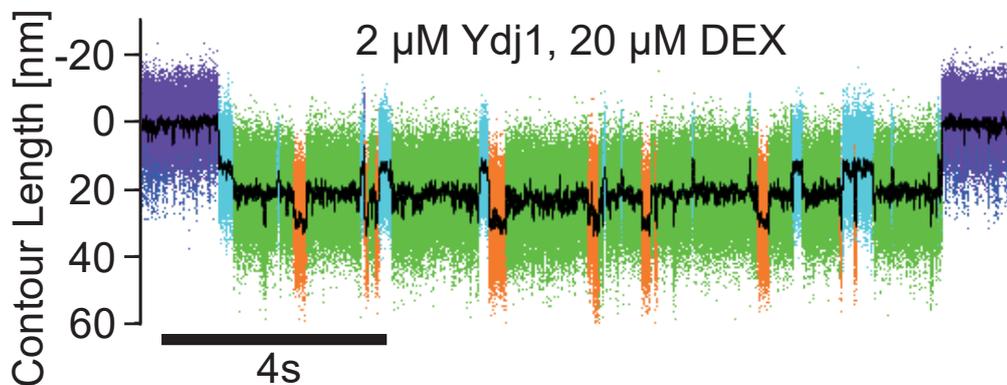


Figure 75: Passive mode trace of GR-LBD in the presence of 2  $\mu\text{M}$  Ydj1 and 20  $\mu\text{M}$  DEX. In the presence of Ydj1, the DEX-unbound GR-LBD populates a new Ydj1-bound state (green) at an unfolded contour length of 19 nm from the N-terminus.

This new Ydj1 bound state (green) was never observed in the absence of Ydj1. Also, Ydj1 could exclusively bind to the apo state of GR-LBD, i.e. the transition to the Ydj1-bound state always started from the DEX-unbound light-blue state. When in the Ydj1-bound state, the GR-LBD was unable to bind DEX. As a result, the binding of Ydj1 and DEX to GR-LBD is mutually exclusive, i.e. a true competition. The discussion in section 10.6 will argue that this Ydj1-bound state is actually an unphysiological state induced by mixing yeast Hsp40 with human GR-LBD and

therefore of only limited interest for further analysis. DEX dissociation rates from the lid-open state (dark-blue) remained the same in the presence as in the absence of Ydj1. Averaging the DEX dissociation rate over 6 different molecules in the presence of Ydj1 rendered  $k_{diss,DEX} \approx 3.7 \cdot s^{-1}$ , which is on the order of the previously published  $k_{diss,DEX} \approx 3 \cdot s^{-1}$ , cf. [44]. DEX rebinding rates to the DEX-unbound state (light-blue) also remained the same in the presence as in the absence of Ydj1. Averaging the DEX binding rate over 6 different molecules in the presence of Ydj1 rendered  $k_{bind,DEX} \approx 0.028 \cdot s^{-1} \mu M^{-1}$ , which is on the order of the previously published  $k_{bind,DEX} \approx 0.033 \cdot s^{-1} \mu M^{-1}$  [44].

The Ydj1 binding rate depended both on the Ydj1 concentration and on the applied force. Already a difference in force of less than 0.5 pN seemed to affect the Ydj1 binding rates heavily. Since the applied force cannot be adjusted with sub-picoNewton precision during an on-going experiment, this makes it challenging to separate the effects of Ydj1 concentration and force on the Ydj1 binding rate. To give at least a rough estimate, averaging the Ydj1 binding rate to the DEX-unbound state (transition from light-blue to green state) over 6 molecules, which were all recorded at a force close to 9 pN, rendered  $k_{bind,Ydj1} \approx 3.8 \cdot s^{-1} \mu M^{-1}$ . The Ydj1 dissociation rate averaged over 6 molecules was  $k_{diss,Ydj1} \approx 1.15 \cdot s^{-1}$ . As a result, the affinity of Ydj1 to the DEX-unbound state of GR-LBD at a force of  $\sim 9$  pN is  $k_D \approx 0.3 \mu M$ . Since Ydj1 can bind to a specific location in apo GR-LBD, a natural follow-up question was whether it could also bind in other places along the sequence of GR-LBD. To find out, we mechanically unfolded the GR-LBD completely and in the presence of Ydj1 and tried to identify unusually long intermediates along the way. However, as the trace in Fig.76 shows, no Ydj1-bound phases could be detected anywhere else apart from the previously discussed binding at 19 nm contour length (green):

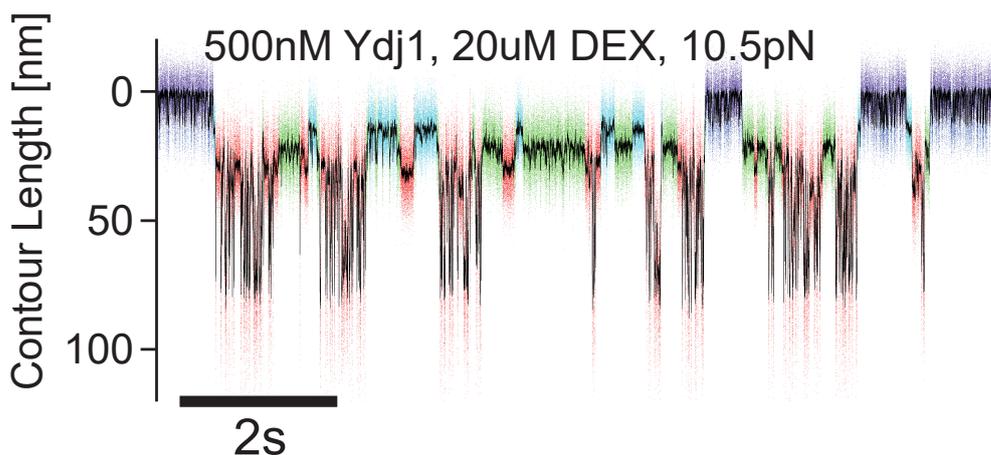


Figure 76: Passive mode trace of GR-LBD in the presence of 2  $\mu M$  Ydj1 and 20  $\mu M$  DEX. The GR-LBD unfolds and refolds in equilibrium under a high mechanical load. Apart from the previously discussed Ydj1 binding at 19 nm, the unfolding intermediates are indistinguishable from intermediates populated in the absence of Ydj1.

## 10.2 Hdj2

Ydj1 stems from yeast, while our experiments were conducted using human GR-LBD and human Hsp70. This "mixing of species" seems to have become the standard in the field, with Ydj1 performing exceptionally well as Hsp40 co-chaperone even in combination with human chaperone systems and human substrate proteins [43] [142]. However, we wanted to test whether human Hsp40 had the same effect as yeast Ydj1 on GR-LBD. In particular, it was of interest whether or not the new Ydj1 bound state could be reproduced with human Hsp40 and if Hsp70/40-induced unfolding of GR-LBD still worked.

To this end, we used Hdj2, the human equivalent of Ydj1, in experiments with GR-LBD. Fig.77 shows the resulting trace when Hdj2 is added to GR-LBD in the absence of Hsp70.

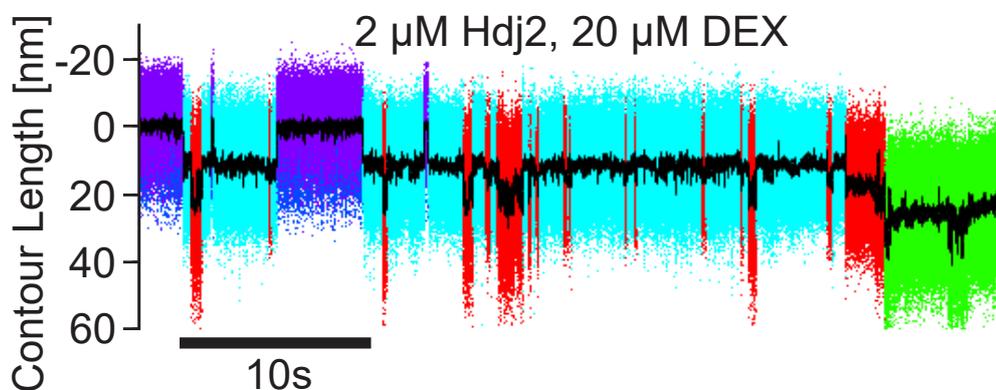


Figure 77: Passive mode trace of GR-LBD in the presence of 2  $\mu\text{M}$  Hdj2 and 20  $\mu\text{M}$  DEX. In the presence of Hdj2, the DEX-unbound GR-LBD does not show the previously observed Ydj1-bound state (cf. Fig.75). However, Hdj2 does interact with apo GR-LBD, trapping it in a partially unfolded state that is unable to refold or rebind DEX indefinitely (green phase at the end of trace).

Interestingly, the previously observed Ydj1-bound state could not be reproduced. Therefore, this state can be seen as an artefact resulting from the mixing of yeast Hsp40 with human GR-LBD.

However, Hdj2, too, interacted with GR-LBD even in the absence of Hsp70, albeit in a less precise and reliable way than Ydj1. As Fig.77 shows, after a few DEX dissociations and rebindings the GR-LBD becomes trapped in a partially unfolded state (green phase) and remains unable to refold or rebind ligand indefinitely. This specific kind of state only occurred in the presence of Hdj2. Increasing the Hdj2 concentration caused this state to appear even earlier, usually directly after the first DEX dissociation.

The fact that the occurrence of this green trapped phase depended on Hdj2 concentration is evidence that it is indeed caused by the presence of Hdj2. Apparently, Hdj2, just like Ydj1, also cannot attack holo GR-LBD, but can and eventually does interact with GR-LBD as soon as DEX dissociates. However, it does so in a totally distinct way from Ydj1. It is important to note this striking difference for anyone who performs experiments using Ydj1.

### 10.3 Hdj1

Hdj1 is a human type II Hsp40 co-chaperone and therefore not a homologue of Yjd1 (type I Hsp40). Since type II Hsp40s lack the G/F-rich region responsible for substrate interaction, it was of great interest to examine whether Hdj1 does not interact with GR-LBD on its own and whether the Hsp70/40-induced unfolding of GR-LBD still worked with a type II Hsp40.

Fig.78 shows a trace of GR-LBD in the presence of 2  $\mu\text{M}$  Hdj1.

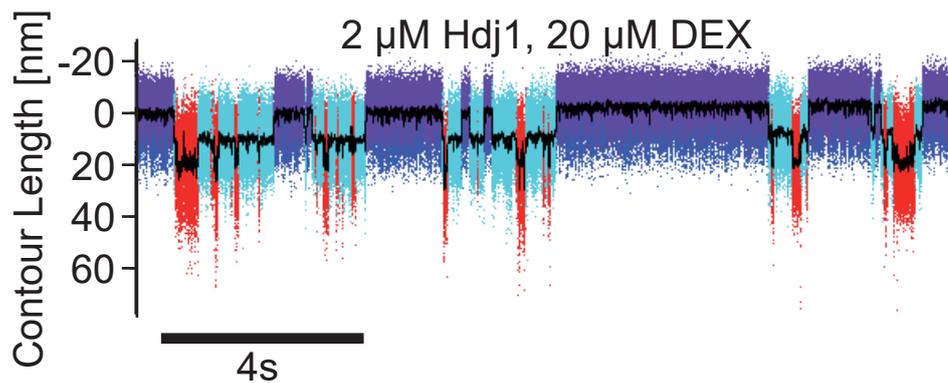


Figure 78: Passive mode trace of GR-LBD in the presence of 2  $\mu\text{M}$  Hdj1 and 20  $\mu\text{M}$  DEX. In our experiments, Hdj1 alone does not affect the GR-LBD.

As expected, Hdj1 in the absence of Hsp70 did not have any detectable effect on GR-LBD in our experiments.

### 10.4 J-Domain

We also conducted experiments using a construct that consisted of only the J-Domain of Ydj1 (JD). This construct was lacking all the regions responsible for substrate recognition in Hsp40. It therefore had the advantage that the stimulating effect of Hsp40 on Hsp70's ATPase activity could be studied in isolation, discarding any substrate interaction that Hsp40 might otherwise show.

Naturally we expected the J-Domain alone to not show any effect on GR-LBD, and this expectation was met by experiment, as Fig.79 illustrates.

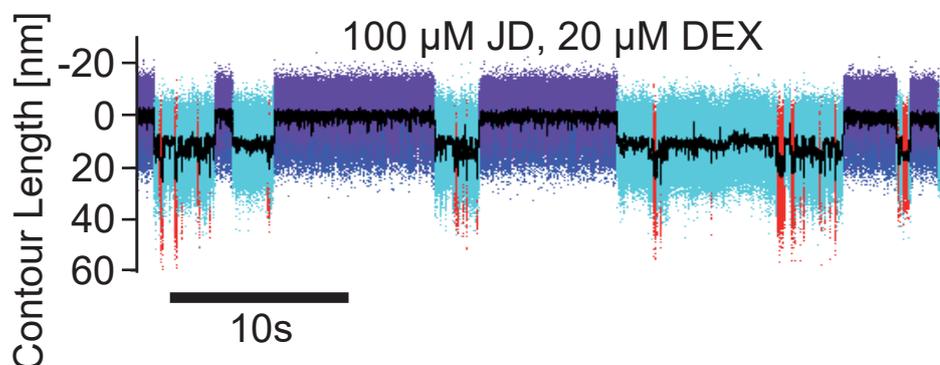


Figure 79: Passive mode trace of GR-LBD in the presence of 100  $\mu\text{M}$  JD and 20  $\mu\text{M}$  DEX. As expected, the JD alone does not affect the GR-LBD.

### 10.5 Hsp70/40 Combination for All Hsp40 Variants

We demonstrated significant differences in how the four Hsp40 variants affected the GR-LBD in the absence of Hsp70. However, in combination with Hsp70, all four Hsp40 variants lead to qualitatively very similar unfolding patterns. They are summarized in Fig.80:

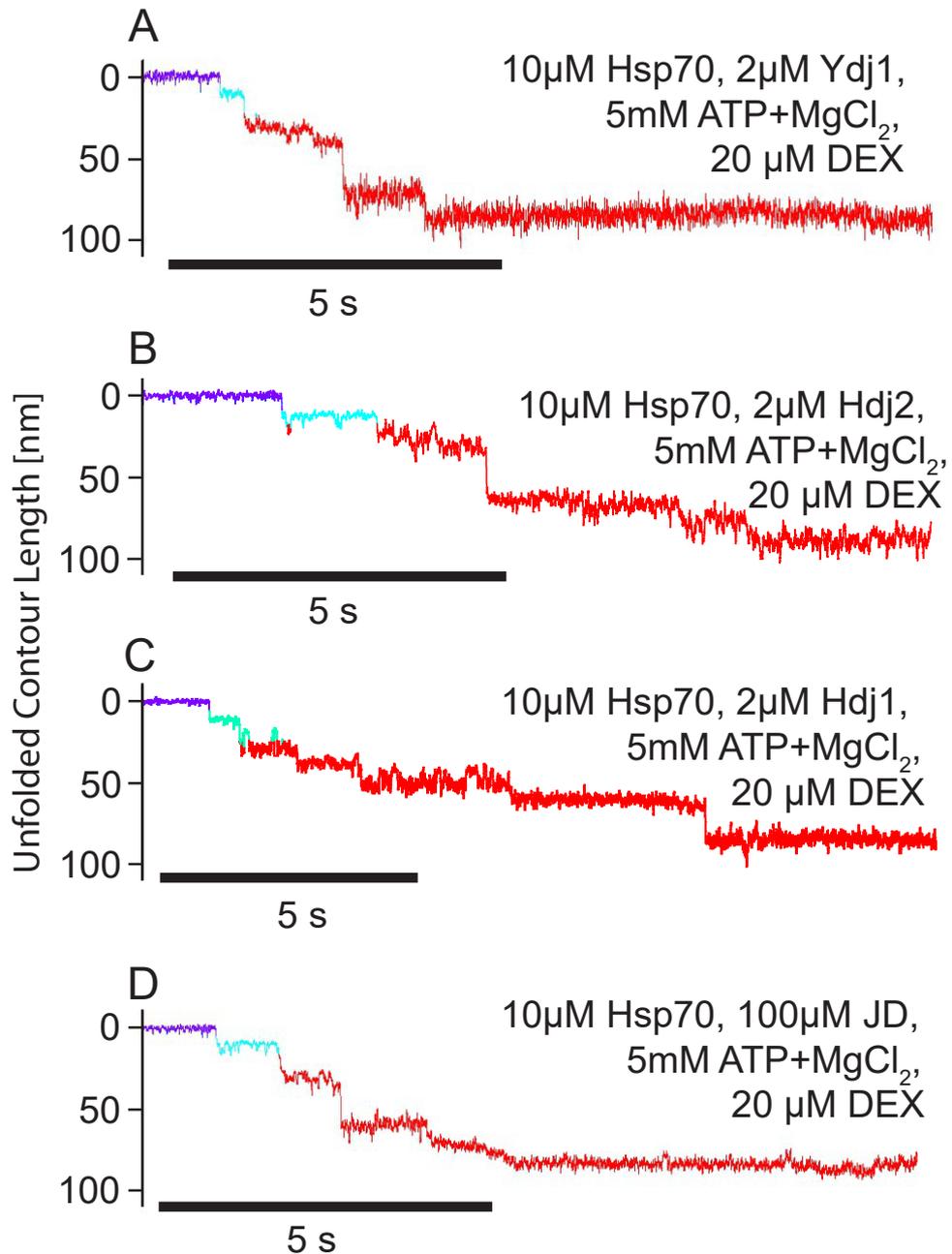


Figure 80: Hsp70/40-induced stepwise unfolding of GR-LBD works with all four types of Hsp40, namely Ydj1 (A), Hdj2 (B) and Hdj1 (C) and JD (D). All unfoldings display well-defined steps and occur on timescales around 5-10 seconds.

Using Ydj1 at an extremely high concentration of 15 µM in combination with 10 µM

Hsp70, in fact the binding of Ydj1 to apo GR-LBD seemed to delay Hsp70/40 induced unfolding, as shown in Fig.81. Since both Ydj1 and Hsp70 can only attack GR-LBD after DEX dissociation, it is conceivable that Ydj1 is blocking an Hsp70 binding site necessary for the initiation of Hsp70/40 induced unfolding:

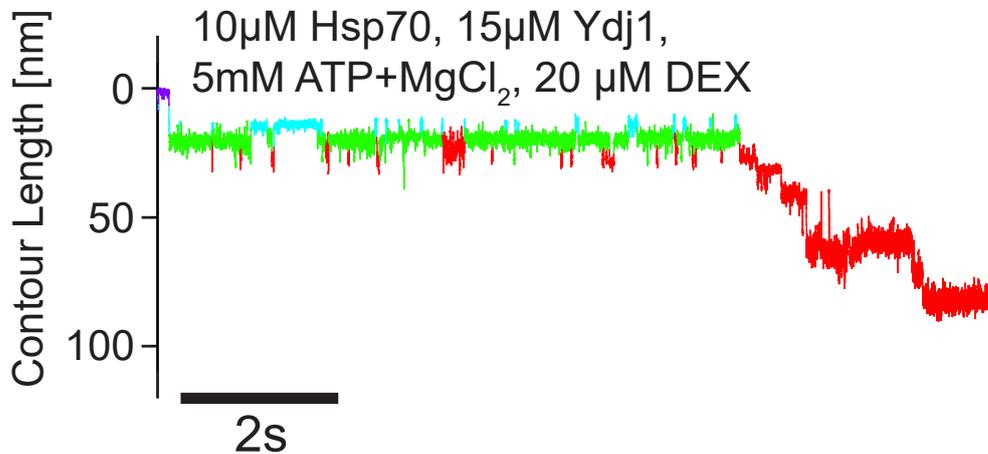


Figure 81: Hsp70/40 induced unfolding at 15  $\mu\text{M}$  Ydj1, 10  $\mu\text{M}$  Hsp70, 5 mM ATP and 5 mM  $\text{MgCl}_2$ . A high concentration of Ydj1 significantly delays Hsp70/40 induced unfolding. Most likely, Ydj1 binding inhibits Hsp70 from accessing an Hsp70 binding site necessary to initiate Hsp70/40 induced unfolding.

## 10.6 Discussion of the Four Hsp40 Variants and Their Effects

For all four different variants of Hsp40 (Yjd1, Hdj2, Hdj1 and JD), when added in combination with Hsp70 and MgATP, qualitatively very similar complete and step-wise unfoldings of the GR-LBD were demonstrated (Fig.80). Also, the stimulation of Hsp70's ATP hydrolysis rate by Hsp40 through variation of JD concentration was directly shown (Figs. 59).

Nonetheless, there were some remarkable differences in the effect of these four Hsp40 variants on GR-LBD.

In the case of Ydj1, we identified a new state where Ydj1 binds to apo GR-LBD (green state in Fig.75), but never to holo GR-LBD. This Ydj1-bound state reproducibly occurred at an unfolded contour length of 19 nm.

However, this state was not populated by Hdj2 and Hdj1 (Figs.77 and 78). Also, at high concentrations of Ydj1 ( $>2 \mu\text{M}$ ), this state in fact seemed to inhibit or delay the unfoldase activity of Hsp70 (Fig.81). Finally, Hsp70/40 unfoldings did not require this particular Ydj1-bound state, since, in combination with Hsp70, this state was not observed for Hdj2, Ydj2 or JD, even though unfolding was induced with similar efficiency (Fig.80). Moreover, even unfoldings with Hsp70/Ydj1 often did not show this Ydj1 bound state on the way (e.g. Fig.54). Therefore, population of this particular Ydj1-bound state does not seem to be necessary for efficient GR-LBD unfolding and is likely an artefact originating in the mixing of yeast and human molecules.

Apart from the Ydj1-bound state at 19 nm, we did not detect any further binding of Ydj1 to the GR-LBD. In force-induced complete unfoldings of the GR-LBD in the presence of Ydj1, GR-LBD did not exhibit any unprecedented long-lived intermediates, but in fact showed identical kinetics as in force-induced unfoldings in the absence of chaperones (Fig.76). If Ydj1 binds to GR-LBD to target Hsp70 to specific

binding sites in the sense of a “tagging mechanism”, this binding must be weak and transient, as we could not observe it in our experiments.

In the case of Hdj2 we observed irreversible binding to unfolded GR-LBD (Fig.77). This finding is consistent with earlier reports [143], where the *E.coli* homologue of Hsp40 DnaJ bound irreversibly to denatured rhodanese. Since much lower concentrations of Hsp40 can already stimulate ATP hydrolysis by Hsp70, this strong and irreversible binding is likely not the physiologically relevant mode of action.

JD, when added on its own, had no effect on the GR-LBD even at very high concentrations of 100  $\mu\text{M}$  (Fig.79). In combination with Hsp70, a striking difference between full length Ydj1 and the truncated JD construct was the fact that we needed 50-fold higher concentrations of JD as compared to Ydj1 to achieve similar unfolding rates at the same Hsp70 concentration. For instance,  $\sim 100 \mu\text{M}$  of JD vs. 2  $\mu\text{M}$  of Ydj1 were necessary in combination with 10  $\mu\text{M}$  Hsp70 for an unfolding rate of  $\sim 0.8 \text{ s}^{-1}$  (see graphs in Figs. 57 and 59). This effect has been reported before [38] [135]. The 50-fold higher efficiency of Ydj1 in promoting Hsp70/40 unfolding of GR-LBD as compared to JD can only be due to a region truncated in the JD mutant.

This finding is consistent with those regions creating transient interactions with the substrate, thus recruiting Hsp70 more efficiently to these sites [159] [160]. So in fact the difference in required concentrations of Ydj1 and JD is an indirect sign for some kind of tagging mechanism at work, even though it does not show up directly in our experiments.

## 11 Conclusion and Outlook

Coming back to the GR-LBD's intricate cycle involving ligand binding and chaperone interaction (Fig.5), the works of Suren and Mößner combined now cover roughly the first half of it with their single-molecule optical trapping experiments. The ligand binding and dissociation kinetics and the protein folding pathway have been analysed in [44]. In this work, the interaction of GR-LBD with Hsp70/40, i.e. the first major part of the chaperone system necessary for GR activation, has been studied.

With the improvements concerning the oxygen scavenging system and buffering conditions, the assay began to work strikingly reliably and has done so for the last three years. This facilitated the acquisition of large amounts of high quality GR-LBD data and enabled significantly more elaborate and complicated experiments.

Our detailed understanding of the GR-LBD's protein folding behavior, its interaction with DEX and Hsp70/40, as well as its biochemical peculiarities now serves as a sound foundation for future experiments.

The natural continuation of this route will include the addition of HOP, Hsp90 and p23 to eventually solve the entire cycle of GR-LBD in optical trapping experiments. First tentative experiments with these chaperones by Suren and Mößner showed promising results.

With the recent acquisition of two C-trap setups [162], the new possibility to perform single-molecule optical trapping experiments in a multi-channel microfluidics setup with additional fluorescence detection opened up for the Rief group. This offers completely novel ways to probe the GR-LBD and will likely prove essential for the complicated and crowded assays with the full chaperone system. It will also be of large interest to re-evaluate and confirm the experiments with Hsp70/40 at the C-trap. Certainly, additional knowledge can be obtained there.

Multiple channels allow for sequential step-by-step imitation of the ligand and chaperones cycle of GR-LBD, following the chronology of ligand and chaperones entering and leaving the process. Experiments with fluorescently labeled chaperones might give further insights into the exact role of Hsp40, in particular with respect to a tagging mechanism, the number of Hsp70s involved, and the precise chronology of binding events. By combining single-molecule optical trapping with single-molecule fluorescence detection, one will be able not just to *feel*, but also to *see*.

# Appendix

## 12 Methods and Protocols

### 12.1 Maleimide Oligos Attachment Protocol

#### Overview:

Our maleimide oligos consist of 34 bases of single-stranded DNA with a maleimide group at their end. For high binding efficiency, the terminal cysteines of the GR-LBD are first reduced by exposing the GR-LBD to a large molar excess of TCEP (at least 10:1 per cysteine) for 1 hour. After a subsequent buffer exchange, maleimide oligos are added at a molar excess of 2:1 (one for each of the two termini). It is crucial to keep the pH between 6.5 and 7.5, since in this pH range the affinity of maleimides to cysteines is 1000-fold higher than its affinity for other sidegroups in proteins. After at least 2 hours of incubation with oligos at room temperature, this sample is run over an HPLC column to extract the protein with two attached oligos.

#### Detailed protocol:

1. Reduce the terminal cysteines of your protein by adding TCEP to your buffer. The molar excess of TCEP should be at least 10x over cystein amount, so 20x the protein concentration. 2mM TCEP is a typical concentration. The pH for this reaction should lie between 6.5 and 7.5. Let this react for at least half an hour at room temperature.
2. Get rid of unreacted TCEP. You can use Amikon filters of a suitable size for your molecule. I use Amikon 10k filters for a molecule of 30kDalton size. Put your sample in the filter and fill it up to 500  $\mu$ L. Then spin it down until you reach a volume of  $\sim$ 100  $\mu$ L. Fill the filter up to 500  $\mu$ L again, and spin it down to  $\sim$ 100  $\mu$ L again. Repeat this another time. With each run you reduce the amount of TCEP by a factor of 1:5, so after three runs, you have diluted the TCEP by a factor of 1:125, resulting in less than 1% of initial TCEP. Another possibility to get rid of unreacted TCEP is to run your sample over an HPLC column and collect your protein peak, but this can lead to a large loss of concentration of protein.
3. Add maleimide oligos (diluted in 10x PBS buffer) to your protein sample at a molar excess of 1:1 per cysteine (i.e. 2:1 per protein). It is crucial to keep the pH between 6.5 and 7.5 during this reaction (therefore the 10x PBS buffer). In this pH range, the affinity of the maleimide for cysteine is 1000-fold higher than for other amino acids or the wrong side groups. Let this incubate at room temperature for at least 2 hours.
4. Run your sample over an HPLC column. Ideally, you will see three peaks. From right to left, these peaks are: 1. unreacted oligos 2. protein with one oligo 3. protein with two oligos. The peak on the left, i.e. protein with two oligos, is the one you should collect. Additional peaks might be due to dimerization of maleimide oligos. Be aware that some of our HPLC machines have a delay volume, so the fractions might not be "what you see is what you get"; but shifted by two fractions or so.

5. Determine the concentration of the fractions of interest by determining the concentration of the single-stranded DNA oligos with nanodrop. Usually, the resulting values are in the order of  $10 \text{ ng } \mu\text{L}^{-1}$ .

## 12.2 Azide Oligos Attachment Protocol

### Overview:

The attachment of azide oligos works via a DBCO-maleimide linker. First, the maleimides of the DBCO-maleimide linker react with the terminal cysteines. In a second reaction step, azide modified DNA oligos react with the DBCO. The azide oligos are similar to the maleimide oligos of the previous section, consisting of 34 bases of single-stranded DNA, but with an azide group at their end.

The protocol again starts with a reduction of the terminal cysteines using TCEP (at least 10:1 per cysteine), as in the previous section on maleimide oligos attachment. After a buffer exchange to remove the TCEP, the maleimide-DBCO linker is added, which binds to the terminal cysteines. After at least 2 hours incubation time and another buffer exchange to remove unreacted DBCO-maleimide, azide oligos are added at a molar ratio 1:1 per cysteine (i.e. 2:1 per protein). This is incubated overnight. On the following day, this sample is run over an HPLC column to extract the protein with two attached oligos.

### Detailed protocol:

1. Reduce the terminal cysteines of your protein by adding TCEP to your buffer. The molar excess of TCEP should be at least 10x over cysteine amount, so 20x the protein concentration. 2mM TCEP is a typical concentration. The pH for this reaction should lie between 6.5 and 7.5. Let this react for at least half an hour at room temperature.
2. Get rid of unreacted TCEP. You can use Amikon filters of a suitable size for your molecule. I use Amikon 10k filters for a molecule of 30kDalton size. Put your sample in the filter and fill it up to 500  $\mu\text{L}$ . Then spin it down until you reach a volume of  $\sim 100 \mu\text{L}$ . Fill the filter up to 500  $\mu\text{L}$  again, and spin it down to  $\sim 100 \mu\text{L}$ . Repeat this another time. With each run you reduce the amount of TCEP by 1:5, so after three runs, 1:125, resulting in less than 1% of initial TCEP. Another possibility to get rid of unreacted TCEP is to run your sample over an HPLC column and collect your protein peak, but this can lead to a large loss of concentration.
3. Freshly dilute maleimide-DBCO in DMSO. Aim for a concentration of 20mM. Then, further dilute the maleimide-DBCO from 20mM to 1mM in your buffer.
4. Add maleimide-DBCO (1mM in buffer) to your protein sample at a molar excess of 5 per protein (i.e. 2.5 per cysteine). Make sure to add it slowly and then mix the sample well by pipetting up and down a few times. Let this incubate at room temperature for at least 2 hours. Check if your protein is precipitating once in a while. If it does, you might consider spinning the sample down during the reaction to get rid of aggregates. If a pellet is forming while you spin it down, you have proof that precipitation is a problem with your protein during this step.
5. Get rid of unreacted maleimide-DBCO. For this, perform the same buffer ex-

change as in step 2.

6. Determine the concentration of protein and DBCO. Both absorb at 280 nm, but DBCO has another peak at 309 nm. DBCO absorbs at 309 nm with an extinction coefficient of  $12\,000\text{ M}^{-1}\text{ cm}^{-1}$ . Therefore, determine the concentration of DBCO at 309 nm first. DBCO absorbs at 280 nm 1.089 times more than at 309 nm. Measure the absorption of the sample at 280 nm and subtract from it the absorption of DBCO at 309 nm multiplied by 1.089. What remains is the absorption of your protein. Using the extinction coefficient of your protein, you can then determine the concentration of your protein. The concentration of DBCO-maleimide should be twice as high as the concentration of your protein (2:1 per cysteine). However, Thomas and I usually get a labeling of 3:1 for some reason. Maybe one of the extinction coefficients we are using is wrong. However, this labeling of 3:1 is usually a good sign, the resulting attachments usually went well.

7. Add azide-oligos at a molar excess of 2:1 per protein (1:1 per cysteine) to the sample and let this incubate over night.

8. Run your sample over an HPLC column. Ideally, you will see three peaks. From right to left, these peaks are: 1. unreacted oligos 2. protein with one oligo 3. protein with two oligos. The peak on the left, protein with two oligos, is the one you should collect. Be aware that some of our HPLC machines have a delay volume, so the fractions might not be "what you see is what you get"; but shifted by two fractions or so.

9. Determine the concentration of the fractions of interest by determining the concentration of the single-stranded DNA oligos with nanodrop. Usually, the resulting values are in the order of  $10\text{ ng }\mu\text{L}^{-1}$ .

### 12.3 Comparison of Maleimide Oligos vs. Azide Oligos

The attachment using Maleimide Oligos works faster, requiring only two hours of incubation time between maleimide and cysteines and one buffer exchange. The attachment via DBCO-maleimide linker and azide oligos, while slower due to the overnight incubation and two buffer exchanges, usually resulted in a better reaction efficiency. It showed a higher percentage of proteins with two attached oligos in the subsequent HPLC run.

## 12.4 DNA Handles Hybridization

Our DNA handles consist of 544 base pairs. Since each base pair is 3.3 Å, this results in a total length of 185 nm of one DNA handle.

The handles are produced in a Polymerase Chain Reaction (PCR) and afterwards diluted or concentrated to the desired storage concentration.

When added to the Protein+Oligo construct, I typically needed ~200 ng for ~10 ng protein.

## 12.5 Quality Control of Oligo Attachment and DNA Handles Hybridization via Agarose Gel

To check whether the end product of the oligo attachment step indeed was a protein with two attached oligos (Protein+Oligos), and also to find the best ratio between the DNA handles and Protein+Oligos construct, an agarose gel was routinely run after the HPLC filtering process. To this end, several samples of the Protein+Oligo construct were incubated with different amounts of DNA handles for 1 hour.

Typically, I needed 200 ng DNA handles for 10 ng or Protein+Oligos construct.

An ideal outcome would be a gel line as in Fig.82:

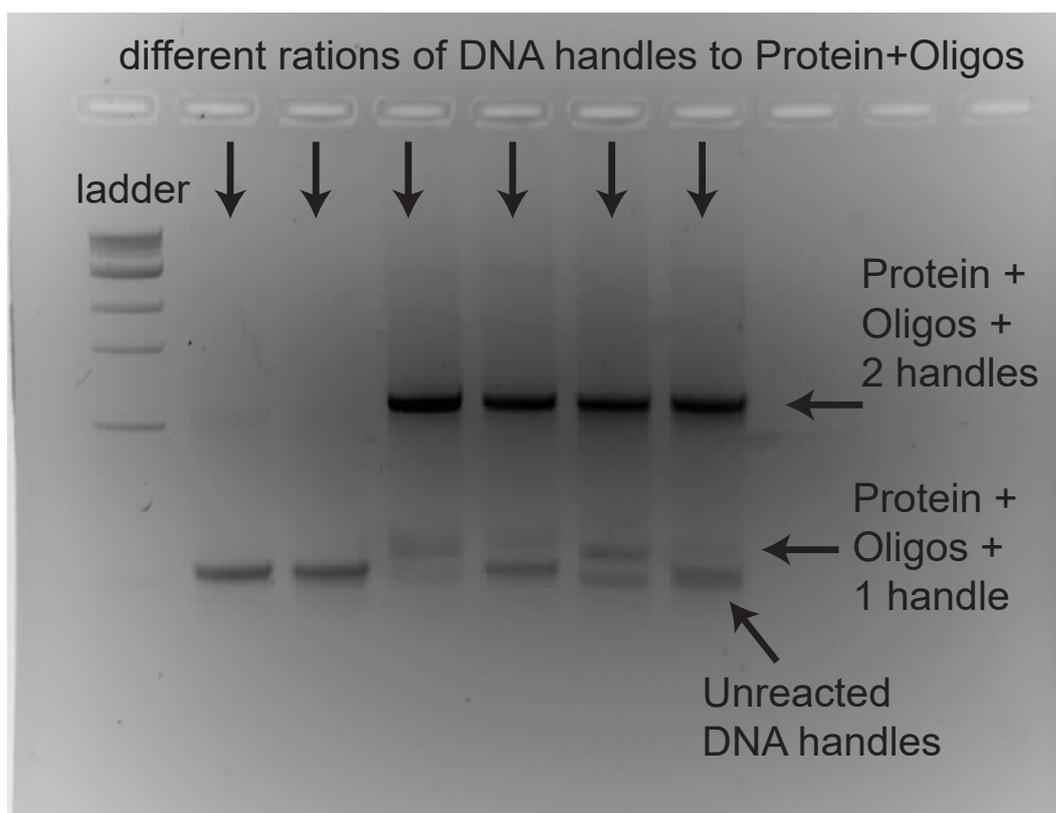


Figure 82: **Agarose control gel:** An agarose gel is routinely run after each oligo attachment to confirm that protein has indeed bound two DNA oligos. Also, the best ratio between DNA handles and Protein+Oligos construct is determined. If the oligo attachment has worked, there are three lines visible on the gel: one that contains only unreacted DNA handles, one that consists of Protein+Oligos+1handle, and the intended line signifying Protein+Oligos+2handles

Each column on the agarose gel shows a weak mark of unreacted handles, another

weak mark where Protein+Oligos reacted with just one handle, and a big population at the length indicative of Protein+Oligos that reacted with two handles. If the amount of handles was too little, the population of Protein+Oligos reacted with one handle was stronger, while the population of Protein+Oligos with two attached handles was weaker. If the amount of handles was too large, the line indicating unreacted handles was very strong. Since no further filtering of the sample between handle attachment and optical trapping experiments was possible, unreacted handles posed a problem, occupying too many binding sites on the beads later on.

## 12.6 Preparation of the Optical Trap Measurements

For tethering the protein to the beads in our optical trap setup, we used a protocol similar to the one described by Cecconi et al. [163]. The protein was incubated with 34 bp 3'-maleimide modified oligonucleotides for 2h at room temperature. The desired oligo-protein-oligo construct was then again purified by size exclusion using a Yarra 3u SEC-3000 column, concentrated to about 0.5  $\mu\text{M}$ , shock-frozen and stored in aliquots at  $-80^\circ\text{C}$ . On each measurement day, the construct was incubated for 1h on ice with 180 nm long dsDNA handles that could hybridize to the oligonucleotides. Proper construct formation including 2 dsDNA handles was checked on an Agarose gel. At the other end, half of the handles were biotin-modified, while the other half were digoxigenin-modified. The whole construct was incubated 20 min with 1  $\mu\text{m}$ -sized streptavidin-coated beads (polysciences, Inc.) before mixing with antidigoxigenin-coated beads. The measurement chambers were made by attaching parafilm (Bemis Company) between two 170  $\mu\text{m}$ -thick coverslips (Carl Roth). A custom built dual-beam optical trap as in [164] was used to trap the two different kinds of beads, one in the fixed beam and the other one in the mobile beam, which can be moved using a piezo mirror. The construct was tethered between the two beads by bringing the bead surfaces together in close proximity. Data was recorded with a sampling frequency of 30 kHz. All measurements were performed at  $23^\circ\text{C}$  in 40 mM Tris, 150 mM NaCl and 1 mM DTT at pH 8.0 with the addition of 0 to 200  $\mu\text{M}$  Dexamethasone (DEX, Sigma D1756). An oxygen scavenging system was added consisting of pyranose oxidase, catalase and glucose as described in [125].

## 12.7 Global Fit Parameters

Table 1: Global fit parameters extracted from the graphs in Figs. 57 and 59:

$k_{on,Hsp70}$	$0.20 \pm 0.03 s^{-1} \mu\text{M}^{-1}$
$k_{on,Ydj1}$	$0.78 \pm 0.20 s^{-1} \mu\text{M}^{-1}$
$k_{on,JD}$	$0.016 \pm 0.004 s^{-1} \mu\text{M}^{-1}$
$k_{max,hydrolysis}$	$> 1000 s^{-1}$

## 12.8 Assembly of Dumbbell Assay

### 12.8.1 Protein+Oligo+Handles (POH)

First, you need to incubate the protein-oligo complex (PO) with DNA handles to get "Protein+Oligo+Handles" (POH). Usually, a combination of 10 ng of protein with 100 ng of handles works well. Thomas and I found that for the GR project, a combination of 8ng protein with 200 ng handles works even better. The ideal values might change depending on the quality and concentration of your protein-oligo attachment and your DNA handles.

POH:

10 ng protein-oligo (typically around 0.5  $\mu$ L)

200 ng handles (typically around 1ul)

Mix well and avoid air bubbles.

Let this incubate for 1h, then fill up to 100  $\mu$ L with buffer and again mix well. Avoid air bubbles.

### 12.8.2 Protein+Oligo+Handles' (POH')

So far, you have POH, where 10 ng of protein reacted with 200 ng handles, and filled this up to 100  $\mu$ L.

POH' is now a dilution of this POH. This is necessary because the POH is still too highly concentrated to directly add it to the beads.

How much you dilute the POH in the POH' step varies strongly and largely depends on the beads you are using, i.e. their number of binding sites etc., and the concentration of your POH. Normally, I dilute the POH by 1:20 or so:

POH':

38  $\mu$ L Buffer

2  $\mu$ L POH

### 12.8.3 Incubation with Streptavidin Beads (S)

In this step, the incubation with streptavidin beads occurs:

S:

10  $\mu$ L Buffer

5  $\mu$ L POH'

Mix well

5  $\mu$ L streptavidin beads diluted 1:30 as compared to main stock, vortex well before use.

Let this incubate for 20min.

#### 12.8.4 Final Solution (F)

For the final solution, we first need to prepare a buffer containing the scavenging system (glucose, catalase, pyranose oxidase) and whatever else we want to have in the chamber (e.g. DTT, hormone, ATP, MgCl<sub>2</sub>, chaperones,...). Then we need to filter this buffer using 20 nm filters, since we want very clean chambers and the pyranose oxidase in particular, but also the Dexamethasone for example, are dirty. Since the 20 nm filters have a certain dead volume (about 50  $\mu$ L), we first prepare 200  $\mu$ L of a first solution F1, then filter it, and then take the amount we need to make the final mixture F2. The F2 is what we eventually fill into the chamber and then measure at the trap. The anti-dig beads are prediluted to 1:150 as compared to the main stock.

F1:

186  $\mu$ L Buffer  
4  $\mu$ L glucose  
4  $\mu$ L catalase  
4  $\mu$ L pyranose oxidase  
2  $\mu$ L DTT at 100 mL

filter with 20 nm filter  $\rightarrow$  F1 filtered

F2:

98.7  $\mu$ L F1 filtered  
0.7  $\mu$ L S  
0.6  $\mu$ L anti-dig beads, diluted 1:150 as compared to main stock, vortex well before use.

$\rightarrow$  Flush measurement chamber first with 200  $\mu$ L ddH<sub>2</sub>O, then 200  $\mu$ L buffer, and then with 100  $\mu$ L F2.

Start your measurement at the optical trap.

## 12.9 Protein Expression and Sequences

All protein constructs were prepared using standard recombinant techniques described below. Purifications were conducted by Ulrike Majdic, Vinay Dahiya, Jannis Lawatschek or Daniel Rutz.

### 12.9.1 GR-LBD

UniProt-P04150 (residues: 521-777)

Protein expression:

Human GR-LBD variants (aa 527-777) were expressed in BL21 (DE3 RIPL) cells at 18°C overnight in ZYM-5052 auto-induction media supplied with 500 µM Dexamethasone (DEX) (Sigma-Aldrich, St. Louis, USA). Cells were harvested by centrifugation for 15 min at 7,000 rpm and 4°C (Beckman Avanti J-26 XP, Beckman Coulter, Brea, California) and washed with ice-cold PBS. Cells were resuspended in Ni-A buffer (50 mM Tris, 2 M Urea, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 10 mM Imidazole, 2 mM β-mercaptoethanol, 50 µM DEX, pH 7.9) supplemented with DNaseI (Roche, Basel, Swiss) and Protease Inhibitor HP (Serva electrophoresis GmbH, Heidelberg, Germany). Cell suspension was lysed by sonication (Bandelin Sonoplus UW2200, Bandelin electronic, Berlin Germany) or french press (Constant Systems Limited, Low March, UK) and centrifuged for 1 hour at 20,000 rpm and 4°C. Cleared lysate was applied onto a Ni-column (5 mL) FF,GE Healthcare, Chalfont St. Giles, Great Britain), pre-equilibrated in Ni-B buffer (50 mM Tris, 500 mM NaCl, 10 mM Imidazole, 10% Glycerol, 2 mM β-mercaptoethanol, 50 µM DEX pH 7.9). The column was then gradient-equilibrated in Ni-B buffer and His6-Halo-GR-LBD was eluted with Ni-C buffer (50 mM Tris, 500 mM NaCl, 350 mM Imidazole, 10% Glycerol, 2 mM β-mercaptoethanol, 50 mM DEX pH 7.9). IMAC-Buffers for the purification of apo-GR-LBD were supplied with 2 mM ATP to prevent binding of E. coli GroE and DnaK. GR-protein containing fractions were pooled, supplemented with His6-tagged TEV protease and dialyzed against 50 mM Tris, 100 mM NaCl, 10% Glycerol, 2 mM β-mercaptoethanol, 0.5% CHAPS, 50 µM DEX pH 7.9 overnight. Then, digested protein was passed through a Ni-column to remove Halo-tag-Fusion and TEV protease. The flow through was concentrated and loaded onto a gel filtration column (Superdex 200, 16/60 pg, GE Healthcare, Chalfont St. Giles, Great Britain) equilibrated in GR-storage buffer (25 mM Tris, 100 mM NaCl, 10% Glycerol, 0.5% CHAPS, 2 mM DTT, 50 µM DEX pH 7.9). GR-proteins were shock-frozen and analyzed by SDS-PAGE.

Sequence:

N-SACK-GR-LBD F602S/C638D-KCL-C (GRSD): SACKQLTPTLVSLLEVIEPE  
VLYAGYDSSVPDSTWRIMTTLNMLGGRQVIAAVKWAKAIPGFRNLHLDDQ  
MTLLQYSWMSLMAFALGWRSYRQSSANLLCFAPDLIINEQRMTPDMYDQ  
CKHMLYVSSSELHRLQVSYEEYLCMKTLTLLSSVPKDGLKSQELFDEIRMTYI  
KELGKAIVKREGNSSQNWQRFYQLTKLLDSMHEVVENLLNYCFQTFLDKT  
MSIEFPEMLAEIITNQIPKYSNGNIKKLLFHQKCL

### 12.9.2 Ydj1

UniProt-P25491

**Protein expression:**

Ydj1 with a cleavable solubility tag (SUMO) was expressed in BL21 (DE3 RIPL) cells overnight at 30°C. Cells were harvested by centrifugation for 15 min at 4,600 rpm and 4°C (Centrifuge Rotanta 460R Hettich Zentrifugen, Tuttlingen) and resuspended in Ni-buffer A (40 mM HEPES, 150 mM KCl, 350 mM NaCl, 20 mM MgCl<sub>2</sub>, 5% Glycerol, 10 mM Imidazole, 1 mM DTT pH 7.5) supplemented with DNaseI (Roche, Basel, Swiss) and Protease Inhibitor AEBSF-HCl (Serva electrophoresis GmbH, Heidelberg, Germany). Cell suspension was lysed by high pressure (Cell disruptor CF1m LTD, Constant Systems, Daventry Northants, United Kingdom) and centrifuged for 1 hour at 38,000 rpm and 4°C. Cleared lysate was applied onto a Ni-column (5 mL FF, GE Healthcare, Chalfont St. Giles, Great Britain), pre-equilibrated in Ni-buffer A. After washing with Ni-buffer A and 6% Ni-buffer B (40 mM HEPES, 150 mM KCl, 350 mM NaCl, 20 mM MgCl<sub>2</sub>, 5% Glycerol, 300 mM Imidazole, 1 mM DTT pH 7.5) the protein was eluted in 100% Ni-buffer B. Ydj1 containing fractions were pooled and 1:4 diluted with ice-cooled H<sub>2</sub>O. The solution was supplemented with DTT (1 mM) and His6-tagged SENP2 protease and dialyzed against GF-buffer (40 mM HEPES, 150 mM KCl, 5 mM MgCl<sub>2</sub> pH 7.5) overnight. Digested protein was passed through a Ni-column to remove SUMO-tag-Fusion and SENP2 protease. The flow through was concentrated and loaded onto a gel filtration column (Superdex 200 Increase 10/300 GL, Sigma-Aldrich, St. Louis, U.S.A) equilibrated in GF-buffer. Ydj1-proteins were shock frozen and analyzed by SDS-PAGE.

**Sequence:**

MVKETKFDYDILGVPVTATDVEIKKAYRKCALKYHPDKNPSEEAEEKFKEA  
 SAAYEILSDPEKRDIYDQFGEDGLSGAGGAGGFPGGGFPGFDDIFSQFFGA  
 GGAQRPRGPQRGKDIKHEISASLEELYKGRATAKLALNKQILCKECEGRGGK  
 KGAVKKCTSCNGQGKIFVTRQMGPMIQRFFQTECDVCHGTGDIIDPKDRCK  
 SCNGKKVENERKILEVHVPEGMKDGQRIVFKGEADQAPDVIPGDVVVIFVSE  
 RPHKSFKRDDGDDLVEAEIDLLTAIAGGEFALEHVSVDWLVKVGIVPGEVIAP  
 GMRKVIEGKGMPIPKYGGYGNLIKFTIKFPENHFTSEENLKKLEELPPRIV  
 PAIPKKATVDECVLADFPKYNRTRASRGGANYSDEEEQGGEGVQCAS  
 Q

**12.9.3 Human Hsp70**

UniProt-P0DMV8

Protein expression: Expressed and purified as described for Ydj1. Sequence:

MAKAAAIGIDLGTITYSCVGVFQHGKVEIANDQGNRTTPSYVAFTDTERLIG  
 DAAKNQVALNPQNTVFDKRLIGRKFQDPVVQSDMKHWPFQVINDGDKP  
 KVQVSYKGETKAFYPEEISSMVLTKMKEIAEAYLGYPTNAVITVPAYFND  
 SQRQATKDAGVIAGLNVLRINEPTAAAIAYGLDRTGKGERNVLIFFDLGGGT  
 FDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNHVFVEEFKRKHKKDISQN  
 KRAVRRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEELCSDL  
 FRSTLEPVEKALRDAKLDKAQIHDLVLVGGSTRIPKVQKLLQDFFNDRDLN  
 KSINPDEAVAYGAAVQAAILMGDKSENVQDLLLLDVAPLSLGLTAGGVMT  
 ALIKRNSTIPTKQTQIFTTYSNQPGLVLIQVYEGERAMTKDNNLLGRFELSGI  
 PPAPRGVPQIEVTFDIDANGILNVTATDKSTGKANKITITNDKGRLSKEEIER  
 MVQEAKEYKAEDDEVQRERVSNALESYAFNMKSAVEDEGLKKGKISEADK  
 KKVLDKCQEVISWLDANTLAEKDEFEHKRKELEQVCNPIISGLYQGAGGPG

PGGFQAQGPKGSGSGPTIEEVD

#### 12.9.4 J-Domain

UniProt-P25491, residues: 1-103

Protein expression: Expressed and purified as described for Ydj1. Sequence:  
GSMVKETKIFYDILGVPVTATDVEIKKAYRKCALKYHPDKNPSEEAEEKFK  
EASAAYEILSDPEKRDIYDQFGEDGLSGAGGAGGFPGGGFGFGDDIFSQFFG  
AGG

#### 12.9.5 Hdj2

UniProt-P31689

Protein expression: Expressed and purified as described for Ydj1, with the exception of the composition of Ni-buffer A (40 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 20 mM Imidazole, 2 mM DTT, 10% Glycerol pH 8.0), Ni-buffer B (40 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 500 mM Imidazole, 2 mM DTT, 10% Glycerol pH 8.0) and GF-buffer (40 mM HEPES, 150 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, pH 8.0).

Sequence:

MVKETTYDVLGVKPNATQEELKKAYRKLALKYHPDKNPNEGEKFKQISQ  
AYEVLSDAKKRELYDKGGEQAIKEGGAGGGFGSPMDIFDMFFGGGGRMQ  
RERRGKNVVHQLSVTLEDLYNGATRKLALQKNVICDKCEGRGGKKGAVEC  
CPNCRGTGMQIRIHQIGPGMVQQIQSVCMECQGHGERISPKDRCKSCNGRK  
IVREKKILEVHIDKGMKDGQKITFHGEGDQEPGLEPGDIIIIVLDQKDHAVFT  
RRGEDLFMCMDIQLVEALCGFQKPISTLDNRTIVITSHPGQIVKHGDIKCVL  
NEGMPIYRRPYEKGRLLIEFKVNFPENGFLSPDKLSLLEKLLPERKEVEETDE  
MDQVELVDFDPNQERRRHYNGEAYEDDEHHPRGGVQCQTS

## 12.10 ChaperISM Algorithm Prediction

The prediction for Hsp70 binding sites in the sequence of GR-LBD by the ChaperISM algorithm [138] [140] using the -qn flag as well as the -ql flag was the following:

-qn flag: (cutoff, default: 2.7)

Sequence: 1M2Z\_1—Chains A, C[auth D]—  
glucocorticoid receptor—Homo sapiens (9606)

POSITION	HEPTAMER	SCORE	
0	FQSPGSI	2.348578	
1	QSPGSIV	2.287828	
2	SPGSIVP	2.193302	
3	PGSIVPA	2.266039	
4	GSIVPAT	2.357289	
5	SIVPATL	2.559595	
6	IVPATLP	2.500008	
7	VPATLPQ	2.377258	
8	PATLPQL	2.408303	
9	ATLPQLT	2.499553	
10	TLPQLTP	2.367228	
11	LPQLTPT	2.367228	
12	PQLTPTL	2.367228	
13	QLTPTLV	2.593911	
14	LTPTLVS	2.558972	
15	TPTLVSL	2.558972	
16	PTLVSLL	2.725449	*
17	TLVSLLE	2.681196	
18	LVSLEEV	2.816629	*
19	VSLLEVI	2.776178	*
20	SLLEVIE	2.505243	
21	LLEVIEP	2.445655	
22	LEVIEPE	2.143675	
23	EVIEPEV	2.112630	
24	VIEPEVL	2.414610	
25	IEPEVLY	2.534778	
26	EPEVLYA	2.449826	
27	PEVLYAG	2.549500	
28	EVLYAGY	2.896350	*
29	VLYAGYD	2.915167	*
30	LYAGYDS	2.748072	*
31	YAGYDSS	2.549932	
32	AGYDSSV	2.429764	
33	GYDSSVP	2.297440	
34	YDSSVPD	2.216582	
35	DSSVPDS	1.929319	
36	SSVPDST	2.046006	
37	SVPDSTW	2.144945	

POSITION	HEPTAMER	SCORE	
38	VPDSTWR	2.395253	
39	PDSTWRI	2.385847	
40	DSTWRIM	2.344930	
41	STWRIMT	2.461617	
42	TWRIMTT	2.493279	
43	WRIMTTL	2.659757	
44	RIMTTLN	2.494541	
45	IMTTLNM	2.143728	
46	MTTLNML	2.184178	
47	TTLNMLG	2.280517	
48	TLNMLGG	2.244688	
49	LNMLGGR	2.463334	
50	NMLGGRQ	2.300133	
51	MLGGRQV	2.533504	
52	LGGRQVI	2.791698	*
53	GGRQVIA	2.666295	
54	GRQVIAA	2.743198	*
55	RQVIAAV	2.914459	*
56	QVIAAVK	2.878636	*
57	VIAAVKW	2.942636	*
58	IAAVKWA	2.848278	*
59	AAVKWAK	2.905074	*
60	AVKWAKA	2.905074	*
61	VKWAKAI	2.990026	*
62	KWAKAIP	2.763344	*
63	WAKAIPG	2.544693	
64	AKAIPGF	2.673598	
65	KAIPGFR	2.851169	*
66	AIPGFRN	2.570407	
67	IPGFRNL	2.695810	
68	PGFRNLH	2.649884	
69	GFRNLHL	2.907611	*
70	FRNLHLD	2.826754	*
71	RNLHLDD	2.513886	
72	NLHLDDQ	2.298516	
73	LHLDDQM	2.264288	
74	HLDDQMT	2.097811	
75	LDDQMTL	2.184188	
76	DDQMTLL	2.184188	
77	DQMTLLQ	2.304151	
78	QMTLLQY	2.676438	
79	MTLLQYS	2.641499	
80	TLLQYSW	2.840943	*
81	LLQYSWM	2.708776	*
82	LQYSWMS	2.510636	
83	QYSWMSL	2.510636	
84	YSWMSLM	2.375193	
85	SWMSLMA	2.160667	
86	WMSLMAF	2.388510	

POSITION	HEPTAMER	SCORE	
87	MSLMAFA	2.362308	
88	SLMAFAL	2.660953	
89	LMAFALG	2.656786	
90	MAFALGW	2.557586	
91	AFALGWR	2.908399	*
92	FALGWRS	2.835662	*
93	ALGWRSY	2.895081	*
94	LGWRSYR	3.072652	*
95	GWRSYRQ	2.909451	*
96	WRSYRQS	2.913618	*
97	RSYRQSS	2.814679	*
98	SYRQSSA	2.637107	
99	YRQSSAN	2.570830	
100	RQSSANL	2.481707	
101	QSSANLL	2.429539	
102	SSANLLC	2.355836	
103	SANLLCF	2.583680	
104	ANLLCFA	2.656417	
105	NLLCFAP	2.524092	
106	LLCFAPD	2.505345	
107	LCFAPDL	2.505345	
108	CFAPDLI	2.464894	
109	FAPDLII	2.661347	
110	APDLIIN	2.367226	
111	PDLIINE	2.190649	
112	DLIINEQ	2.285175	
113	LIINEQR	2.620508	
114	IINEQRM	2.321863	
115	INEQRMT	2.195837	
116	NEQRMTL	2.236288	
117	EQRMTLP	2.242977	
118	QRMTLPC	2.308053	
119	RMTLPCM	2.172610	
120	MTLPCMY	2.209565	
121	TLPCMYD	2.225045	
122	LPCMYDQ	2.228322	
123	PCMYDQC	1.991418	
124	CMYDQCK	2.265490	
125	MYDQCKH	2.416016	
126	YDQCKHM	2.416016	
127	DQCKHML	2.326893	
128	QCKHMLY	2.699180	
129	CKHMLYV	2.831336	*
130	KHMLYVS	2.870100	*
131	HMLYVSS	2.655615	
132	MLYVSSE	2.440012	
133	LYVSSEL	2.738657	*
134	YVSSELH	2.652279	
135	VSSELHR	2.615325	

POSITION	HEPTAMER	SCORE	
136	SSELHRL	2.646370	
137	SELHRLQ	2.681309	
138	ELHRLQV	2.848403	*
139	LHRLQVS	2.952244	*
140	HRLQVSY	3.041367	*
141	RLQVSYE	2.825764	*
142	LQVSYEE	2.471615	
143	QVSYEEY	2.560738	
144	VSYEEYL	2.723940	*
145	SYEEYLC	2.518081	
146	YEEYLCM	2.417576	
147	EEYLCMK	2.344798	
148	EYLCMKT	2.480301	
149	YLCMCTL	2.782282	*
150	LCMKTLL	2.693159	
151	CMKTLLL	2.693159	
152	MKTLLLL	2.930062	*
153	KTLLLLS	3.030567	*
154	TLLLLSS	2.816082	*
155	LLLLSSV	2.951514	*
156	LLLLSVP	2.693787	
157	LLSSVPK	2.710132	*
158	LSSVPKD	2.426968	
159	SSVPKDG	2.224662	
160	SVPKDGL	2.422802	
161	VPKDGLK	2.637286	
162	PKDGLKS	2.470192	
163	KDGLKSQ	2.564718	
164	DGLKSQE	2.246393	
165	GLKSQEL	2.529556	
166	LKSQELF	2.761567	*
167	KSQELFD	2.478403	
168	SQELFDE	2.160078	
169	QELFDEI	2.317767	
170	ELFDEIR	2.533136	
171	LFDEIRM	2.536472	
172	FDEIRMT	2.369995	
173	DEIRMTY	2.429414	
174	EIRMTYI	2.672127	
175	IRMTYIK	2.990452	*
176	RMTYIKE	2.728922	*
177	MTYIKEL	2.676754	
178	TYIKELG	2.773092	*
179	YIKELGK	2.955915	*
180	IKELGKA	2.741389	*
181	KELGKAI	2.741389	*
182	ELGKAIV	2.693998	
183	LGKAIVK	3.012324	*
184	GKAIVKR	3.064492	*

POSITION	HEPTAMER	SCORE	
185	KAIVKRE	2.964818	*
186	AIVKREG	2.746167	*
187	IVKREGN	2.607153	
188	VKREGNS	2.449464	
189	KREGNSS	2.282370	
190	REGNSSQ	2.102823	
191	EGNSSQN	1.786238	
192	GNSSQNW	1.989018	
193	NSSQNWQ	2.028123	
194	SSQNWQR	2.344708	
195	SQNWQRF	2.572552	
196	QNWQRFY	2.859815	*
197	NWQRFYQ	2.859815	*
198	WQRFYQL	3.124232	*
199	QRFYQLT	3.056955	*
200	RFYQLTK	3.236501	*
201	FYQLTKL	3.184333	*
202	YQLTKLL	3.154629	*
203	QLTKLLD	2.782342	*
204	LTKLLDS	2.747403	*
205	TKLLDSM	2.448758	
206	KLLDSMH	2.528858	
207	LLDSMHE	2.210533	
208	LDSMHEV	2.179488	
209	DSMHEVV	2.148443	
210	SMHEVVE	2.129626	
211	MHEVVEN	2.063349	
212	HEVVENL	2.361994	
213	EVVENLL	2.448372	
214	VVENLLN	2.485935	
215	VENLLNY	2.606103	
216	ENLLNYC	2.400245	
217	NLLNYCF	2.731929	*
218	LLNYCFQ	2.833145	*
219	LNYPFQT	2.666667	
220	NYCFQTF	2.696372	
221	YCFQTFL	2.960788	*
222	CFQTFLD	2.588501	
223	FQTFLDK	2.841750	*
224	QTFLDKT	2.645568	
225	TFLDKTM	2.510125	
226	FLDKTMS	2.478463	
227	LDKTMSI	2.408308	
228	DKTMSIE	2.106327	
229	KTMSIEF	2.419195	
230	TMSIEFP	2.145123	
231	MSIEFPE	2.009620	
232	SIEFPEM	2.009620	
233	IEFPEML	2.207760	

POSITION	HEPTAMER	SCORE	
234	EFPEMLA	2.122808	
235	FPEMLAE	2.122808	
236	PEMLAEI	2.052653	
237	EMLAEII	2.269929	
238	MLAEIIT	2.405432	
239	LAEIITN	2.439660	
240	AEIITNQ	2.276459	
241	EIITNQI	2.361411	
242	IITNQIP	2.405664	
243	ITNQIPK	2.462460	
244	TNQIPKY	2.592034	
245	NQIPKYS	2.560371	
246	QIPKYSN	2.560371	
247	IPKYSNG	2.521266	
248	PKYSNGN	2.297300	
249	KYSNGNI	2.514577	
250	YSNGNIK	2.514577	
251	SNGNIKK	2.441799	
252	NGNIKKL	2.639939	
253	GNIKKLL	2.904355	*
254	NIKKLLF	3.136366	*
255	IKKLLFH	3.314405	*
256	KKLLFHQ	3.191654	*
257	KLLFHQK	3.191654	*

-ql flag: (qualitative, default: 0.2) Sequence: 1M2Z\_1—Chains A, C[auth D]—glucocorticoid receptor—Homo sapiens (9606)

POSITION	HEPTAMER	SCORE	
0	FQSPGSI	0.064203	
1	QSPGSIV	0.030571	
2	SPGSIVP	0.029481	
3	PGSIVPA	0.050952	
4	GSIVPAT	0.037988	
5	SIVPATL	0.201908	*
6	IVPATLP	0.218834	*
7	VPATLPQ	0.110975	
8	PATLPQL	0.136777	
9	ATLPQLT	0.123813	
10	TLPQLTP	0.119269	
11	LPQLTPT	0.119269	
12	PQLTPTL	0.119269	
13	QLTPTLV	0.196821	
14	LTPTLVS	0.178805	
15	TPTLVSL	0.178805	
16	PTLVSL	0.295123	*

POSITION	HEPTAMER	SCORE	
17	TLVSLLE	0.226769	*
18	LVSLELV	0.317284	*
19	VSLELVI	0.322879	*
20	SLLEVIE	0.176975	
21	LLEVIEP	0.193901	
22	LEVIEPE	0.022194	
23	EVIEPEV	-0.003609	
24	VIEPEVL	0.168099	
25	IEPEVLY	0.266376	*
26	EPEVLYA	0.161971	
27	PEVLYAG	0.169759	
28	EVLVYAGY	0.345587	*
29	VLYAGYD	0.335563	*
30	LYAGYDS	0.241085	*
31	YAGYDSS	0.120805	
32	AGYDSSV	0.022527	
33	GYDSSVP	0.017983	
34	YDSSVPD	0.000171	
35	DSSVPDS	-0.192584	
36	SSVPDST	-0.127170	
37	SVPDSTW	0.090103	
38	VPDSTWR	0.231280	*
39	PDSTWRI	0.262677	*
40	DSTWRIM	0.225254	*
41	STWRIMT	0.290668	*
42	TWRIMTT	0.294631	*
43	WRIMTTL	0.410949	*
44	RIMTTLN	0.193052	
45	IMTTLNM	0.031379	
46	MTTLNML	0.025784	
47	TTLNMLG	0.002642	
48	TLNMLGG	-0.044960	
49	LNMLGGR	0.092254	
50	NMLGGRQ	-0.010011	
51	MLGGRQV	0.085091	
52	LGGRQVI	0.231462	*
53	GGRQVIA	0.132652	
54	GRQVIAA	0.197762	
55	RQVIAAV	0.335879	*
56	QVIAAVK	0.285999	*
57	VIAAVKW	0.485256	*
58	IAAVKWA	0.412249	*
59	AAVKWAK	0.377670	*
60	AVKWAKA	0.377670	*
61	VKWAKAI	0.482075	*
62	KWAKAIP	0.404523	*
63	WAKAIPG	0.269588*	
64	AKAIPGF	0.180424	
65	KAIPGFR	0.300130	*

POSITION	HEPTAMER	SCORE	
66	AIPGFRN	0.208210	*
67	IPGFRNL	0.307020	*
68	PGFRNLH	0.234175	*
69	GFRNLHL	0.337529	*
70	FRNLHLD	0.319717	*
71	RNLHLDD	0.130156	
72	NLHLDDQ	0.006996	
73	LHLDDQM	-0.012877	
74	HLDDQMT	-0.129195	
75	LDDQMTL	-0.061945	
76	DDQMTLL	-0.061945	
77	DQMTLLQ	0.017522	
78	QMTLLQY	0.271729	*
79	MTLQYS	0.253713	*
80	TLLQYSW	0.491482	*
81	LLQYSWM	0.467023	*
82	LQYSWMS	0.346743	*
83	QYSWMSL	0.346743	*
84	YSWMSLM	0.30823	*
85	SWMSLMA	0.136946	
86	WMSLMAF	0.265056	*
87	MSLMAFA	0.069253	
88	SLMAFAL	0.210030	*
89	LMAFALG	0.166391	
90	MAFALGW	0.263384	*
91	AFALGWR	0.425056	*
92	FALGWRS	0.403585	*
93	ALGWRSY	0.468231	*
94	LGWRSYR	0.587937	*
95	GWRSYRQ	0.485673	*
96	WRSYRQS	0.529311	*
97	RSYRQSS	0.312038	*
98	SYRQSSA	0.192332	
99	YRQSSAN	0.191709	
100	RQSSANL	0.119234	
101	QSSANLL	0.098338	
102	SSANLLC	-0.031076	
103	SANLLCF	0.097034	
104	ANLLCFA	0.118504	
105	NLLCFAP	0.113960	
106	LLCFAPD	0.053133	
107	LCFAPDL	0.053133	
108	CFAPDLI	0.058728	
109	FAPDLII	0.296001	*
110	APDLIIN	0.167268	
111	PDLIINE	0.094370	
112	DLIINEQ	0.095460	
113	LIINEQR	0.298087	*
114	IINEQRM	0.157311	

POSITION	HEPTAMER	SCORE	
115	INEQRMT	0.035398	
116	NEQRMTL	0.029803	
117	EQRMTLP	0.047353	
118	QRMTLPC	-0.012618	
119	RMTLPCM	-0.051130	
120	MTLPCMY	0.000449	
121	TLPCMYD	-0.040506	
122	LPCMYDQ	-0.026453	
123	PCMYDQC	-0.258131	
124	CMYDQCK	-0.183761	
125	MYDQCKH	-0.019333	
126	YDQCKHM	-0.019333	
127	DQCKHML	-0.091807	
128	QCKHMLY	0.162399	
129	CKHMLYV	0.238861	*
130	KHMLYVS	0.350259	*
131	HMLYVSS	0.258962	*
132	MLYVSSE	0.154505	
133	LYVSSEL	0.295281	*
134	YVSSELH	0.228031	*
135	VSSELHR	0.176453	
136	SSELHRL	0.202255	*
137	SELHRLQ	0.220272	*
138	ELHRLQV	0.314749	*
139	LHRLQVS	0.366176	*
140	HRLQVS	0.438651	*
141	RLQVSYE	0.334193	*
142	LQVSYEE	0.141590	
143	QVSYEEY	0.214065	*
144	VSYEEYL	0.316329	*
145	SYEEYLC	0.110453	
146	YEEYLCM	0.089957	
147	EEYLCMK	-0.011501	
148	EYLCMKT	0.043888	
149	YLCMCTL	0.215595	*
150	LCMKTLL	0.143121	
151	CMKTLLL	0.143121	
152	MKTLLLL	0.374799	*
153	KTLLLLS	0.395295	*
154	TLLLLSS	0.303999	*
155	LLLLSSV	0.394514	*
156	LLLLSSVP	0.291160	*
157	LLSSVPK	0.262176	*
158	LSSVPKD	0.080444	
159	SSVPKDG	-0.083475	
160	SVPKDGL	0.036805	
161	VPKDGLK	0.128102	
162	PKDGLKS	0.033624	
163	KDGLKSQ	0.034713	

POSITION	HEPTAMER	SCORE	
164	DGLKSQE	-0.108010	
165	GLKSQEL	0.073722	
166	LKSQELF	0.245470	*
167	KSQELFD	0.063738	
168	SQELFDE	-0.078985	
169	QELFDEI	0.046891	
170	ELFDEIR	0.170051	
171	LFDEIRM	0.200982	*
172	FDEIRMT	0.084664	
173	DEIRMTY	0.149310	
174	EIRMTYI	0.336636	*
175	IRMTYIK	0.479359	*
176	RMTYIKE	0.302057	*
177	MTYIKEL	0.281161	*
178	TYIKELG	0.258018	*
179	YIKELGK	0.345352	*
180	IKELGKA	0.174068	
181	KELGKAI	0.174068	
182	ELGKAIV	0.177249	
183	LGKAIVK	0.319972	*
184	GKAIVKR	0.340868	*
185	KAIVKRE	0.333080	*
186	AIVKREG	0.198145	
187	IVKREGN	0.176051	
188	VKREGNS	0.050175	
189	KREGNSS	-0.044302	
190	REGNSSQ	-0.117583	
191	EGNSSQN	-0.259383	
192	GNSSQNW	0.009317	
193	NSSQNWQ	0.070972	
194	SSQNWQR	0.212772	*
195	SQNWQRF	0.340882	*
196	QNWQRFY	0.533637	*
197	NWQRFYQ	0.533637	*
198	WQRFYQL	0.654541	*
199	QRFYQLT	0.441230	*
200	RFYQLTK	0.514511	*
201	FYQLTKL	0.493615	*
202	YQLTKLL	0.485786	*
203	QLTKLLD	0.231579	*
204	LTKLLDS	0.213563	*
205	TKLLDSM	0.072787	
206	KLLDSMH	0.121855	
207	LLDSMHE	-0.020868	
208	LDSMHEV	-0.046671	
209	DSMHEVV	-0.072474	
210	SMHEVVE	-0.062449	
211	MHEVVEN	-0.063073	
212	HEVVENL	0.077704	

POSITION	HEPTAMER	SCORE	
213	EVVENLL	0.144953	
214	VVENLLN	0.195757	
215	VENLLNY	0.294034	*
216	ENLLNYC	0.088158	
217	NLLNYCF	0.267695	*
218	LLNYCFQ	0.286334	*
219	LNYPFQT	0.170016	
220	NYCFQTF	0.177845	
221	YCFQTFI	0.298749	*
222	CFQTFID	0.044543	
223	FQTFIDK	0.247237	*
224	QTFIDKI	0.123090	
225	TFDIKTI	0.084578	
226	FDIKTMI	0.080616	
227	LDIKTMI	0.078382	
228	DIKTMSI	-0.093326	
229	KTMSIEF	0.096235	
230	TMSIEFP	0.021865	
231	MSIEFPE	-0.033524	
232	SIEFPEM	-0.033524	
233	IEFPEML	0.086757	
234	EFPEMLA	-0.017648	
235	FPEMLAE	-0.017648	
236	PEMLAEI	-0.019882	
237	EMLAEII	0.089066	
238	MLAEIIT	0.144456	
239	LAEIITN	0.164328	
240	AEIITNQ	0.062064	
241	EIITNQI	0.166468	
242	IITNQIP	0.234822	*
243	ITNQIPK	0.200243	*
244	TNQIPKY	0.267123	*
245	NQIPKYS	0.263160	*
246	QIPKYSN	0.263160	*
247	IPKYSNG	0.201505	*
248	PKYSNGN	0.075006	
249	KYSNGNI	0.183955	
250	YSNGNIK	0.183955	
251	SNGNIKK	0.082496	
252	NGNIKKL	0.202777	*
253	GNIKKLL	0.323681	*
254	NIKKLLF	0.495429	*
255	IKKLLFH	0.549084	*
256	KKLLFHQ	0.441224	*
257	KLLFHQK	0.441224	*

# Bibliography

## References

- [1] Emily R Weikum, Matthew T Knuesel, Eric A Ortlund, and Keith R Yamamoto. Glucocorticoid receptor control of transcription: precision and plasticity via allostery. *Nature reviews Molecular cell biology*, 18(3):159, 2017. (Cited on page 1.)
- [2] George P Chrousos and Tomoshige Kino. Glucocorticoid signaling in the cell: expanding clinical implications to complex human behavioral and somatic disorders. *Annals of the New York Academy of Sciences*, 1179:153, 2009. (Cited on page 1.)
- [3] Vincent Giguère, Stanley M Hollenberg, Michael G Rosenfeld, and Ronald M Evans. Functional domains of the human glucocorticoid receptor. *Cell*, 46(5):645–652, 1986. (Cited on pages 1 and 4.)
- [4] François Tronche, Christoph Kellendonk, Holger M Reichardt, and Günther Schütz. Genetic dissection of glucocorticoid receptor function in mice. *Current opinion in genetics & development*, 8(5):532–538, 1998. (Cited on pages 1 and 5.)
- [5] Ulrike Baschant and Jan Tuckermann. The role of the glucocorticoid receptor in inflammation and immunity. *The Journal of steroid biochemistry and molecular biology*, 120(2-3):69–75, 2010. (Cited on page 1.)
- [6] Robert M Sapolsky, L Michael Romero, and Allan U Munck. How do glucocorticoids influence stress responses? integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine reviews*, 21(1):55–89, 2000. (Cited on page 1.)
- [7] Brian M Necela and John A Cidlowski. Mechanisms of glucocorticoid receptor action in noninflammatory and inflammatory cells. *Proceedings of the American Thoracic Society*, 1(3):239–246, 2004. (Cited on page 1.)
- [8] Karen Jui Lin Choo, F Estelle R Simons, and Aziz Sheikh. Glucocorticoids for the treatment of anaphylaxis. *Evidence-Based Child Health: A Cochrane Review Journal*, 8(4):1276–1294, 2013. (Cited on page 1.)
- [9] Sabine Chourbaji and Peter Gass. Glucocorticoid receptor transgenic mice as models for depression. *Brain research reviews*, 57(2):554–560, 2008. (Cited on page 1.)
- [10] Roger Murayi and Prashant Chittiboina. Glucocorticoids in the management of peritumoral brain edema: a review of molecular mechanisms. *Child's Nervous System*, 32(12):2293–2302, 2016. (Cited on page 1.)
- [11] Peer B Jacobson, Thomas W von Geldern, Lars Oehman, Marie Oesterland, Jiahong Wang, Bradley Zinker, Denise Wilcox, Phong T Nguyen, Amanda Mika, and Steven Fung. Hepatic glucocorticoid receptor antagonism is sufficient to reduce elevated hepatic glucose output and improve glucose control in

- animal models of type 2 diabetes. *Journal of Pharmacology and Experimental Therapeutics*, 314(1):191–200, 2005. (Cited on page 1.)
- [12] Ana R Sousa, Stephen J Lane, John A Cidlowski, Dontcho Z Staynov, and Tak H Lee. Glucocorticoid resistance in asthma is associated with elevated in vivo expression of the glucocorticoid receptor  $\beta$ -isoform. *Journal of Allergy and Clinical Immunology*, 105(5):943–950, 2000. (Cited on page 1.)
- [13] K Renner, MJ Ausserlechner, and R Kofler. A conceptual view on glucocorticoid-induced apoptosis, cell cycle arrest and glucocorticoid resistance in lymphoblastic leukemia. *Current molecular medicine*, 3(8):707–717, 2003. (Cited on page 1.)
- [14] Ernesto Canalis, Gherardo Mazziotti, Andrea Giustina, and John P Bilezikian. Glucocorticoid-induced osteoporosis: pathophysiology and therapy. *Osteoporosis International*, 18(10):1319–1328, 2007. (Cited on page 1.)
- [15] Donald E Hricik, Wassim Y Almawi, and Terry B Strom. Trends in the use of glucocorticoids in renal transplantation. *Transplantation*, 57(7):979–989, 1994. (Cited on page 1.)
- [16] Anna Kleiman, Sabine Hübner, Jan M Rodriguez Parkitna, Anita Neumann, Stefan Hofer, Markus A Weigand, Michael Bauer, Wolfgang Schmid, Gunter Schütz, and Claude Libert. Glucocorticoid receptor dimerization is required for survival in septic shock via suppression of interleukin-1 in macrophages. *The FASEB Journal*, 26(2):722–729, 2012. (Cited on page 1.)
- [17] Machiko Ikegami, Alan H Jobe, John Newnham, Daniel H Polk, Karen E Willet, and Peter Sly. Repetitive prenatal glucocorticoids improve lung function and decrease growth in preterm lambs. *American journal of respiratory and critical care medicine*, 156(1):178–184, 1997. (Cited on page 1.)
- [18] Peter J Barnes. Therapeutic strategies for allergic diseases. *Nature*, 402(6760):31–38, 1999. (Cited on page 1.)
- [19] RF Laan, TL Jansen, and PL Van Riel. Glucocorticosteroids in the management of rheumatoid arthritis. *Rheumatology (Oxford, England)*, 38(1):6–12, 1999. (Cited on page 1.)
- [20] Harold L Mason, Willard M Hoehn, and Edward C Kendall. Chemical studies of the suprarenal cortex: Iv. structures of compounds c, d, e, f, and g. *Journal of Biological Chemistry*, 124(2):459–474, 1938. (Cited on page 1.)
- [21] T von Reichstein and CW Shoppee. *The hormones of the adrenal cortex*, volume 1, pages 345–413. Elsevier, 1943. (Cited on page 1.)
- [22] Marjet D Heitzer, Irene M Wolf, Edwin R Sanchez, Selma F Witchel, and Donald B DeFranco. Glucocorticoid receptor physiology. *Reviews in endocrine and metabolic disorders*, 8(4):321–330, 2007. (Cited on page 1.)
- [23] Lewis H Sarett and Everett S Wallis. The chemistry of the steroids. *Annual review of biochemistry*, 16(1):655–688, 1947. (Cited on page 1.)
- [24] Philip S Hench, Edward C Kendall, Charles H Slocumb, and Howard F Polley. The effect of a hormone of the adrenal cortex (17-hydroxy-11-dehydrocorticosterone: compound e) and of pituitary adrenocortical hormone

- in arthritis: preliminary report. *Annals of the rheumatic diseases*, 8(2):97–104, 1949. (Cited on page 1.)
- [25] FD Hart. Corticosteroid therapy in the rheumatic disorders. 1980. (Cited on page 1.)
- [26] Christopher M Burns. The history of cortisone discovery and development. *Rheumatic Disease Clinics*, 42(1):1–14, 2016. (Cited on page 1.)
- [27] Evangelia Charmandari, Tomoshige Kino, and George P Chrousos. Glucocorticoids and their actions: an introduction. *Annals of the New York Academy of Sciences*, 1024(1):1–8, 2004. (Cited on page 1.)
- [28] Cinzia Solinas, Laura Perra, Marco Aiello, Edoardo Migliori, and Nicola Petrosillo. A critical evaluation of glucocorticoids in the treatment of severe covid-19. *Cytokine & growth factor reviews*, 2020. (Cited on page 1.)
- [29] Chun-Yang Fan, Soojin Lee, and Douglas M Cyr. Mechanisms for regulation of hsp70 function by hsp40. *Cell stress & chaperones*, 8(4):309, 2003. (Cited on pages 1 and 90.)
- [30] Heike Schäcke, Wolf-Dietrich Doecke, and Khusru Asadullah. Mechanisms involved in the side effects of glucocorticoids. *Pharmacology & therapeutics*, 96(1):23–43, 2002. (Cited on page 1.)
- [31] Merih Oray, Khawla Abu Samra, Nazanin Ebrahimiadib, Halea Meese, and C Stephen Foster. Long-term side effects of glucocorticoids. *Expert opinion on drug safety*, 15(4):457–465, 2016. (Cited on page 1.)
- [32] Tobias Seitz, Ralf Thoma, Guillaume A Schoch, Martine Stihle, Joerg Benz, Brigitte D’Arcy, Andrea Wiget, Armin Ruf, Michael Hennig, and Reinhard Sterner. Enhancing the stability and solubility of the glucocorticoid receptor ligand-binding domain by high-throughput library screening. *Journal of molecular biology*, 403(4):562–577, 2010. (Cited on page 1.)
- [33] Oliver Robin Lorenz, Lee Freiburger, Daniel Andreas Rutz, Maike Krause, Bettina Karolina Zierer, Sara Alvira, Jorge Cuéllar, José María Valpuesta, Tobias Madl, and Michael Sattler. Modulation of the hsp90 chaperone cycle by a stringent client protein. *Molecular cell*, 53(6):941–953, 2014. (Cited on pages 1, 2, and 4.)
- [34] WB Pratt, Y Morishima, M Murphy, and M Harrell. Chaperoning of glucocorticoid receptors. *Molecular chaperones in health and disease*, pages 111–138, 2006. (Cited on pages 1 and 4.)
- [35] Iwona Grad and Didier Picard. The glucocorticoid responses are shaped by molecular chaperones. *Molecular and cellular endocrinology*, 275(1-2):2–12, 2007. (Cited on page 1.)
- [36] Patricija van Oosten-Hawle, Yael Bar-Lavan, Netta Shemesh, and Anat Ben-Zvi. Chaperone families and interactions in metazoa. *Essays in biochemistry*, 60(2):237–253, 2016. (Cited on page 1.)
- [37] Ricksen S Winardhi, Qingnan Tang, Huijuan You, Michael Sheetz, and Jie Yan. The holdase function of escherichia coli hsp70 (dnak) chaperone. *bioRxiv*, page 305854, 2018. (Cited on pages 1 and 82.)

- [38] Anne S Wentink, Nadinath B Nillegoda, Jennifer Feufel, Gabrièle Ubartaitė, Carolyn P Schneider, Paolo De Los Rios, Janosch Hennig, Alessandro Barducci, and Bernd Bukau. Molecular dissection of amyloid disaggregation by human hsp70. *Nature*, 587(7834):483–488, 2020. (Cited on pages 1, 67, 82, and 97.)
- [39] Elizabeth A Craig. Hsp70 at the membrane: driving protein translocation. *BMC biology*, 16(1):1–11, 2018. (Cited on page 1.)
- [40] Taras Makhnevych and Walid A Houry. The role of hsp90 in protein complex assembly. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1823(3):674–682, 2012. (Cited on page 2.)
- [41] Bernd Bukau, Jonathan Weissman, and Arthur Horwich. Molecular chaperones and protein quality control. *Cell*, 125(3):443–451, 2006. (Cited on page 2.)
- [42] Andrija Finka and Pierre Goloubinoff. Proteomic data from human cell cultures refine mechanisms of chaperone-mediated protein homeostasis. *Cell Stress and Chaperones*, 18(5):591–605, 2013. (Cited on page 2.)
- [43] Elaine Kirschke, Devrishi Goswami, Daniel Southworth, Patrick R Griffin, and David A Agard. Glucocorticoid receptor function regulated by coordinated action of the hsp90 and hsp70 chaperone cycles. *Cell*, 157(7):1685–1697, 2014. (Cited on pages 2, 4, 6, 7, 8, 72, 82, 83, 90, and 93.)
- [44] Thomas Suren, Daniel Rutz, Patrick Moessmer, Ulrich Merkel, Johannes Buchner, and Matthias Rief. Single-molecule force spectroscopy reveals folding steps associated with hormone binding and activation of the glucocorticoid receptor. *Proceedings of the National Academy of Sciences*, 115(46):11688–11693, 2018. (Cited on pages 2, 35, 37, 38, 39, 47, 49, 50, 64, 73, 74, 79, 80, 90, 91, 92, and 98.)
- [45] Thomas E Creighton. *Proteins: structures and molecular properties*. Macmillan, 1993. (Cited on page 3.)
- [46] Michael Levitt. Nature of the protein universe. *Proceedings of the National Academy of Sciences*, 106(27):11079–11084, 2009. (Cited on page 3.)
- [47] GJ Mulder. Ueber die zusammensetzung einiger thierischen substanzen. *Journal für praktische Chemie*, 16(1):129–152, 1839. (Cited on page 3.)
- [48] Keith Roberts, Bruce Alberts, Alexander Johnson, Peter Walter, and Tim Hunt. Molecular biology of the cell. *New York: Garland Science*, 2002. (Cited on page 3.)
- [49] Charles Tanford and Jacqueline Reynolds. *Nature’s robots: a history of proteins*. OUP Oxford, 2003. (Cited on page 3.)
- [50] Jeffrey L Hansen, T Martin Schmeing, Peter B Moore, and Thomas A Steitz. Structural insights into peptide bond formation. *Proceedings of the National Academy of Sciences*, 99(18):11670–11675, 2002. (Cited on page 3.)
- [51] Christian B Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973. (Cited on page 3.)

- [52] Cyrus Levinthal. Are there pathways for protein folding? *Journal de chimie physique*, 65:44–45, 1968. (Cited on page 3.)
- [53] José Nelson Onuchic and Peter G Wolynes. Theory of protein folding. *Current opinion in structural biology*, 14(1):70–75, 2004. (Cited on page 4.)
- [54] David Whitley, Steven P Goldberg, and William D Jordan. Heat shock proteins: a review of the molecular chaperones. *Journal of vascular surgery*, 29(4):748–751, 1999. (Cited on page 4.)
- [55] Kathryn J Howard, SJ Holley, Keith R Yamamoto, and Clark W Distelhorst. Mapping the hsp90 binding region of the glucocorticoid receptor. *Journal of Biological Chemistry*, 265(20):11928–11935, 1990. (Cited on page 4.)
- [56] Randy K Bledsoe, Valerie G Montana, Thomas B Stanley, Chris J Delves, Christopher J Apolito, David D McKee, Thomas G Consler, Derek J Parks, Eugene L Stewart, and Timothy M Willson. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell*, 110(1):93–105, 2002. (Cited on pages 4 and 5.)
- [57] D Ricketson, U Hostick, L Fang, KR Yamamoto, and BD Darimont. A conformational switch in the ligand-binding domain regulates the dependence of the glucocorticoid receptor on hsp90. *Journal of molecular biology*, 368(3):729–741, 2007. (Cited on page 4.)
- [58] F602s gr-lbd pdb. <https://www.rcsb.org/structure/1M2Z>, 2021. [Online; accessed 20-January-2022]. (Cited on pages 4 and 5.)
- [59] Mahita Kadmiel and John A Cidlowski. Glucocorticoid receptor signaling in health and disease. *Trends in pharmacological sciences*, 34(9):518–530, 2013. (Cited on pages 5 and 6.)
- [60] WL Miller and JB Tyrrell. The adrenal cortex. chap 12. *Endocrinology and metabolism*. 3rd ed. New York (NY): McGraw-Hill, pages 659–678, 1995. (Cited on page 5.)
- [61] Simon C Biddie, Becky L Conway-Campbell, and Stafford L Lightman. Dynamic regulation of glucocorticoid signalling in health and disease. *Rheumatology*, 51(3):403–412, 2012. (Cited on page 5.)
- [62] Andrew R Clark and Maria G Belvisi. Maps and legends: the quest for dissociated ligands of the glucocorticoid receptor. *Pharmacology & therapeutics*, 134(1):54–67, 2012. (Cited on page 5.)
- [63] Miguel Beato, Peter Herrlich, and Günther Schütz. Steroid hormone receptors: many actors in search of a plot. *Cell*, 83(6):851–857, 1995. (Cited on page 6.)
- [64] Jacques Drouin, YL Sun, M Chamberland, Y Gauthier, A De Lean, M Nemer, and TJ Schmidt. Novel glucocorticoid receptor complex with dna element of the hormone-repressed pomc gene. *The EMBO journal*, 12(1):145–156, 1993. (Cited on page 6.)
- [65] Carsten Jonat, Hans J Rahmsdorf, Kun-Koo Park, Andrew CB Cato, Stephan Gebel, Helmut Ponta, and Peter Herrlich. Antitumor promotion and anti-

- inflammation: down-modulation of ap-1 (fos/jun) activity by glucocorticoid hormone. *Cell*, 62(6):1189–1204, 1990. (Cited on page 6.)
- [66] Sofie Vandevyver, Lien Dejager, and Claude Libert. On the trail of the glucocorticoid receptor: into the nucleus and back. *Traffic*, 13(3):364–374, 2012. (Cited on pages 6 and 7.)
- [67] Joselyn Rojas, Mervin Chávez-Castillo, Mayela Cabrera, and Valmore Bermúdez. Glucocorticoid-induced death of pancreatic beta cells: an organized chaos. *JOP. Journal of the Pancreas*, 16(1):11–19, 2015. (Cited on page 6.)
- [68] Ray Yu-Ruei Wang, Chari M Noddings, Elaine Kirschke, Alexander G Myasnikov, Jill L Johnson, and David A Agard. Gr chaperone cycle mechanism revealed by cryo-em: inactivation of gr by gr: Hsp90: Hsp70: Hop client-loading complex. *bioRxiv*, 2020. (Cited on pages 6, 83, and 90.)
- [69] Ferruccio Ritossa. A new puffing pattern induced by temperature shock and dnp in drosophila. *Experientia*, 18(12):571–573, 1962. (Cited on page 6.)
- [70] Milton J Schlesinger. Heat shock proteins. *Journal of Biological Chemistry*, 265(21):12111–12114, 1990. (Cited on page 7.)
- [71] Milton J Schlesinger. Heat shock proteins: the search for functions. *The Journal of cell biology*, 103(2):321–325, 1986. (Cited on page 7.)
- [72] William B Pratt and David O Toft. Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Experimental biology and medicine*, 228(2):111–133, 2003. (Cited on page 7.)
- [73] Iain J McEwan, William B Pratt, Mario D Galigniana, Yoshihiro Morishima, and Patrick JM Murphy. Role of molecular chaperones in steroid receptor action. *Essays in biochemistry*, 40:41–58, 2004. (Cited on page 7.)
- [74] Kurt D Dittmar, Kevin A Hutchison, Janet K Owens-Grillo, and William B Pratt. Reconstitution of the steroid receptor· hsp90 heterocomplex assembly system of rabbit reticulocyte lysate. *Journal of Biological Chemistry*, 271(22):12833–12839, 1996. (Cited on page 7.)
- [75] Yoshihiro Morishima, Patrick JM Murphy, Da-Pei Li, Edwin R Sanchez, and William B Pratt. Stepwise assembly of a glucocorticoid receptor· hsp90 heterocomplex resolves two sequential atp-dependent events involving first hsp70 and then hsp90 in opening of the steroid binding pocket. *Journal of Biological Chemistry*, 275(24):18054–18060, 2000. (Cited on page 7.)
- [76] David F Smith, BA Stensgard, WJ Welch, and DO Toft. Assembly of progesterone receptor with heat shock proteins and receptor activation are atp mediated events. *Journal of Biological Chemistry*, 267(2):1350–1356, 1992. (Cited on page 7.)
- [77] MP Mayer and Bernd Bukau. Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and molecular life sciences*, 62(6):670–684, 2005. (Cited on pages 8 and 9.)

- [78] Matthias P Mayer. Hsp70 chaperone dynamics and molecular mechanism. *Trends in biochemical sciences*, 38(10):507–514, 2013. (Cited on pages 8, 10, and 87.)
- [79] Jürgen Radons. The human hsp70 family of chaperones: where do we stand? *Cell Stress and Chaperones*, 21(3):379–404, 2016. (Cited on page 9.)
- [80] Deepak Sharma and Daniel C Masison. Hsp70 structure, function, regulation and influence on yeast prions. *Protein and peptide letters*, 16(6):571–581, 2009. (Cited on pages 9 and 82.)
- [81] Rina Rosenzweig, Nadinath B Nillegoda, Matthias P Mayer, and Bernd Bukau. The hsp70 chaperone network. *Nature Reviews Molecular Cell Biology*, 20(11):665–680, 2019. (Cited on pages 9 and 10.)
- [82] Henry Pelham, M Lewis, and Susan Lee Lindquist. Expression of a drosophila heat shock protein in mammalian cells: transient association with nucleoli after heat shock. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 307(1132):301–307, 1984. (Cited on page 9.)
- [83] María Rosario Fernández-Fernández and José María Valpuesta. Hsp70 chaperone: a master player in protein homeostasis. *F1000Research*, 7, 2018. (Cited on page 9.)
- [84] Matthias P Mayer and Lila M Gierasch. Recent advances in the structural and mechanistic aspects of hsp70 molecular chaperones. *Journal of Biological Chemistry*, 294(6):2085–2097, 2019. (Cited on pages 9, 10, and 90.)
- [85] Paolo De Los Rios and Alessandro Barducci. Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption. *Elife*, 3:e02218, 2014. (Cited on page 10.)
- [86] Stefan Rüdiger, Lothar Germeroth, Jens Schneider-Mergener, and Bernd Bukau. Substrate specificity of the dnaK chaperone determined by screening cellulose-bound peptide libraries. *The EMBO journal*, 16(7):1501–1507, 1997. (Cited on page 11.)
- [87] Xiaotian Zhu, Xun Zhao, William F Burkholder, Alexander Gragerov, Craig M Ogata, Max E Gottesman, and Wayne A Hendrickson. Structural analysis of substrate binding by the molecular chaperone dnaK. *Science*, 272(5268):1606–1614, 1996. (Cited on page 11.)
- [88] Benjamin Misselwitz, Oliver Staack, and Tom A Rapoport. J proteins catalytically activate hsp70 molecules to trap a wide range of peptide sequences. *Molecular cell*, 2(5):593–603, 1998. (Cited on pages 11 and 90.)
- [89] Arthur Ashkin. Acceleration and trapping of particles by radiation pressure. *Physical review letters*, 24(4):156, 1970. (Cited on page 14.)
- [90] Arthur Ashkin and JM Dziedzic. Optical levitation by radiation pressure. *Applied Physics Letters*, 19(8):283–285, 1971. (Cited on page 14.)
- [91] Arthur Ashkin. Trapping of atoms by resonance radiation pressure. *Physical Review Letters*, 40(12):729, 1978. (Cited on page 14.)

- [92] Arthur Ashkin. Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophysical journal*, 61(2):569–582, 1992. (Cited on pages 15 and 17.)
- [93] Arthur Ashkin and James M Dziedzic. Optical trapping and manipulation of viruses and bacteria. *Science*, 235(4795):1517–1520, 1987. (Cited on page 16.)
- [94] Josep M Huguet, Cristiano V Bizarro, Núria Fornis, Steven B Smith, Carlos Bustamante, and Felix Ritort. Single-molecule derivation of salt dependent base-pair free energies in dna. *Proceedings of the National Academy of Sciences*, 107(35):15431–15436, 2010. (Cited on page 16.)
- [95] Michelle D Wang, Mark J Schnitzer, Hong Yin, Robert Landick, Jeff Gelles, and Steven M Block. Force and velocity measured for single molecules of rna polymerase. *Science*, 282(5390):902–907, 1998. (Cited on page 16.)
- [96] Melanie Brunnbauer, Felix Mueller-Planitz, Süleyman Koesem, Thi Hieu Ho, Renate Dombi, J Christof M Gebhardt, Matthias Rief, and Zeynep Oekten. Regulation of a heterodimeric kinesin-2 through an unprocessive motor domain that is turned processive by its partner. *Proceedings of the National Academy of Sciences*, 107(23):10460–10465, 2010. (Cited on page 16.)
- [97] Giovanni Cappello, Paolo Pierobon, Clémentine Symonds, Lorenzo Busoni, J Christof, M Gebhardt, Matthias Rief, and Jacques Prost. Myosin v stepping mechanism. *Proceedings of the National Academy of Sciences*, 104(39):15328–15333, 2007. (Cited on page 16.)
- [98] Amit D Mehta, Matthias Rief, James A Spudich, David A Smith, and Robert M Simmons. Single-molecule biomechanics with optical methods. *Science*, 283(5408):1689–1695, 1999. (Cited on page 16.)
- [99] Ciro Cecconi, Elizabeth A Shank, Frederick W Dahlquist, Susan Marqusee, and Carlos Bustamante. Protein-dna chimeras for single molecule mechanical folding studies with the optical tweezers. *European Biophysics Journal*, 37(6):729–738, 2008. (Cited on page 16.)
- [100] Jeffrey R Moffitt, Yann R Chemla, David Izhaky, and Carlos Bustamante. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proceedings of the National Academy of Sciences*, 103(24):9006–9011, 2006. (Cited on page 16.)
- [101] Johannes Stigler, Fabian Ziegler, Anja Gieseke, J Christof M Gebhardt, and Matthias Rief. The complex folding network of single calmodulin molecules. *Science*, 334(6055):512–516, 2011. (Cited on page 16.)
- [102] Benjamin Pelz, Gabriel Žoldák, Fabian Zeller, Martin Zacharias, and Matthias Rief. Subnanometre enzyme mechanics probed by single-molecule force spectroscopy. *Nature communications*, 7(1):1–9, 2016. (Cited on page 16.)
- [103] Nobel prize for arthur ashkin. <https://www.nobelprize.org/prizes/physics/2018/ashkin/facts/>, 2021. [Online; accessed 09-September-2021]. (Cited on page 17.)

- [104] Karel Svoboda and Steven M Block. Biological applications of optical forces. *Annual review of biophysics and biomolecular structure*, 23(1):247–285, 1994. (Cited on pages 17, 19, and 20.)
- [105] Keir C Neuman and Steven M Block. Optical trapping. *Review of scientific instruments*, 75(9):2787–2809, 2004. (Cited on pages 19 and 20.)
- [106] Tsvi Thusty, Amit Meller, and Roy Bar-Ziv. Optical gradient forces of strongly localized fields. *Physical review letters*, 81(8):1738, 1998. (Cited on page 20.)
- [107] Antonio AR Neves, Adriana Fontes, Liliana de Y Pozzo, Andre A De Thomaz, Enver Chillce, Eugenio Rodriguez, Luiz C Barbosa, and Carlos L Cesar. Electromagnetic forces for an arbitrary optical trapping of a spherical dielectric. *Optics Express*, 14(26):13101–13106, 2006. (Cited on page 20.)
- [108] Benjamin Pelz. *Enzyme mechanics studied by single molecule force spectroscopy*. Macmillan, 2014. (Cited on page 20.)
- [109] Yann von Hansen, Alexander Mehlich, Benjamin Pelz, Matthias Rief, and Roland R Netz. Auto-and cross-power spectral analysis of dual trap optical tweezer experiments using bayesian inference. *Review of Scientific Instruments*, 83(9):095116, 2012. (Cited on pages 20 and 21.)
- [110] Markita P Landry, Patrick M McCall, Zhi Qi, and Yann R Chemla. Characterization of photoactivated singlet oxygen damage in single-molecule optical trap experiments. *Biophysical journal*, 97(8):2128–2136, 2009. (Cited on pages 20, 21, 24, and 56.)
- [111] João MJM Ravasco, Hélio Faustino, Alexandre Trindade, and Pedro MP Gois. Bioconjugation with maleimides: a useful tool for chemical biology. *Chemistry—A European Journal*, 25(1):43–59, 2019. (Cited on page 23.)
- [112] John F Marko and Eric D Siggia. Stretching dna. *Macromolecules*, 28(26):8759–8770, 1995. (Cited on page 28.)
- [113] Michelle D Wang, Hong Yin, Robert Landick, Jeff Gelles, and Steven M Block. Stretching dna with optical tweezers. *Biophysical journal*, 72(3):1335–1346, 1997. (Cited on page 28.)
- [114] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. (Cited on pages 29 and 31.)
- [115] Johannes Stigler and Matthias Rief. Hidden markov analysis of trajectories in single-molecule experiments and the effects of missed events. *ChemPhysChem*, 13(4):1079–1086, 2012. (Cited on page 29.)
- [116] Sean A McKinney, Chirlmin Joo, and Taekjip Ha. Analysis of single-molecule fret trajectories using hidden markov modeling. *Biophysical journal*, 91(5):1941–1951, 2006. (Cited on page 29.)
- [117] G David Forney. The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, 1973. (Cited on page 31.)

- [118] Andrew Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, 13(2):260–269, 1967. (Cited on page 31.)
- [119] Leonard E Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, 3(1):1–8, 1972. (Cited on page 31.)
- [120] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970. (Cited on page 31.)
- [121] Hendrik Dietz and Matthias Rief. Protein structure by mechanical triangulation. *Proceedings of the National Academy of Sciences*, 103(5):1244–1247, 2006. (Cited on pages 43 and 78.)
- [122] Biomers. <https://www.biomers.net/de/index.html?lang=de>, 2021. [Online; accessed 16-November-2021]. (Cited on page 46.)
- [123] Colin Echeverría Aitken, R Andrew Marshall, and Joseph D Puglisi. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical journal*, 94(5):1826–1835, 2008. (Cited on pages 57 and 58.)
- [124] Colin Echeverria Aitken, R Andrew Marshall, and Joseph D Puglisi. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical journal*, 94(5):1826–1835, 2008. (Cited on page 57.)
- [125] Marko Swoboda, Joerg Henig, Hsin-Mei Cheng, Dagmar Brugger, Dietmar Haltrich, Nicolas Plumeré, and Michael Schlierf. Enzymatic oxygen scavenging for photostability without ph drop in single-molecule experiments. *ACS nano*, 6(7):6364–6369, 2012. (Cited on pages 60 and 103.)
- [126] Bjoern Kauppi, Clarissa Jakob, Mathias Faernegaardh, Jie Yang, Harri Ahola, Maria Alarcon, Karin Calles, Owe Engstroem, John Harlan, and Steven Muchmore. The three-dimensional structures of antagonistic and agonistic forms of the glucocorticoid receptor ligand-binding domain: Ru-486 induces a transconformation that leads to active antagonism. *Journal of Biological Chemistry*, 278(25):22748–22754, 2003. (Cited on page 61.)
- [127] Kelly Suino-Powell, Yong Xu, Chenghai Zhang, Yong-guang Tao, W David Tolbert, S Stoney Simons Jr, and H Eric Xu. Doubling the size of the glucocorticoid receptor ligand binding pocket by deacylcortivazol. *Molecular and cellular biology*, 28(6):1915–1923, 2008. (Cited on page 61.)
- [128] Stephen H McLaughlin and Sophie E Jackson. Folding and stability of the ligand-binding domain of the glucocorticoid receptor. *Protein Science*, 11(8):1926–1936, 2002. (Cited on page 61.)
- [129] Isoelectric.org isoelectric calculator. <http://isoelectric.org/calculate.ph>, 2021. [Online; accessed 09-September-2021]. (Cited on page 62.)

- [130] Expasy.org isoelectric calculator. [https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/), 2021. [Online; accessed 09-September-2021]. (Cited on page 62.)
- [131] Protpi.ch isoelectric calculator. <https://www.protpi.ch/Calculator/ProteinTool>, 2021. [Online; accessed 09-September-2021]. (Cited on page 62.)
- [132] Kemper Talley and Emil Alexov. On the ph-optimum of activity and stability of proteins. *Proteins: Structure, Function, and Bioinformatics*, 78(12):2699–2706, 2010. (Cited on page 62.)
- [133] Sophia Diamant and Pierre Goloubinoff. Temperature-controlled activity of dnak- dnaj- grpe chaperones: Protein-folding arrest and recovery during and after heat shock depends on the substrate protein and the grpe concentration. *Biochemistry*, 37(27):9688–9694, 1998. (Cited on pages 63 and 83.)
- [134] Si Wu, Liu Hong, Yuqing Wang, Jieqiong Yu, Jie Yang, Hong Zhang, and Sarah Perrett. Kinetics of the conformational cycle of hsp70 reveals the importance of the dynamic and heterogeneous nature of hsp70 for its function. *Proceedings of the National Academy of Sciences*, 117(14):7814–7823, 2020. (Cited on page 63.)
- [135] Krzysztof Liberek, Daniel Wall, and Costa Georgopoulos. The dnaj chaperone catalytically activates the dnak chaperone to preferentially bind the sigma 32 heat shock transcriptional regulator. *Proceedings of the National Academy of Sciences*, 92(14):6224–6228, 1995. (Cited on pages 67 and 97.)
- [136] Joost Van Durme, Sebastian Maurer-Stroh, Rodrigo Gallardo, Hannah Wilkinson, Frederic Rousseau, and Joost Schymkowitz. Accurate prediction of dnak-peptide binding via homology modelling and experimental data. *PLoS Comput Biol*, 5(8):e1000475, 2009. (Cited on pages 69, 77, 78, 80, 85, and 88.)
- [137] J-H Ha, ER Johnson, DB McKay, MC Sousa, S Takeda, and SM Wilbanks. *Structure and mechanism of Hsp70 proteins*, pages 651–688. CRC Press, 1999. (Cited on page 70.)
- [138] MBB Gutierrez, CBC Bonorino, and MM Rigo. Chaperism: improved chaperone binding prediction using position-independent scoring matrices. *Bioinformatics*, 36(3):735–741, 2020. (Cited on pages 77, 79, and 109.)
- [139] Limbo algorithm by switchlab. <https://limbo.switchlab.org>, 2021. [Online; accessed 16-November-2021]. (Cited on pages 77 and 78.)
- [140] Chaperism algorithm. <https://github.com/BioinfLab/ChaperISM>, 2021. [Online; accessed 16-November-2021]. (Cited on pages 79 and 109.)
- [141] Sergey V Slepnev and Stephan N Witt. The unfolding story of the escherichia coli hsp70 dnak: is dnak a holdase or an unfoldase? *Molecular microbiology*, 45(5):1197–1206, 2002. (Cited on page 82.)
- [142] Vinay Dahiya, Ganesh Agam, Jannis Lawatscheck, Daniel Andreas Rutz, Don C Lamb, and Johannes Buchner. Coordinated conformational processing of the tumor suppressor protein p53 by the hsp70 and hsp90 chaperone machineries. *Molecular cell*, 74(4):816–830. e7, 2019. (Cited on pages 82 and 93.)
- [143] Ruth Kellner, Hagen Hofmann, Alessandro Barducci, Bengt Wunderlich, Daniel Nettels, and Benjamin Schuler. Single-molecule spectroscopy reveals

- chaperone-mediated expansion of substrate protein. *Proceedings of the National Academy of Sciences*, 111(37):13355–13360, 2014. (Cited on pages 82, 85, and 97.)
- [144] Rui Sousa, Hsien-Shun Liao, Jorge Cuéllar, Suping Jin, Jose M Valpuesta, Albert J Jin, and Eileen M Lafer. Clathrin-coat disassembly illuminates the mechanisms of hsp70 force generation. *Nature structural & molecular biology*, 23(9):821–829, 2016. (Cited on page 82.)
- [145] Walter Neupert and Johannes M Herrmann. Translocation of proteins into mitochondria. *Annu. Rev. Biochem.*, 76:723–749, 2007. (Cited on page 82.)
- [146] Fernanda Rodriguez, Florence Arsène-Ploetze, Wolfgang Rist, Stefan Rüdiger, Jens Schneider-Mergener, Matthias P Mayer, and Bernd Bukau. Molecular basis for regulation of the heat shock transcription factor sigma32 by the dnak and dnaj chaperones. *Molecular cell*, 32(3):347–358, 2008. (Cited on page 82.)
- [147] Ellen AA Nollen, Jeanette F Brunsting, Jaewhan Song, Harm H Kampinga, and Richard I Morimoto. Bag1 functions in vivo as a negative regulator of hsp70 chaperone activity. *Molecular and cellular biology*, 20(3):1083–1088, 2000. (Cited on page 83.)
- [148] Gian Paolo Rossini and Claudia Malaguti. Nanomolar concentrations of untransformed glucocorticoid receptor in nuclei of intact cells. *The Journal of steroid biochemistry and molecular biology*, 51(5-6):291–298, 1994. (Cited on page 83.)
- [149] Rick Russell, A Wali Karzai, Andrew F Mehl, and Roger McMacken. Dnaj dramatically stimulates atp hydrolysis by dnak: insight into targeting of hsp70 proteins to polypeptide substrates. *Biochemistry*, 38(13):4165–4176, 1999. (Cited on page 83.)
- [150] Thomas Laufen, Matthias P Mayer, Christian Beisel, Dagmar Klostermeier, Axel Mogk, Jochen Reinstein, and Bernd Bukau. Mechanism of regulation of hsp70 chaperones by dnaj cochaperones. *Proceedings of the National Academy of Sciences*, 96(10):5452–5457, 1999. (Cited on page 83.)
- [151] Jean-Pierre Changeux and Stuart Edelman. Conformational selection or induced fit? 50 years of debate resolved. *F1000 biology reports*, 3, 2011. (Cited on page 87.)
- [152] Daniel Schmid, Antonio Baici, Heinz Gehring, and Philipp Christen. Kinetics of molecular chaperone action. *Science*, 263(5149):971–973, 1994. (Cited on page 87.)
- [153] Serge M Gisler, Ezra V Pierpaoli, and Philipp Christen. Catapult mechanism renders the chaperone action of hsp70 unidirectional. *Journal of molecular biology*, 279(4):833–840, 1998. (Cited on page 87.)
- [154] Basile Nguyen, David Hartich, Udo Seifert, and Paolo De Los Rios. Thermodynamic bounds on the ultra-and infra-affinity of hsp70 for its substrates. *Biophysical journal*, 113(2):362–370, 2017. (Cited on page 87.)

- [155] Pierre Goloubinoff and Paolo De Los Rios. The mechanism of hsp70 chaperones:(entropic) pulling the models together. *Trends in biochemical sciences*, 32(8):372–380, 2007. (Cited on page 88.)
- [156] Alireza Mashaghi, Sergey Bezrukavnikov, David P Minde, Anne S Wentink, Roman Kityk, Beate Zachmann-Brand, Matthias P Mayer, Günter Kramer, Bernd Bukau, and Sander J Tans. Alternative modes of client binding enable functional plasticity of hsp70. *Nature*, 539(7629):448–451, 2016. (Cited on page 88.)
- [157] Shruti Sharma, Kausik Chakraborty, Barbara K Müller, Nagore Astola, Yun-Chi Tang, Don C Lamb, Manajit Hayer-Hartl, and F Ulrich Hartl. Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell*, 133(1):142–153, 2008. (Cited on page 88.)
- [158] Roman Kityk, Jürgen Kopp, and Matthias P Mayer. Molecular mechanism of j-domain-triggered atp hydrolysis by hsp70 chaperones. *Molecular cell*, 69(2):227–237. e4, 2018. (Cited on page 90.)
- [159] Douglas M Cyr, Thomas Langer, and Michael G Douglas. Dnaj-like proteins: molecular chaperones and specific regulators of hsp70. *Trends in biochemical sciences*, 19(4):176–181, 1994. (Cited on pages 90 and 97.)
- [160] Michael E Cheetham and Avrom J Caplan. Structure, function and evolution of dnaj: conservation and adaptation of chaperone function. *Cell stress & chaperones*, 3(1):28, 1998. (Cited on pages 90 and 97.)
- [161] Chun-Yang Fan, Soojin Lee, and Douglas M Cyr. Mechanisms for regulation of hsp70 function by hsp40. *Cell stress & chaperones*, 8(4):309, 2003. (Cited on page 90.)
- [162] Lumicks. <https://lumicks.com/>, 2021. [Online; accessed 26-October-2021]. (Cited on page 98.)
- [163] Ciro Cecconi, Elizabeth A Shank, Carlos Bustamante, and Susan Marqusee. Direct observation of the three-state folding of a single protein molecule. *Science*, 309(5743):2057–2060, 2005. (Cited on page 103.)
- [164] Marco Grison, Ulrich Merkel, Julius Kostan, Kristina Djinović-Carugo, and Matthias Rief. alpha-actinin/titin interaction: A dynamic and mechanically stable cluster of bonds in the muscle z-disk. *Proceedings of the National Academy of Sciences*, 114(5):1015–1020, 2017. (Cited on page 103.)