Dissertation zur Erlangung des Doktorgrades
an der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Hsp70 resolves misfolded states and accelerates productive folding of multi-domain proteins

Rahmi İmamoğlu

aus

Fatih/Istanbul, Turkey

2021

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung von 28. November 2011 von Herrn Prof. Dr. F. Ulrich Hartl betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 18/01/2021

Rahmi İmamoğlu

Dissertation eingereicht am:	08.02.2021
1. Gutachter	Prof. Dr. F. Ulrich Hartl
2. Gutachter:	Prof. Dr. Don C. Lamb
Mündliche Prüfung am:	07.07.2021

Acknowledgement

I would foremost like to thank my supervisors, Prof. Dr. Franz Ulrich Hartl and Dr. Manajit Hayer-Hartl, for providing me with the opportunity to conduct my Ph.D. research on this exciting project in collaboration with very talented people. I am very thankful to them for the continuous support, advice and guidance, which has helped me to grow as a scientist. Working with them taught me how to ask good biological questions, a skill that will be invaluable to my future career.

I am truly indebted and grateful to my Postdoc supervisor Dr. David Balchin for his very valuable guidance throughout the research. Without his guidance, it would have been impossible to conduct this research successfully. Also, I would like to thank him for many great discussions on scientific and non-scientific topics.

Besides my supervisors, I would like to thank Dr. Javaid Y. Bhat for friendship, moral support, scientific and non-scientific advice for my career. Also, I would like to wish a very special thanks to Dr. Andreas Bracher for critical discussions on my research project.

I would like to thank to all Hartl-fighters, especially Dr. Patricia Yuste Checa, Dr. Goran Miličić, Amit. K. Singh, Dr. Xiao Yan, Huping Wang, Dr. Liang Zhao, Dr. Cole S. Sitron, Dr. Shubhasis Haldar, Dr. Amit Jean Gupta, Michael Gropp, Martin Müller, Mirkko Flecken, Alonso I. C. Alvarez, Omar Zaki.

I thankfully acknowledge the strong support and help of the Hartl-lab technical staff: Darija Pompino, Emmanuel Burghardt, Albert Ries, Nadine Wischnewski, Silvia Gaertner, Romy Lange and Anastasia Jungclaus.

My sincere thanks also go to members of the QBM office, Dr. Filiz Civril, Mara Kieke, Dr. Julia Schlehe, Mirely Kollmannsberger, Dr. Markus Hohle and Dr. Michael Mende. I would like to thank my QBM friends, especially Andrea Luckas, Roberto Cortini, Vicente Yepez, Gökçen Eraslan, Başak Eraslan, Alena Khmelinskaia, Linda Krause, Laia Pascual Ponce, Madlin Schenk, Daniela Garcia, Elisavet Chatzopoulou.

Last but not the least, my deepest thanks go to my family and especially my wife, Deniz İmamoğlu and my daughter, Didem İmamoğlu. Without their constant support and patience this work would not exist.

Contents

List of abbreviations

ADP	Adenosine 5’-diphosphate
AFM	Atomic force microscopy
AUC	Analytical ultracentrifugation
ATP	Adenosine 5’-triphosphate
BAG1	Bcl2-associated athanogene 1
CHIP	C-terminus of HSC70-interacting protein
Cys	Cysteine
dcFCCS	Dual-colour fluorescence cross-correlation spectroscopy
DNA	Deoxyribonucleic acid
DnaJ	*E. coli* homologue of Hsp40
DnaK	*E. coli* homologue of Hsp70
DOL	Degree of labelling
DTT	Dithiothreitol
E. coli	*Escherichia coli*
FCS	Fluorescence correlation spectroscopy
FLuc	Firefly luciferase
FRET	Fluorescence/Förster resonance energy transfer
GrpE	*E. coli* homologue of nucleotide exchange factor
GuHCl	Guanidine hydrochloride
Hepes	2-[4-(2-hydroxythyl)piperazin-1-yl]ethanesulfonic acid
Hsp	Heat shock protein
Hsc70	Heat shock cognate 70-kDa protein
Hsp70	The 70-kDa heat shock protein
IPTG	Isopropyl β-D-thiogalactopyranoside
LB	Lysogenic broth
NAC	Nascent polypeptide-associated complex
NBD	N-terminal nucleotide binding domain of Hsp70
PBT	Phenylbenzothiazole
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate

PFD	Prefoldin
PIE	Pulsed interleaved excitation
RAC	Ribosome-associated complex
SBD	C-terminal substrate binding domain of Hsp70
SDS-PAGE	Sodium dodecylsulfate polyacrylamid gelelectrophoresis
SEC	Size-exclusion chromatography
smFRET	Single-molecule FRET
TCEP	Tris(2-carboxyethyl)phosphine
TCSPC	Time-correlated single photon counting
TCP-1	Tailless complex polypeptide-1
TRiC	TCP-1 ring complex
TF	Trigger factor
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol

1 Summary

The 70-kDa heat shock protein (Hsp70) is a ubiquitously expressed molecular chaperone that plays an essential role in protein folding and protein homeostasis. The chaperone function of Hsp70 is achieved via cooperation with J domain proteins and nucleotide exchange factors. The Hsp70 system is proposed to act as a passive chaperone of protein folding. However, the protein folding mechanism of the Hsp70 chaperone system is poorly defined, and it is unknown whether Hsp70 plays an active role in folding multi-domain proteins that are genuine clients of the chaperone.

Here, we show that the *E. coli* homologue of Hsp70 (DnaK) and its cognate co-chaperones (DnaJ and GrpE) cooperate to accelerate folding of the model multi-domain client firefly luciferase (FLuc) at least 20-fold over the spontaneous rate. Using single-molecule Förster resonance energy transfer (smFRET), we identify a compact inter-domain misfold as the cause of slow spontaneous folding of FLuc, and demonstrate that DnaK resolves the kinetic trap by selective conformational expansion of the misfolded region. DnaK/DnaJ/GrpE alter the pathway of FLuc folding, and achieve accelerated folding in a single round of client engagement. Furthermore, accelerated folding by the Hsp70 system can be leveraged to maintain the native state of FLuc, even under conditions that would otherwise be denaturing. These findings establish a novel function for the Hsp70 chaperone system in protein folding. The concept of chaperone-accelerated folding has broad implications for the role of Hsp70 in both protein biogenesis and conformational maintenance of the proteome under stress conditions.

2 Introduction

2.1 The formation of proteins

It is unclear how, where and when life began. Also unsolved is how amino acids, the building blocks of proteins, originated on the Hadean Earth. The Earth was formed about 4.5 billion years ago and life on Earth began approximately 3.8 billion years ago. Life is basically characterized by compartmentalization, replication and metabolism (Kitadai and Maruyama, 2018). These three functions are operated by macromolecules such as nucleic acids, proteins, carbohydrates and phospholipids. It is assumed that the monomeric subunits of macromolecules were formed abiotically, condensed into polymers, melded together, and evolved into self-sustaining organisms (Kitadai and Maruyama, 2018). Nucleic acids carry raw genetic information that is translated into proteins that are involved in most life processes such as nucleic acid synthesis, metabolic regulation, translocation of proteins to other locations, transferring oxygen to tissues, etc. Proteins are responsible for nearly all biological functions.

2.2 Protein structure

Proteins are chains of amino acids that are composed of a central carbon (alpha-carbon) atom bonded to an amino group (-NH2), a carboxyl group (-COOH), a hydrogen and a fourth group (-R) that distinguishes and characterizes the chemical properties of the 20 different amino acids. In principle, amino acids can be categorized as hydrophobic, polar and charged according to the chemical properties of the –R group. Amino acids (except glycine) conformationally occur in two optical isomer forms that are either L or D since they contain the chiral carbon atom and only L-amino acids are incorporated into proteins in cells (van Heijenoort, 2001). Amino acids are connected together by peptide bonds to form proteins. Peptide bonds covalently form between two amino acids in a condensation reaction, where the amino group ($–NH_2$) of one amino acid reacts with the carboxyl group (–COOH) of another amino acid, releasing a molecule of water (H_2O). Peptide bonds are very stable and planar because of their partial double bond character (Figure 2.1). The linear, elongated chain of amino acids within a protein is called the primary sequence. Much of the polypeptide backbone of proteins forms defined secondary structures due to the hydrogen bond between the main chain N-H and C=O groups. α-helices, β-strands and turns are the most commons secondary structure elements. The α-helix contains 3.6 residues per turn with hydrogen bonds between the -COOH group of residue n and the $–NH_2$ group of residues

(n+4). Their right-handed nature forms a dipole moment. β-strands are the other most important secondary structure element formed by the hydrogen bond between the carbonyl oxygen of one amino acid and the amide proton of the second amino acid. There exist two types of β-sheet; parallel or antiparallel. The secondary structural elements can pack together and form long-range interactions for tertiary conformations, which are stabilized, by non-covalent interactions such as hydrogen bonds, ionic contacts, Van der Waals interactions and π-π, π-cation interactions. Some proteins are formed by more than one polypeptide and these subunits come together to form more complex quaternary structures. The main driving force for the formation of these structural elements is to shield hydrophobic residues from the polar environment. The process by which the defined structural elements of a protein form is called protein folding.

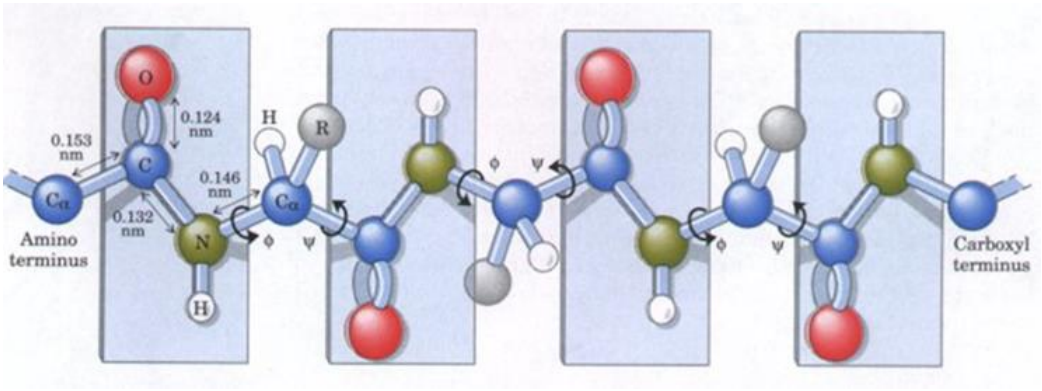


Figure 2.1 The polypeptide backbone
A simplified illustration of a short peptide with four amino acids. The dotted line represents one amino acid. The planer geometry of peptide bond restricts the movement (blue surfaces). The dihedral angles between Cα–C and N–Cα are named as ψ and φ, respectively. R is the functional group of amino acids. Modified from http://www.bioinfo.org.cn/book/biochemistry/chapt07/bio1.htm.

2.3 Protein folding

Most proteins must fold into a unique three-dimensional structure to fulfil their biological role (Balchin et al., 2016). One of the most important milestones in protein folding science is the thermodynamic hypothesis of Christen Anfinsen who explicitly showed in the 1950s that the native conformation of a protein under physiological conditions is the thermodynamically most stable state and is encoded in the amino acid sequence (Anfinsen, 1973a). Although, Anson and Mirsky already observed in the 1930s that protein denaturation can be reversed (Anson and

Mirsky, 1931), it was Anfinsen who demonstrated the significance and the consequence of reversible folding. Anfinsen and colleagues tested the reversibility of fully denatured ribonuclease with reductive cleavage of four disulphide bonds. They observed that the native state was reached with the correct combination of oxidized disulfide bonds, which represents one of 105 possible pairings of the eight sulfhydryl groups (Anfinsen, 1973a; Anfinsen and Haber, 1961; Anfinsen et al., 1961; Sela et al., 1957; White, 1961). Thus, the pioneering experiment of Anfinsen shows that the primary amino acid sequence of a protein contains all the information for its final native structure, which is the most thermodynamically stable state. Christen Anfinsen received the Nobel Prize in Chemistry for this finding in 1972. However, despite elucidating the principle of protein folding, his finding gave rise to the major question concerning the folding pathway of a polypeptide, specifically whether folding occurs as a random search process or along a guided pathway. In 1968, Cyrus Levinthal demonstrated through theoretical considerations that even a small 100-residue polypeptide could adopt 2^{100} different conformations because of the very large number of degrees of chain freedom present in the unfolded polypeptide. Even if each amino acid within the protein can switch conformations in one picosecond, it would take 2^{100} ps to reach the native state (Levinthal, 1968). This would mean that this random search would take much longer than the age of the universe (the Levinthal paradox). However, proteins, particularly small proteins <100 residues, fold very fast in the timescale of milliseconds. Accordingly, Levinthal proposed that protein folding could not be a random search process, but it must occur via a preferred pathway or pathways. What this means is that instead of acquiring the most thermodynamically stable native conformation via random search, a protein must reach its native state via metastable structure(s) (local energy minimum). The following models have been proposed to explain how a protein reaches its native state without sampling all the possible conformations.

(i) Framework model: The so-called framework model suggests that protein folding is a hierarchical process. The initial step of protein folding is the formation of secondary structures. These elements come together by diffusion and collision to form the tightly packed native structure (Karplus and Weaver, 1976) (Figure 2.2). (ii) Nucleation model: This model postulates that some neighbouring residues form native like contacts and secondary structures which further act as nucleation centre to propagate the native structure (Wetlaufer, 1973) (Figure 2.2). (iii) Hydrophobic collapse model: This model assumes that protein will rapidly collapse around its hydrophobic residues as a result of the hydrophobic effect, followed by the subsequent formation of

secondary structural elements. Therefore, the tertiary structure forms first and proceeds with the formation of secondary structural elements (Dolgikh et al., 1981) (Figure 2.2). In the beginning of the 1990s, Sophie Jackson and Alan Ferhst showed that chymotrypsin inhibitor II (Cl2), a 64 residue polypeptide, folds by an apparent two-state mechanism (Jackson and Fersht, 1991), and thus concluded that during the folding transition state, all elements (secondary and tertiary structures) are formed simultaneously. Based on this finding and the analysis of other small proteins combined with molecular simulations, the so-called nucleation-condensation model was proposed in the 1995 (Itzhaki et al., 1995). According to this model folding is initiated by the formation of a folding nucleus that is characterized by the simultaneous formation of secondary and tertiary structural elements (Nölting & Agard, 2008), that is subsequently followed by the formation of the native structure (Figure 2.2).

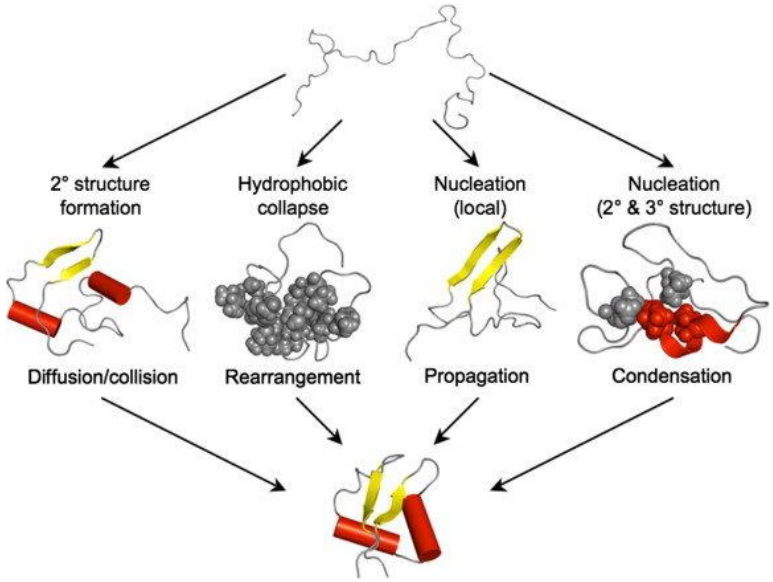


Figure 2.2 Different proposed model of protein folding mechanisms
(1) The framework model proposes the initial formation of local secondary structures. These elements diffuse and collide to form correct tertiary structures. The formation of tertiary structures was suggested as the rate-limiting step. (2) The hydrophobic collapse model suggests that folding starts with the collapse of hydrophobic side chains, followed by the rearrangement of the structural elements to a "molten globule" intermediate. (3) The nucleation propagation model states that local native-like secondary structures act as a nucleus for the propagation of the native structure. (4) The nucleation condensation model proposes that the folding of the nucleus occurs from the simultaneous formation of secondary and tertiary structural elements, with the native structure being formed by the rapid condensation of these elements. Modified from (Nickson and Clarke, 2010).

2.4 The energy landscape and folding funnels

In the 1990s, “folding funnels” were proposed as a multidimensional representation of the folding process by Ken A. Dill and Hue Sun Chan (Dill and Chan, 1997). The folding funnel hypothesis assumed that the energetically favourable stable native state with the lowest free energy is located at the bottom of the funnel, whereas an unfolded protein chain with a high degree of freedom is located at the top of the funnel. The native-like contacts form as protein chains progress along a downhill path toward the native state. Although small single domain proteins can often fold rapidly without being trapped in local energy minima, the folding of large, multi-domain proteins usually populates kinetically stable folding intermediates that need to cross substantial energy barriers to allow rearrangement and progression to the native state. Exposed hydrophobic regions of stable folding intermediates can give rise to aggregation. Aggregation is mediated by concentration dependent inter-molecular contacts between exposed hydrophobic regions of proteins. Although many aggregates are small and amorphous, in some cases they may form highly ordered amyloid fibrils, which are structurally defined by beta strands running perpendicular to the long axis of the fibril. Amyloid fibrils may be toxic to cells and are associated with neurodegenerative diseases such as Alzheimer’s, Huntington’s and Parkinson’s diseases (Figure 2.3).

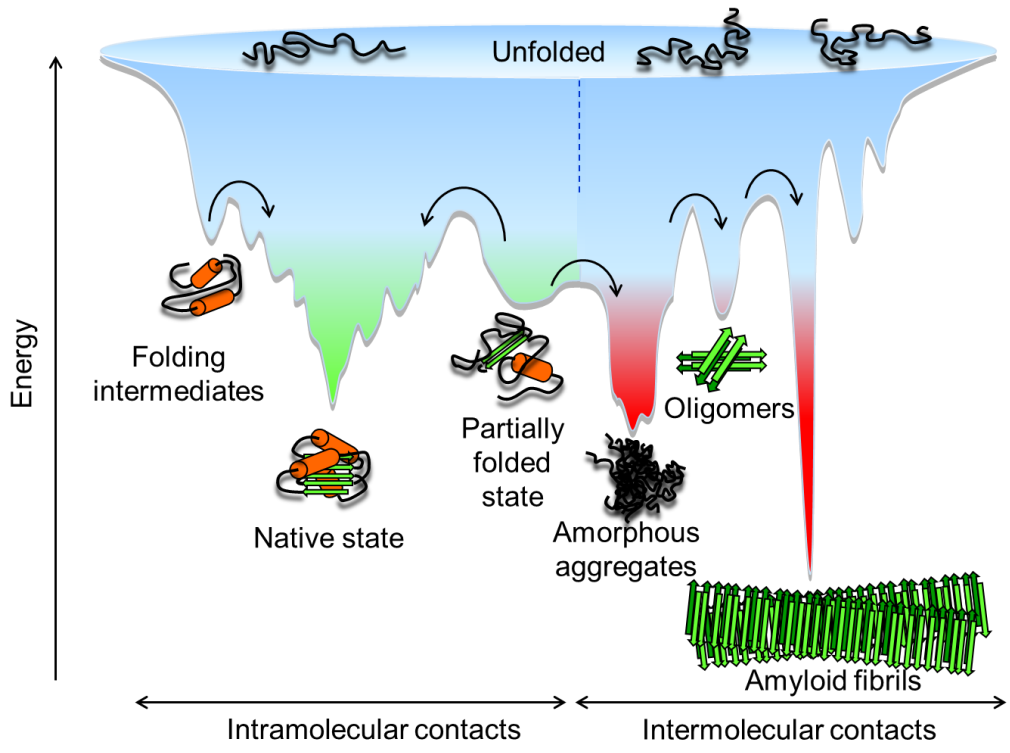


Figure 2.3 The folding energy landscape

Nascent polypeptide chains sample various conformations as they progress toward the energetically favourable native state. As the polypeptide chain progresses on a downhill path toward the native state, intramolecular contacts stabilize the conformation. Large and multi-domain proteins tend to populate energetically stable, non-native folding intermediates and partially folded states that have exposed hydrophobic regions. Inter-molecular contacts of these hydrophobic regions cause aggregation (amorphous aggregates, oligomers and amyloid fibrils). Chaperones bind to exposed hydrophobic regions and prevent aberrant inter-molecular interactions. Modified from (Hartl et al., 2011).

2.5 Protein folding in the cell

Anfinsen showed in his pioneering experiments that the three-dimensional conformation of proteins is encoded by the primary amino acid sequence, and thus that folding can proceed spontaneously, at least in diluted solution in the test tube (Anfinsen, 1973a). However, the cell cytosol is tightly packed with proteins and other macromolecules (300-400 g/L), resulting in excluded volume effects and macromolecular crowding (Zimmerman and Trach, 1991). This has a profound effect on both the thermodynamic and kinetics properties of proteins (Ellis, 2001). Importantly, in such an environment, non-specific interactions are favoured, particularly for large polypeptides. Therefore, crowding strongly enhances non-native interactions among proteins, causing aggregation. Polypeptide chains must reach their native state under these challenging environmental conditions to become functionally active.

Another important difference between *in vitro* and *in vivo* folding is that the latter occurs in the context with the vectorial translation of the polypeptide chain on the ribosome, whereby N-terminal parts of the chain could initiate folding, followed by more C-terminal segments (Kim et al., 2013). Physical separation of domains by the ribosome can prevent intra- or inter-domain misfolding, thus promoting productive folding (Agashe et al., 2004; Netzer and Hartl, 1998). Interestingly, the translation rate in eukaryotes (4 amino acids per second) is about 5-times slower than that of bacteria (20 amino acids per second) (Hartl and Hayer-Hartl, 2009). It is thought that this evolutionary adaptation of the translation speed provides eukaryotic proteins with more time for productive folding. Note that the average size of proteins increased from ~ 35 kDa in bacteria to ~52 kDa in humans (Kaiser et al., 2011; Kim et al., 2013).

2.6 The proteostasis network (PN)

Cells have evolved diverse mechanisms to ensure protein homeostasis (proteostasis) and that protein synthesis, folding, conformational maintenance, and degradation are in proper balance under a range of environmental conditions (Klaips et al., 2018). About 10% of the human genome is devoted to proteins that play a role in maintain proteostasis, a major investment (Hipp et al., 2019) (Figure 2.4).

Human cells possesses about 10 million ribosome, each producing a new protein every 1-2 minutes, with ~10.000 different proteins being expressed at any given time (Kulak et al., 2017). Many newly synthesized proteins have to obtain their native structure co-translationally or by the assistance of folding helper proteins, termed molecular chaperones. Furthermore, after initial folding, the native conformation of proteins has to be maintained for them to fulfil their biological activities. The proteins that are unable to fold efficiently or unable to maintain the native conformation can form terminally misfolded proteins, which require degradation by the ubiquitin-proteasome system (UPS) or autophagy-lysosome system.

The proteostasis network (PN), including molecular chaperones and protein degradation machinery, ensures the maintenance of healthy proteome, but sustaining proteome balance is a challenging task because of the high level of endogenous and external stresses (Balchin et al., 2016). It is assumed that the capacity of the PN significantly declines during aging, thereby increasing the risk of the formation of toxic protein aggregates. Therefore, impairment of the PN predisposes for numerous diseases, including neurological disorders such as Alzheimer's, Parkinson's and Huntington's diseases. Understanding how cells normally maintain proteostasis can inform on the development of new therapeutic strategies.

Here, I will mainly focus on the role of protein folding by molecular chaperones.

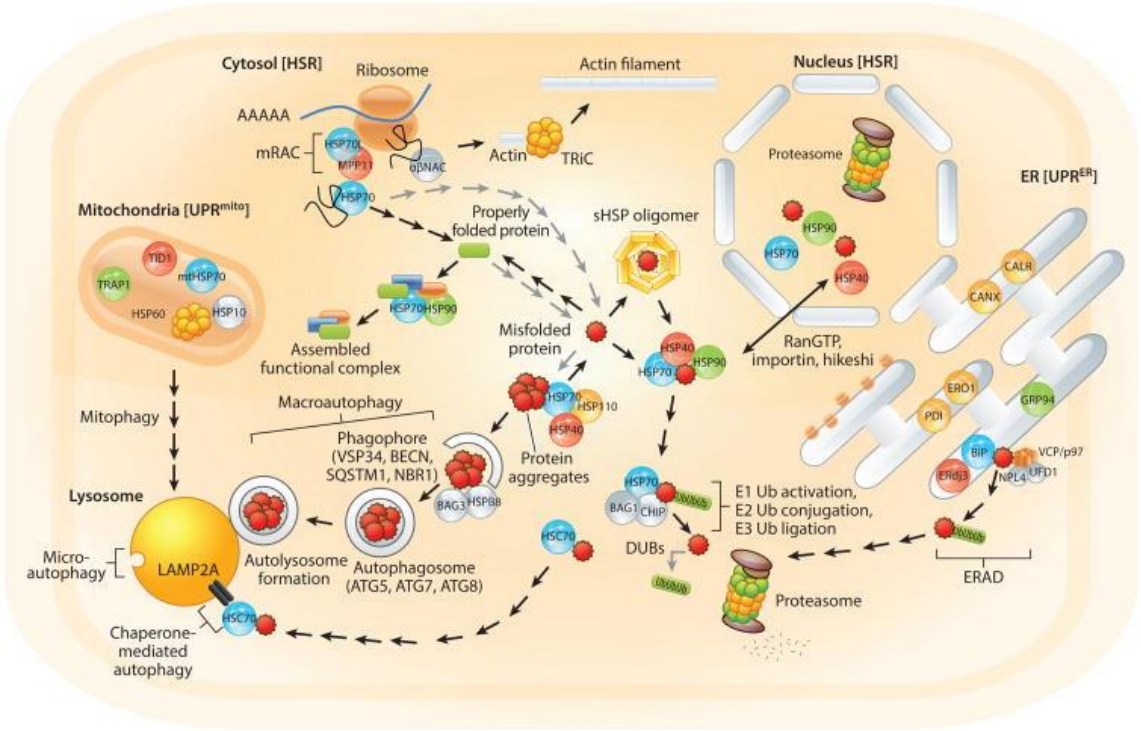


Figure 2.4 General overview of the proteostasis network

The Hsp70 (blue spheres), Hsp40 (red spheres) and Hsp90 (green spheres) chaperone families are found in the cytosol and within organelles. They work together with co-chaperones (gray spheres) to assist the folding of nascent chains and folding intermediates, assembly of protein complexes, refolding and disaggregation of misfolded proteins (serrated red spheres) and degradation of terminally misfolded proteins by the proteasome and in some cases by autophagy. Small heat shock protein (sHsp) associate with misfolded proteins and hold them in a reversible, folding-competent state for the Hsp70 chaperone system to work on. If refolding is not achieved, Hsp70 and its co-chaperones, Bcl2-associated athanogene 1 (BAG1) and the E3 ubiquitin ligase C-terminus of HSC70-interacting protein (CHIP), can direct substrate proteins to the proteasome for clearance. Tailless complex polypeptide-1 (TCP-1) ring complex (TRiC) in the cytosol plays a crucial role for the folding of folding of cytoskeletal proteins and the Hsp60 system is required for mitochondrial proteostasis (with the co-chaperone Hsp10). Under stress conditions, stress response pathways, such as the heat shock response and unfolded protein responses of the ER (UPR_{ER}) and mitochondria (UPR_{mito}), can be triggered and they induce the synthesis of chaperones. Hsp110 synergizes with Hsp70/Hsp40 to removes individual proteins from the aggregates, thus acting as a disaggregase. The degradation of persistence protein aggregates is executed by the ubiquitin-proteasome system (UPS). The lysosome through autophagy might be required for the degradation of larger aggregates. Modified from (Labbadia and Morimoto, 2015).

2.7 Molecular chaperones

In the early 1960s, Ferrucio Ritossa was studying the puffing pattern of the salivary gland chromosomes of the fruitfly *Drosophila busckii*. Interestingly, he observed that special regions of polytene chromosomes, chromosomal puffs, are highly induced upon heat shock, suggesting that the transcriptional activation serves to produce specialized proteins under temperature stress (Ritossa, 1962). Twelve years after Rittossa's observation, Alfred Tissieres and his colleagues identified the six proteins that are most strongly induced in the salivary gland of *Drosophila melanogaster* by heat shock (Tissières et al., 1974). Therefore, these proteins were called heat shock proteins and named according to their molecular weight. Following studies revealed that the heat shock response is highly conserved in all three kingdoms (Figure 2.5) and that the overexpression of heat shock proteins is induced not only by thermal upshift, but also by other stress factors such as oxidative stresses, pH shifts and heavy metal exposure (Lindquist, 1986; Morimoto, 1993).

In 1978, Ron Laskey identified an acidic protein required for the assembly of nucleosomes (Laskey et al., 1978). This acidic protein binds to histones and transfers it to DNA for proper formation of the nucleosome. He called this acidic protein a "molecular chaperone", as it neither became a part of the final assembled complex. Two years later, in 1980, John Ellis saw an unexpected result while they were studying ribulose biphosphate carboxylase (Rubisco), the CO_2-fixing enzyme complex in chloroplast. Rubisco consists of large and small subunits that need to be properly assembled for the formation of the active enzyme. Interestingly, Ellis and colleagues found that assembly occurs only in the presence of another protein, later called a chaperonin (Barraclough and Ellis, 1980). In 1986, Hugh Pelham suggested that Hsp70 and Hsp90 are not only overexpressed under stress conditions but may be more generally involved in the assembly and disassembly of proteins (Pelham, 1986). While the chaperone concept (Ellis, 1987) was first focused on the role of these machineries in assembly and disassembly reactions (Cheng et al., 1989; Goloubinoff et al., 1989a; Goloubinoff et al., 1989b), the breakthrough study in 1989 showed that chaperones are crucial for the folding of some proteins (Ostermann et al., 1989), which changed in the long held-belief that protein folding in the cell occurs spontaneously.

Molecular chaperones are highly conserved from bacteria to human. Many of them are heat shock proteins (Hsp). We currently define a molecular chaperone as a protein that interacts with another protein, and assist its folding and/or assembly without being a part of final native structure (Kim

et al., 2013). A major function of chaperones is to prevent off-pathway reactions in folding that lead to aggregation (Kim et al., 2013). In addition to their role in *de novo* protein folding, chaperones can assist complex assembly, protein transport and disaggregation reactions. To perform these tasks, in general, chaperones bind to exposed hydrophobic residues of non-native proteins, either in an ATP-dependent or ATP-independent manner. The most important chaperone classes include ribosome-binding chaperones, small Hsp (sHsp), Hsp60, Hsp70, Hsp90 and Hsp100. The chaperone pathways in the cytosol are highly conserved throughout all domains of life (Kim et al., 2013).

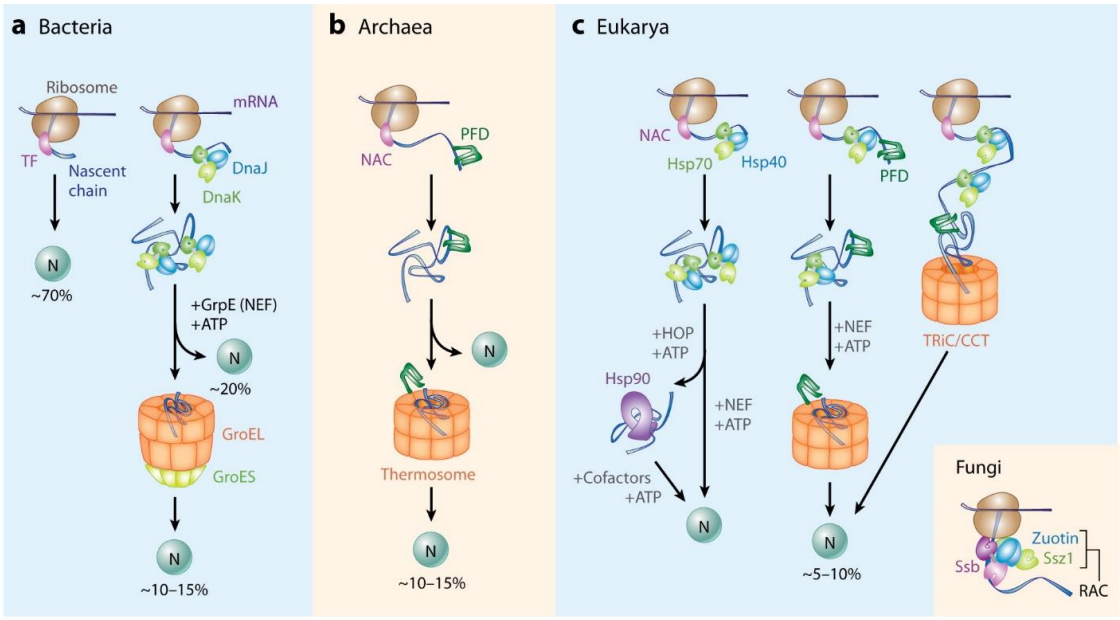


Figure 2.5 Cytosolic chaperone pathways in all domains of life
In bacteria (a), archaea (b), and eukarya (c), ribosome-bound chaperones (trigger factor (TF) in bacteria, nascent-chain-associated complex (NAC) in archaea and eukarya) assist in cotranslational folding of the nascent chains as they emerge from the ribosome. Cytosolic, non-ribosomal-bound members of the Hsp70 family (DnaK in bacteria and Hsp70 in eukarya), together with two co-chaperones, which are Hsp40s and nucleotide exchange factors (NEFs), assist co- and post-translational folding and thus function as a second chaperone system. In archaea, prefoldin (PFD) assists in folding as a second chaperone system, since archaea lacks of Hsp70. Partially folded substrates may need to be transferred to another downstream chaperone class, called the chaperonin (GroEL/ES in bacteria, thermosome in archaea, TRiC in eukarya). The Hsp90 system with might mediate the folding and maturation of Hsp70-released substrates. The ribosome-associated complex (RAC), in fungi is an additional ribosome-binding chaperone system, which consists of Ssz1 (a specialized Hsp70) and zuotin (Hsp40). Cooperation with another Hsp70 isoform (Ssb), they assist nascent chain folding. Percentages indicates the relative amount of proteins in whole proteome, interacting for a given chaperone. Modified from (Kim et al., 2013).

2.7.1 Ribosome-associated chaperones

Protein synthesis on the ribosome occurs in a vectorial manner from N- to C- terminus, long-range interactions are thus completed post-translationally. As the nascent chain emerges from the ribosome exit tunnel, ribosome-binding chaperones prevent it from unfavorable intra- and inter-molecular interactions by binding to exposed hydrophobic regions of the emerging nascent chain.

In bacteria, trigger factor (TF), (~ 50 kD), a bacterial ribosome-binding chaperone, consists of three domains. The N-terminal domain of TF includes a helix-loop-helix motif for ribosome binding. The middle domain have catalytic activity as peptidyl-prolyl cis/trans isomerase (PPIase) as well as C-terminal domain displays chaperone activity by binding exposed hydrophobic regions of emerging nascent chains in an ATP-independent manner when nascent chains reach ~100 amino acids in length (Hoffmann et al., 2010; Kim et al., 2013). Although TF alone is not essential for bacteria at 37°C, the deletion of both TF and DnaK is lethal at temperatures above 30°C, indicating that TF and DnaK have redundant functions (Deuerling et al., 1999; Teter et al., 1999). TF slows down the rate of co-translational folding, thereby preventing premature folding (O'Brien et al., 2012). Interestingly, single-molecule studies with TF suggested that TF can also guide unfolded proteins toward the native state, thereby modulating the energy landscape, which is in contrast to a model in which the chaperone only prevents aggregation (Mashaghi et al., 2013).

Ribosome-associated complex (RAC) and nascent polypeptide-associated complex (NAC) are the main ribosome-binding chaperones in eukaryotes. RAC is a heterodimer consisting of an atypical Hsp70 (Ssz1 in yeast; Hsp70L1 in mammals) and a J domain protein (Zuo1 in yeast; MPP11 in mammals) (Kramer et al., 2019; Weyer et al., 2017). Additionally, Ssb1 or Ssb2, ribosome-binding Hsp70 proteins, are functionally linked to RAC and form a functional unique triad on the ribosome in yeast. The Ssz1/Zuo1 complex stimulates the ATPase activity of Ssb. Consequently, SSB deletion causes aggregation which is enhanced by the deletion of NAC (Koplin et al., 2010). However, mammalian RAC lacks Ssb-like proteins, instead cytosolic Hsp70 is recruited to the RAC complex (Preissler and Deuerling, 2012). The second ribosome-associated complex is a highly conserved heterodimeric (α- and, β-subunits) protein, NAC (Kramer et al., 2019). In general, NAC is similar to TF in that it is ATP-independent, interacts stably with a wide variety of nascent chains and is of high abundance (Kramer et al., 2019). However, the exact role and mechanism of NAC is not yet clear (Kim et al., 2013). NAC is

essential in metazoans but is dispensable in yeast; however, simultaneous deletion of Ssb and NAC is lethal, similar to the situation of TF and DnaK in *E. coli* (Kramer et al., 2019; Markesich et al., 2000). Together, ribosome-associated chaperones protect nascent chains and prevent them from unfavourable interactions as they emerge from the ribosome.

2.7.2 The Hsp70 system

Hsp70 chaperones are a ubiquitously expressed and highly conserved chaperone family found in bacteria, some archaea, the eukaryotic cytosol and in organelles. They function as a central hub of the homeostasis network and play an important role in a wide range of cellular functions, including *de novo* protein folding, refolding, disaggregation, oligomer assembly and protein transfer to cellular compartments. Thus, they are involved in all stages of the life of proteins (Balchin et al., 2016; Rosenzweig et al., 2019). This high versatility is achieved by Hsp70 binding to short hydrophobic peptide segments, consisting of five hydrophobic amino acids flanked by positively charged residues (Rudiger et al., 1997).

Hsp70s amount to 0.5-2% of total protein mass under normal cellular conditions (~ 30 µM in *E. coli*) and increase around two-fold under stress conditions (Goloubinoff, 2017). The *E. coli* homologue of Hsp70, DnaK is the most extensively studied Hsp70 family member (Dahiya and Buchner, 2019). DnaK shares about 50% identical amino acids with eukaryotic Hsp70s (Daugaard et al., 2007). In eukaryotic cells, there are two different types of Hsp70, which are constitutively expressed Hsp70 (Hsc70) and stress inducible Hsp70s (Dahiya and Buchner, 2019). Additionally, Hsp70 homologues are found in mitochondria (mortalin/HSPA9) and in the endoplasmic reticulum (ER) (BiP/HSPA5) (Dahiya and Buchner, 2019; Kampinga and Craig, 2010). Overall, human cells express 13 Hsp70 homologues (in cytosol, nucleus, ER and mitochondria) (Rosenzweig et al., 2019). Mutations in Hsp70s are associated with several human diseases.

E. coli DnaK has been shown to interact with ~700 different proteins (15-20% of *E. coli* proteome), which are mostly multi-domain proteins (Calloni et al., 2012). Deletion of TF leads to an increase in the DnaK interactome to ~1000 proteins (Calloni et al., 2012).

2.7.2.1 The structure of Hsp70

The Hsp70 chaperone members are typically ~ 640-650 residues in length, showing a high degree of amino acid sequence similarity of ~50% even in distantly related organisms (prokaryotes and

eukaryotes) (Flaherty et al., 1990). Hsp70s consist of two highly conserved functional domains: a ~44-kDa N-terminal nucleotide binding domain (NBD) (1-383) and a ~25-kDa C-terminal substrate-binding domain (SBD) (396-638), connected by a conserved hydrophobic linker (384-395) (Bertelsen et al., 2009) (Figure 2.6). The NBD is a member of the sugar kinase superfamily, which includes hexokinase and also the cytoskeletal protein actin (Bauer et al., 2018), sharing the same fold topology, even though the overall sequence similarity is low. The NBD is composed of two lobes (I and II) which are separated by the ATP binding core, where nucleotides and ions (Mg^{2+}, K^{+}) bind. Each lobe has two subdomains (A and B), with the four subdomains being IA (1-39, 116-118), IB (40-115), IIA (189-228, 307-385), and IIB (229-306) (Bertelsen et al., 2009; Flaherty et al., 1990). The SBD consists of two subdomains, a twisted β-sandwich subdomain (SBDβ) (396-502), where peptide substrate binds, and an α-helical subdomain (SBDα) (503-638), which acts like a lid and closes over the bound substrate. SBDα consist of 5 α-helices (αA, αB, αC, αD, αE). Once the substrate peptide binds to the SBD, SBDα closes over the SBDβ to form the substrate binding cavity, through hydrophobic contacts, salt bridges and hydrogen bonds (Zhu et al., 1996). The SBDα ends with a conserved ~ 30 residue unstructured region. Even though SBDα does not directly contact the bound peptide, DnaK with a partially deleted lid cannot efficiently refold denatured firefly luciferase (Mayer et al., 2000b). Additionally, in many cytosolic and nuclear Hsp70s in eukaryotes, the C-terminal unstructured region ends with a conserved charged motif (Glu-Glu-Val-Asp; EEVD) that mediates the interaction with the co-chaperones CHIP, HIP and HOP by binding to the tetratricopeptide repeat (TPR) domains of these proteins (Rosenzweig et al., 2019). Although the bacterial Hsp70, DnaK, does not contain an EEVD motif, mutation of the C-terminal region leads to impairment of protein folding activity without changing the co-chaperone, peptide binding affinity and allosteric mechanism of DnaK (Smock et al., 2011). It was proposed that this region has a secondary substrate binding site, thereby keeping the substrate in the vicinity of the protein at the high local concentrations in the cell, which facilitates repeated cycles of binding and release of proteins to the SBD for efficient folding (Smock et al., 2011). SBDβ is composed of an eight-stranded β-sandwich containing the substrate binding cavity in the central hydrophobic pocket (Kityk et al., 2015). The crystal structure of the SBD of DnaK complexed with a model peptide (NRLLLTG) indicated that the peptide interacts with β-strand 1, β-strand 2 including $L_{1,2}$ and $L_{3,4}$ via several Van der Waals contacts, hydrogen bonds and hydrophobic interactions (Zhu et al., 1996). NBD and SBD are connected

with a conserved hydrophobic linker that mediates the inter-domain communication and couples ATP binding and hydrolysis to the conformational change of the SBD (Swain et al., 2007).

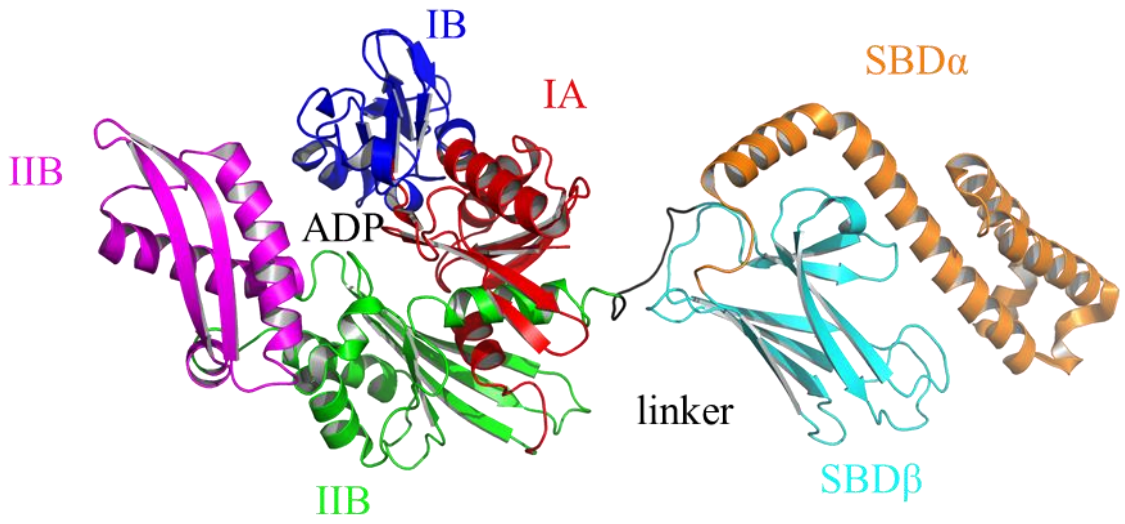


Figure 2.6 Structure of bacterial Hsp70
Structure of ADP-bound DnaK. Subdomains IA (red), IB (blue), IIA (magenta), IIB (green) are shown. The NBD is connected to the SBD via a conserved hydrophobic linker (black). The SBD is composed of SBDβ (cyan), containing the substrate binding site and SBDα lid segment (orange) (PDB: 2KHO).

2.7.2.2 Molecular mechanism of Hsp70

The chaperone function of Hsp70 involves an allosteric communication between the NBD and the SBD (Mayer, 2013) that is best understood for DnaK. The regulation of affinity for a substrate is provided by ATP binding and its hydrolysis (Mayer, 2013). Therefore, DnaK exists at least in two states which are called the “open” and a “closed” conformations (Mayer, 2013). By comparison with the ADP bound state where DnaK is in the “closed” conformation, ATP binding to the NBD results in a ~20° rotation of subdomain IB towards the subdomain IIB (Bhattacharya et al., 2009). This rotation leads to the opening of the linker binding cleft between subdomain IA and IIA. When the linker binds to the NBD, the two domains dock and SBDα detaches from SBDβ, thus opening the substrate binding cleft. The interaction between NBD and SBD results in overall conformational changes on the subdomains and this docked conformation is called the “open” conformation (Kityk et al., 2015). The formation of the “open” state leads to an increase in the

peptide dissociation rate (~1000-fold) and association rate (~100-fold), resulting in low affinity for peptide substrate. Thus, the overall peptide binding affinity of DnaK decreases 10 to 50-fold in the ATP bound state (Mayer and Gierasch, 2019; Mayer et al., 2000b; Schmid et al., 1994). When a substrate binds to the SBD in the presence of ATP, conformational changes resulting from binding to the SBD cause the NBD to adopt a conformation for efficient ATP hydrolysis. Substrate binding to the SBD triggers the dissociation of SBDβ from the NBD, which results in a rotation of the NBD subdomains. Thus, this conformational change stimulates ATP hydrolysis ~4-fold (Kityk et al., 2018; Kityk et al., 2015). Upon ATP hydrolysis to ADP, the two domains separate and SBDα closes over the substrate, increasing the substrate binding affinity by ~400-fold (Mayer, 2013) (Figure 2.7).

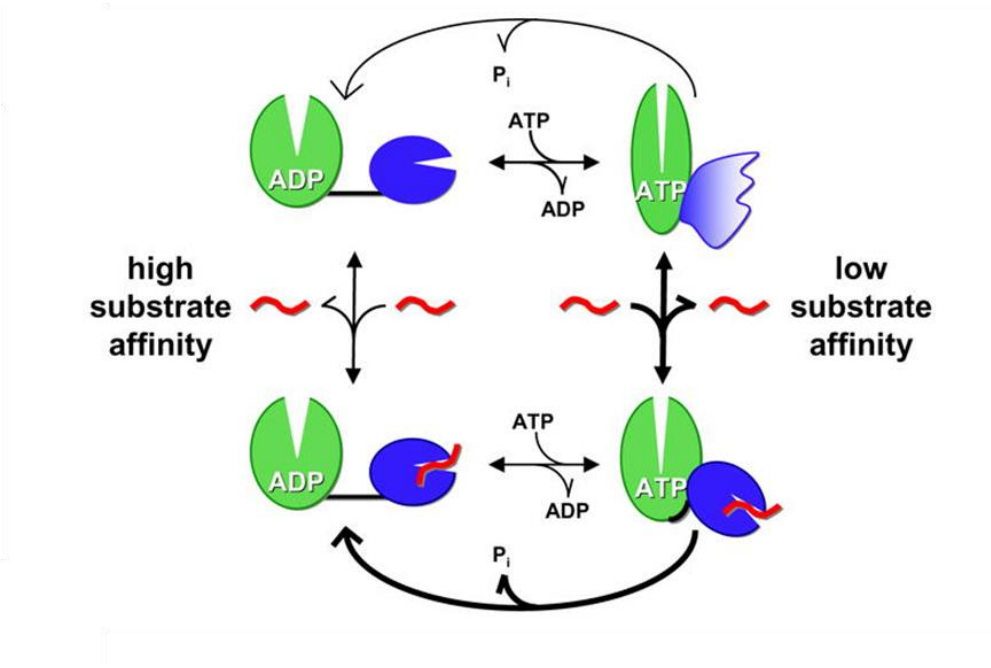


Figure 2.7 Structure of bacterial Hsp70
Two domains (green: NBD; blue: SBD) act independently connected by a flexible hydrophobic linker (black) in the ADP-bound state. ATP binding results in the docking of the two domains, followed by conformational changes, which reduce the substrate (red) binding affinity of the SBD. Substrate binding to the SBD transmits a signal to the NBD via the hydrophobic linker that increases the ATPase rate. Hydrolysis of ATP to ADP stimulates the separation of the domains. Modified from (Swain et al., 2007).

Hsp70s do not work alone in protein folding. The activity of Hsp70 is regulated by two co-chaperones which are Hsp40s, J-domain proteins (DnaJ in *E. coli*) and nucleotide exchange factors (GrpE in *E. coli*). In the current model of DnaK protein folding cycle, DnaJ first binds to the unfolded substrate protein and transfers it to DnaK in the ATP state, followed by the formation of a transient ternary complex among DnaK, DnaJ and the substrate protein (Han and Christen, 2003; Karzai and McMacken, 1996; Laufen et al., 1999). DnaJ, together with the substrate protein stimulate the ATPase activity of DnaK up to 1000-fold (Barouch et al., 1997; Karzai and McMacken, 1996; Laufen et al., 1999; Mayer, 2013; Misselwitz et al., 1998; Silberg et al., 2004) (Mayer, 2013). This generates DnaK in the ADP state with loosely associated NBD and SBD and the alpha helical lid being closed, resulting in substrate binding to DnaK with high affinity. Binding of the nucleotide exchange factor GrpE to the NBD of DnaK then stimulates ADP release, which leads to the opening of the alpha helical lid, allowing substrate release (Figure 2.8).

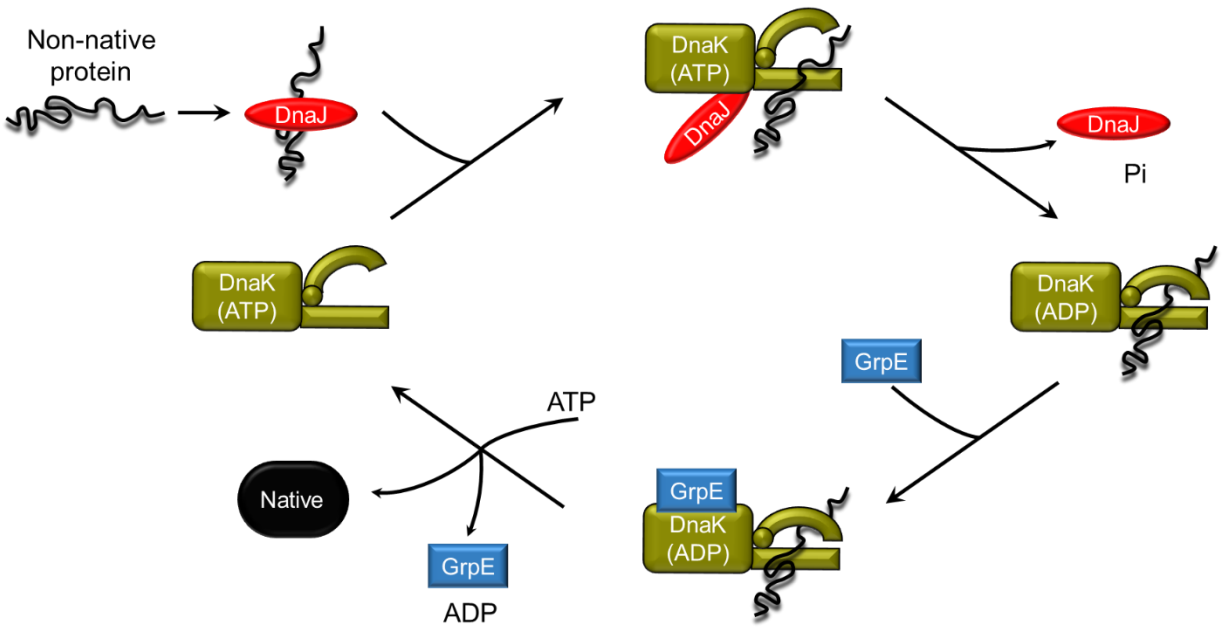


Figure 2.8 The bacterial Hsp70 system (DnaK/DnaJ/GrpE) reaction cycle
DnaJ (red) first binds to unfolded substrate protein (black) and transfers it to ATP-bound DnaK (yellow), followed by transient complex formation of DnaK, DnaJ and substrate protein. DnaJ and substrate protein synergistically trigger ATP hydrolysis by DnaK, leading to the formation of a stable complex between the substrate protein and DnaK. GrpE (blue) catalyses ADP-ATP exchange. ATP re-binding to DnaK stimulates the release of bound substrate.

2.7.2.3 The co-chaperones of Hsp70

Hsp70 cooperates with co-chaperones to fulfil its biological roles. The most important co-chaperones regulating the folding activity are J-domain proteins (Hsp40) and nucleotide exchange factors (NEF).

2.7.2.3.1 Hsp40

The Hsp40 chaperones comprise a diverse family of ATP-independent chaperones with 41 members in human (Kampinga and Craig, 2010) compared to 13 Hsp70s. These Hsp70s can interact with different Hsp40s, so it has been assumed that different classes of Hsp40s regulate the functional diversity of Hsp70s by recruiting them to different cellular contexts (Dahiya and Buchner, 2019; Kampinga and Craig, 2010). The main role of J-domain proteins is to stimulate ATPase activity of Hsp70. They are characterized by a highly conserved ~70 amino acid J domain with four alpha helices containing a conserved His, Pro, Asp (HPD) three amino acid region between the two main helices (II-III) (Kampinga and Craig, 2010). This HPD motif of Hsp40 stimulates the ATPase activity of Hsp70 several fold. Hsp40s are divided into classes I, II and III. Class I Hsp40s consist of a N-terminal J domain, followed by a Gly-Phe (G/F) rich domain, a cysteine rich zinc finger motif and a C-terminal domain where substrate proteins bind. Hsp40s typically function as dimers, held together by interactions at the extreme C terminus. Class II Hsp40s lack the zinc-finger domain. All other Hsp40s are assigned to class III (Kampinga and Craig, 2010). There are large structural and functional diversities among Hsp40s and some Hsp40s may not bind to substrate protein (Figure 2.9).

The bacterial Hsp40 (DnaJ) acts as a holdase; it binds to a motif consisting of eight hydrophobic residues enriched in aromatic and large aliphatic residues and arginine to prevent aggregation (Rudiger et al., 2001). A similar preference for binding to substrate was also observed for other Hsp40s such as Ydj1, Sis1 (yeast Hsp40s) and DNAJA1 (a human Hsp40) (Fan et al., 2004; Terada and Oike, 2010). Even though DnaK binding motifs are not the same as DnaJ binding motifs, most of the binding sites are shared (Rudiger et al., 2001) . Interestingly, Hsp70s cannot bind to proteins properly in the absence of Hsp40 (Kellner et al., 2014; Szabo et al., 1994). Although many studies have shown that Hsp40s bind and prevent the aggregation of substrate proteins, how Hsp70s effect the conformation of the substrate protein remains elusive. Hsp40s also play a critical role in preventing toxic protein aggregation. For example, DNAJB6 and DNAJB8

efficiently prevent the aggregation polyQ proteins like huntingtin and of amyloid-β peptide (Aβ42) (Gillis et al., 2013; Rosenzweig et al., 2019).

Biochemical and biophysical studies have defined the critical residues of Hsp40 that associate with the NBD and the linker region of Hsp70 (Jiang et al., 2007; Kumar et al., 2011; Suh et al., 1998). Interestingly, the recent crystal structure of the J domain of bacterial Hsp40 in a complex with ATP-bound Hsp70 indicated that Hsp40 directly interacts with the NBD to stimulate the ATPase activity of Hsp70, thereby stimulating substrate handover from Hsp40 to Hsp70 (Kityk et al., 2018). Additionally, some eukaryotic Hsp40s also interact with the conserved C- terminal EEVD motif of Hsp70. A study showed that eukaryotic Hsp70 with deleted EEVD motif cannot fold firefly luciferase, because of the inefficient interaction between the Hsp70 and Hsp40 (Yu et al., 2015).

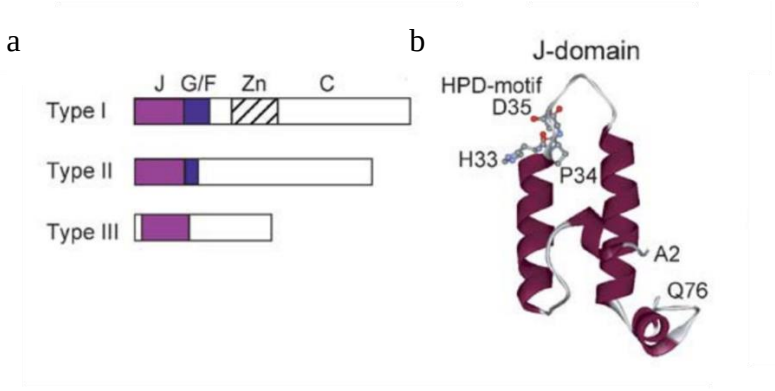


Figure 2.9 Classes of J domain proteins
Domain structures of different classes of J domain proteins are shown. (a) Magenta, purple and dashed line are J domain, G/F rich domain and Zn^{+2} binding domain, respectively. (b) The structure of the J domain with the conserved HPD motif is indicated (PDB:1XBL). Modified from (Mayer and Bukau, 2005).

2.7.2.3.2 Nucleotide exchange factors

Another class of Hsp70 co-chaperone are the nucleotide exchange factors (NEFs), which are mainly responsible for nucleotide release from Hsp70s. NEFs facilitate the release of ADP by opening the nucleotide binding cleft (Rosenzweig et al., 2019). 11 NEFs exist in human cells. They are categorized into four different types (GrpE-type, Hsp110-type, Bag-type, Armadillo-

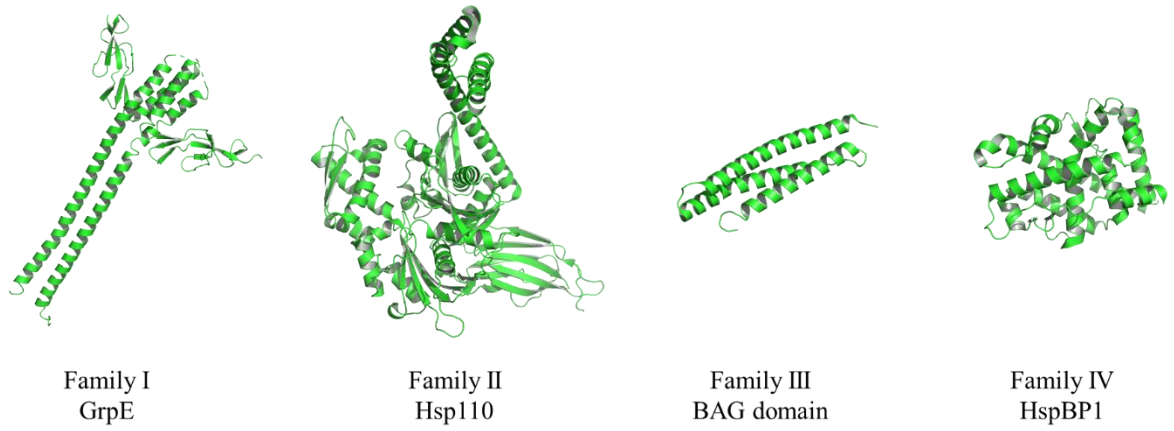
type) (Figure 2.10a). These families share no sequence similarity and interact with the NBD of Hsp70 in different ways. For the *E. coli* homologue, DnaK, ADP dissociation is strongly stimulated up to 5,000-fold by GrpE (Packschies et al., 1997). The GrpE-type NEFs are homodimers found in bacteria, mitochondria and chloroplasts. The structure of bacterial NEFs from different organisms were solved (Harrison et al., 1997; Nakamura et al., 2010; Wu et al., 2012). All bacterial GrpE-type NEFs consist of a long N-terminal α helical dimerization domain that has unstructured hydrophobic and charged residues at the extreme N-terminus, and of a C-terminal β-sheet domain (Harrison et al., 1997) (Figure 2.10a). The long α-helices from a parallel dimer and the C-terminal β-sheet domain of one GrpE monomer binds to the NBD of DnaK, triggering a tilt of subdomain IIB to facilitate ADP-release (Harrison et al., 1997; Melero et al., 2015; Rosenzweig et al., 2019) (Figure 2.10b). Interestingly studies suggested that the extreme N-terminal residues (1-33) act as a pseudo substrate for DnaK to displace bound substrate from the DnaK SBD. Hence, it was observed that GrpE (Δ1-33) cannot efficiently fold the denatured firefly luciferase *in vitro* (Melero et al., 2015). Additionally, similar to the bacterial Hsp70 system, pseudo substrate action was observed for the yeast NEF, Bap (Rosam et al., 2018). Together, GrpE regulates the release of both nucleotide and the substrate.

The eukaryotic Hsp110 (in the cytosol)/Grp170 (in the ER-lumen)-type NEFs are atypical Hsp70 protein that, like Hsp70, consist of an NBD and an SBD (Figure 2.10a). The only difference relative to the NBD of Hsp70 is the presence of an additional segment of variable length in the β-sandwich domain and an insertion in the C-terminal helical lid domain of the SBD (Rosenzweig et al., 2019). In addition to functioning as a NEF, they have holdase activity and thus can bind non-native proteins to prevent aggregation. Although they bind to substrate proteins and ATP, they have no folding activity. Human cells have three homologues of Hsp110: Hsp105, Apg1 and Apg2. Sse1 and Sse2 are the yeast homologues of Hsp110.

Another family of NEFs are the Bcl-2-associated athanogene (BAG) domain proteins, containing a ~85 amino acid BAG domain (Figure 2.10a). There are six homologues in human, BAG1 to BAG6. Snl1 is the BAG homologue in yeast. It is the most complex NEF family, and is currently being studied intensively. BAG proteins have a role in protein degradation and autophagy beyond catalyzing nucleotide exchange on Hsp70 (Kampinga and Craig, 2010).

Armodillo-type NEFs are found in the human cytosol (HSPBP1) and the ER (SIL1) (Figure 2.10a). Fes1 and Sil1 are cytosolic and ER Armodillo-type NEF homologue in yeast, respectively. In addition to ADP release stimulation, Armodillo-type NEFs can inhibit rebinding of substrate protein to Hsp70 by occupying the SBD of Hsp70s.

a



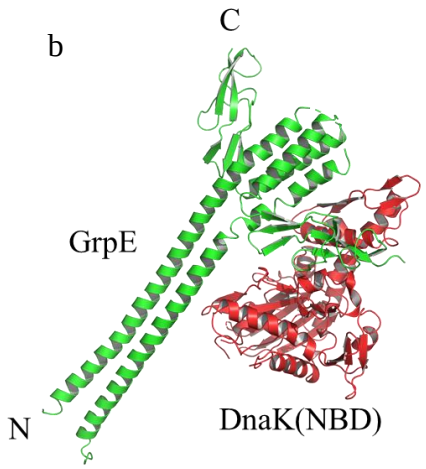


Figure 2.10 Structure of four classes of nucleotide exchange factors (NEFs)
(a) GrpE, the NEF of DnaK, consisting of a long α-helical domain forming a dimer (PDB: 1DKG). Sse1 (yeast homologue of Hsp110) is structurally very similar to Hsp70, containing an NBD and an SBD (PDB: 3D2F). The BAG domain consists of three α-helices and exists in all BAG-domain proteins except BAG5 (PDB: 1HXI). The core domain of HspBP1 is all α-helical (PDB: 1XQS). (b) Dimeric GrpE bound to the NBD of DnaK (PDB: 1DKG). GrpE monomers and the NBD of DnaK are coloured green and red, respectively. The N and the C are the N-terminal domain and the C-terminal domain of GrpE, respectively.

2.7.2.4 The functions of Hsp70s

Hsp70s function in de novo protein folding, disaggregation, protein translocation and the (dis-)assembly of protein complexes.

2.7.2.4.1 The effect of Hsp70 on their substrate proteins and Hsp70 folding mechanisms

The Hsp70s are a versatile class of chaperones that interact with a large number of substrates, including nascent polypeptide chains, folding intermediates, misfolded proteins and nearly native proteins (Rosenzweig et al., 2019). This great versatility of Hsp70s is due to their recognition of exposed hydrophobic regions of non-native proteins (Mayer, 2013). Hsp70 binding sites occur in proteins on average every 36 residues (Rudiger et al., 1997). Such sites are mostly buried in the native state, allowing Hsp70s to bind to non-native proteins but not to their native counterparts.

Multiple studies have shown that bacterial and eukaryotic Hsp70s can refold both heat- and chemically denatured substrates (Moran Luengo et al., 2018; Rampelt et al., 2012; Schröder et al., 1993b; Szabo et al., 1994). According to these studies, Hsp70s recognize exposed hydrophobic residues of non-native proteins in ATP-dependent manner, thereby preventing off-pathway aggregation of the client protein. Binding of Hsp70s to non-native proteins reduces the ability of the protein to aggregate. Release of Hsp70 from the hydrophobic regions allows folding to proceed. Substrate proteins reach their native state spontaneously when they are released from Hsp70s. Rebinding of Hsp70 occurs when the released substrate protein does not reach the native state. Thus, ATP-regulated binding and release of Hsp70 prevents aggregation and promotes protein folding. By contrast, some recent studies suggest that Hsp70 induces unfolding of substrate proteins, thus helping them overcome kinetic barriers (De Los Rios et al., 2006; Palleros et al., 1994; Rodriguez et al., 2008; Sharma et al., 2010; Sousa et al., 2016). This view suggests that Hsp70s use the energy of ATP to destabilize the kinetically trapped intermediates, misfolded proteins and even aggregated proteins (Goloubinoff et al., 2018). In support of Hsp70s possessing unfoldase activity, a study showed that the Hsp70 inactivates GR through partial unfolding, which facilitates delivery of GR to Hsp90. (Kirschke et al., 2014). Additionally, recent biochemical and biophysical studies observed that Hsp70-bound substrates exhibit more unfolded conformations without long-range, tertiary structures (Kellner et al., 2014; Sekhar et al., 2015, 2016;

Sharma et al., 2010). How exactly substrate binding and release by Hsp70 leads to productive folding is not yet understood, however (Imamoglu et al., 2020).

2.7.2.4.2 Hsp70 in disaggregation and complex disassembly

A role of Hsp70 in disaggregation was first shown for the bacterial homologue of Hsp70, DnaK (Diamant et al., 2000). The disaggregation capacity of Hsp70 is improved by the recruitment ring-shaped hexameric AAA+ disaggregases of the Hsp100 class, ClpB in bacteria and Hsp104 in yeast. These chaperones can unfold proteins by pulling the sequence through their central pore (Mogk et al., 2015). Hsp70 binding to the surface of a protein aggregate, mediated by J-domain proteins, recruits Hsp100 (Mogk et al., 2015; Nillegoda and Bukau, 2015). Additionally, Hsp70 regulates the ATPase cycle of Hsp100 via direct interaction with the middle domain of Hsp100. The threading action of Hsp100 disentangles polypeptide chains, trapped in the aggregate (Mogk et al., 2015).

In contrast to bacteria, fungi and plants, metazoans lack Hsp100-type disaggregases. Instead, Hsp70, in cooperation with a specific J-domain protein and Hsp110 isoforms of NEFs form a Hsp70-dependent protein disaggregation machine. Three isoforms (Hsp105, Apg1, Apg2) of Hsp110 in human play a crucial role in disaggregation *in vitro* (Rampelt et al., 2012; Yamagishi et al., 2011). Other NEFs of Hsp70 cannot function as part of the disaggregation machinery. The reason for the specific function of Hsp110 NEFs in disaggregation is not clear, but it has been proposed that their ability to bind peptide segments enriched in aromatic residue play a role. J-domain proteins bind first to aggregated proteins and determine the substrate specificity of Hsp70-based disaggregation (Nillegoda and Bukau, 2015). Class A J-proteins target only small aggregates (Mattoo et al., 2013; Nillegoda and Bukau, 2015), whereas class B J-proteins are involved in the solubilization of large aggregates (Rampelt et al., 2012).

Related to the role in disaggregation is the function of the Hsp70 system in the regulatory disassembly of specific protein complexes. The Hsp70 system participates in several different disassembly process of oligomeric protein complexes such as the disassembly of the replication complex of bacteriophage λ (Zylicz et al., 1989) and virus capsid during penetrating a host membrane (Ravindran et al., 2015) , despite the mostly studied role of the Hsp70 system on uncoating clathrin-coated vesicles (Eisenberg and Greene, 2007; Rapoport et al., 2008; Sousa and Lafer, 2015; Sousa et al., 2016; Ungewickell et al., 1995). During endocytosis, cytosolic clathrin trimers,

which is a coat protein, surrounds the vesicle by forming a polyhedral lattice and assisting vesicle budding for intracellular trafficking (Rosenzweig et al., 2019). After clathrin-mediated intracellular trafficking occurs, the heat-shock cognate protein (Hsc70) chaperone, together with its co-chaperone auxilin (DNAJC6) catalyze the disassembly of clathrin-coated vesicle (Sousa and Lafer, 2015) by recognizing the QLMLT motif clathrin (Rapoport et al., 2008).

2.7.2.4.3 Hsp70 in protein translocation

Most of mitochondrial and chloroplast proteins are synthesized on cytoplasmic ribosomes in the cytosol. Targeting signals at the N-terminus of these proteins guide them to the respective organelles. Translocation occurs through specific translocon machineries and requires that the protein to be transported be stabilized in the cytosol in an unfolded state by Hsp70 binding. Organellar Hsp70 proteins receive the translocating chain on the trans side of the translocon and assist in the translocation process. These include the Hsp70 of mitochondrial matrix, the chloroplast stroma and the endoplasmic reticulum lumen. An entropic pulling model was proposed as a mechanism to explain how Hsp70s drive translocation of polypeptides through translocons (Goloubinoff and De Los Rios, 2007; Mayer, 2013). According to this model, Hsp70 bound to polypeptide close to the trasnlocon has restricted diffusional mobility that leads to a state of low entropy. Brownian motion of the peptide toward the inside of the organelle increases the distance of Hsp70-bound peptide to the membrane (Mayer, 2013). This leads to greater conformational freedom of Hsp70, thus increasing the entropy of the system. Therefore, an increase in entropy creates the driving force for the import as a consequence of the second law of thermodynamics (Goloubinoff and De Los Rios, 2007; Mayer, 2013).

2.7.3 The Hsp60 molecular chaperone (Chaperonins)

The Hsp60 chaperones (chaperonins) are nano-cages for a single protein to fold unimpaired by aggregation (Balchin et al., 2016). They are large double-ring complexes of ~800 kDa formed by 7-9 subunits of ~60 kDa. Two groups of related chaperonins can be distinguished. The group I chaperonins are composed of 7 same subunits per ring and exist in bacteria (GroEL), mitochondria (Hsp60) and chloroplasts (Cpn60). They cooperate with lid-shaped co-chaperones (GroES in bacteria, Hsp10 in mitochondria, and Cpn10/Cpn20 in chloroplasts) (Balchin et al., 2016; Horwich et al., 2007; Kim et al., 2013). The group II chaperonins are present in archea (thermosome) and the eukaryotic cytosol (TRiC, also known as CCT). In contrast to group I chaperonins,

the group II chaperonins consist of eight or nine subunits per ring and they do not have an additional lid-shaped co-chaperone. TRiC is involved in folding of some essential proteins such as actin and tubulin (Horwich et al., 2007; Kim et al., 2013). Interestingly, although the group I and group II chaperonins are similar in structure, their substrates are not interchangeable (Balchin et al., 2016). For example, a recent study indicated that GroEL can bind and encapsulate unfolded actin but cannot mediate folding. TRiC efficiently folds actin by taking advantage of subunit-specific binding and an asymmetric conformational cycle (Balchin et al., 2018).

GroEL is essential for *E. coli* at all temperatures (Fayet et al., 1989; Georgopoulos, 2006). GroEL participates in folding 10% of *E. coli* proteome, particularly for proteins with complex domain topologies such as the TIM barrel fold, which are often obligate GroEL substrates. Notably, these proteins cannot be folded by the binding and release cycle of Hsp70, although Hsp70 binding stabilize them against aggregation and transfers them to the downstream chaperonin, GroEL (Kellner et al., 2014; Sekhar et al., 2015, 2016). GroEL substrates can encapsulate proteins up to ~60 kDa, which can fit into the nano-cage of GroEL.

GroEL consists of two homo-heptameric ring which are stacked back to back. The subunits of GroEL consist of three domains: an equatorial ATPase domain, an intermediate hinge domain and an apical domain (Figure 2.11a). The apical domain is located at the entrance to the GroEL cavity and mediates substrate binding by capturing the exposed hydrophobic regions of substrates. The intermediate domain mediates nucleotide dependent conformational changes from the equatorial domain to the apical domain. The folding cycle of GroEL is regulated by GroES and ATP, with the two rings of GroEL binding and releasing GroES sequentially. The free ring of the GroEL-GroES complex (the so-called trans ring) binds the substrate protein, resulting in a conformational expansion of the collapsed substrate protein (Lin et al., 2008; Sharma et al., 2008). Subsequent ATP binding to the trans-ring causes the dissociation of GroES from the opposite ring (the cis-ring) through a long-range, negatively cooperative allosteric action. GroES binding to the trans-ring then forms a new cis-ring with encapsulated substrate protein. This step of the process is associated with major conformational changes in the cis-ring that form a hydrophilic, negatively charged folding cage. Substrate proteins up to ~60 kDa in size are free to fold inside this cage for the time required to hydrolyse the 7 ATP in the GroEL ring (Singh et al., 2020). ATP binding to the trans-ring then causes the unbinding of GroES and opens the cage, whereupon folded substrate protein leaves. Incompletely folded protein rebinds and goes through another

folding cycle (Figure 2.11b). How exactly GroEL mediates protein folding is a longstanding debate. Three mechanisms have been proposed (i) the passive cage model, (ii) the active cage model, and (iii) iterative annealing model. The passive cage (also known as "Anfinsen cage") model suggests that GroEL creates an isolated chamber for folding in which aggregation is prevented. Consequently, the protein folds spontaneously as under conditions of infinite dilution *in vitro*. This model predicts that the rate folding inside and outside the cage are identical as long as aggregation dose not limit the folding rate in the spontaneous reaction. Scientists who support the passive cage model claim that the apparent acceleration of protein folding by GroEL results from the prevention of reversible-aggregation (Ellis, 1994; Ellis and Hartl, 1996; Horwich et al., 2009). The iterative annealing model suggests that repeated cycles of substrate binding and release are required to drive folding (Thirumalai and Lorimer, 2001; Yang et al., 2013). Unfolding removes kinetic traps and release allows the protein to partition between a productive folding trajectory to the native state and re-forming the kinetically trapped, misfolded intermediate. Consequently, successive cycles of substrate binding, unfolding and release for folding inside or outside the cage must occur for folding with high yield. In contrast, the active cage model posits that the physical environment of the GroEL cavity promotes folding (Brinker et al., 2001; Chakraborty et al., 2010; Tang et al., 2008; Tang et al., 2006). An effect of steric confinement of folding intermediates in the cage is believed to shorten the search time for native intramolecular contacts for proteins with complex topologies. The development of biophysical techniques provides scientists with the ability to analyse the role of GroEL in the absence of transient aggregation. Recent studies using single-molecule techniques have shown that GroEL significantly accelerates the folding of proteins (Georgescauld et al., 2014; Gupta et al., 2014; Weaver et al., 2017). However, it is still not clear how GroEL accelerates the folding of substrates (Balchin et al., 2016).

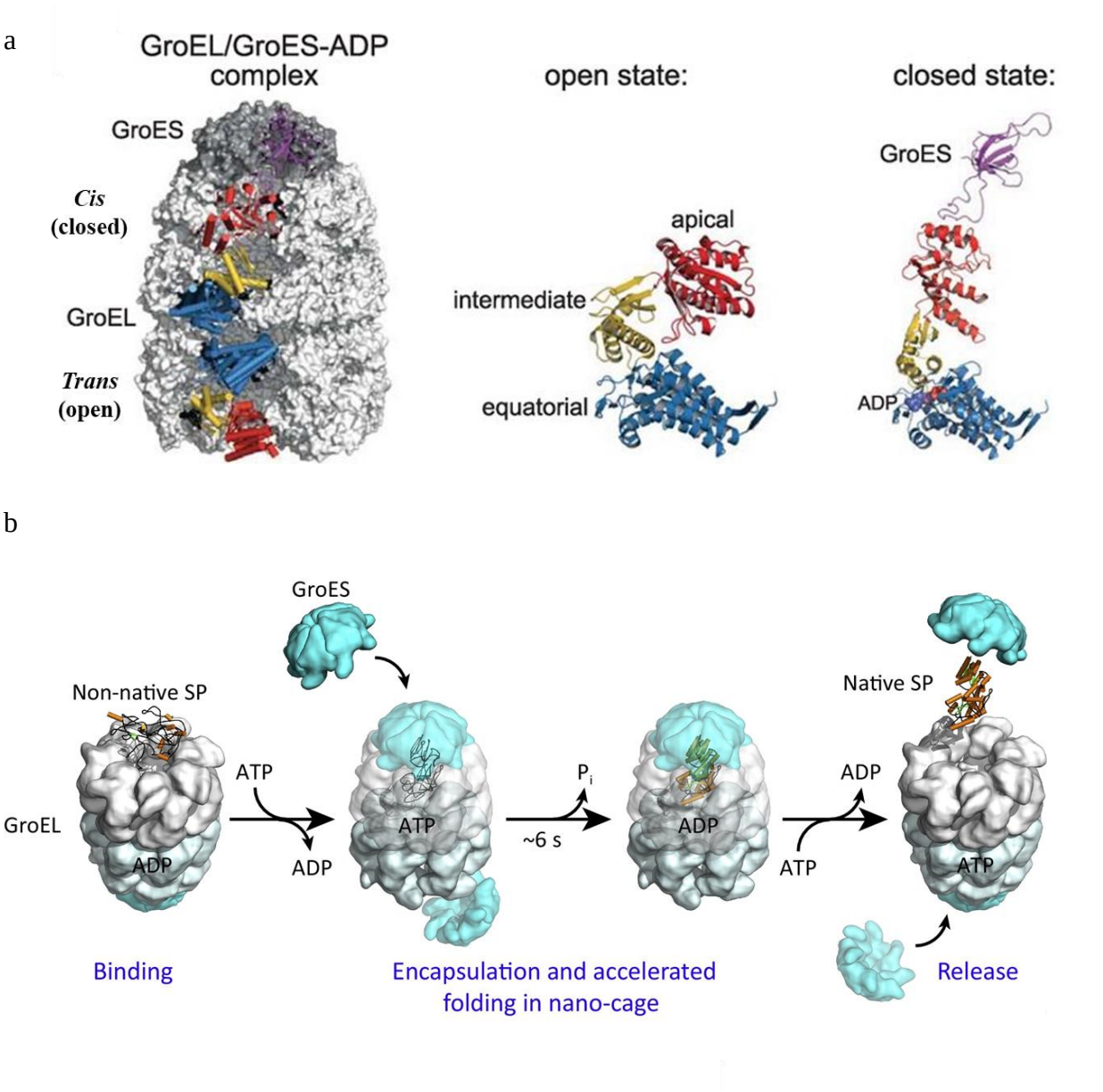


Figure 2.11 The structure and GroEL reaction cycle

(a) The structure of GroEL/ES was shown (PDB:1PFQ). One subunit of a ring was displayed for both open and GroES-bound closed conformation with the equatorial domain (blue), the intermediate domain (yellow) and the apical domain (red). Modified from (Balchin et al., 2016). (b) Hsp70 system or TF released non-native substrate protein (SP) are delivered to GroEL. GroEL reaction cycle is initiated by binding of non-native SP to the open ring of GroEL. The release of SP onto a nano-cage cavity is triggered by the binding of ATP and GroES. During the time required for ATP hydrolysis, the encapsulated proteins reach its native conformation. The folding of substrate is actively promoted by GroEL cavity. Binding of GroES and ATP to the opposite ring results in the cage opening that stimulates the release of SP. Modified from (Hayer-Hartl et al., 2016)

2.8 Fluorescence

Fluorescence is a phenomenon of the emission of light from an excited molecule. The molecular absorption of a photon results in an electronically excited state. The excited molecule releases its energy in the form of light to return to a lower energy state. There are three forms of relaxation processes, which are internal conversion (IC), intersystem crossing (ISC), and fluorescence (Figure 2.12).

The relationship between absorption and emission of light can be illustrated by the Jablonski diagram. The electronica ground state (S_0) is the lowest energy state. When a photon of light is absorbed, electrons are excited to a higher energy state (S_1, S_2, S_N). Electronic energy levels of each state consist of a number of vibrational energy levels. After excitation, a molecule is usually excited to a higher vibration level of S_N, whereas the electron returns to the ground state from the lowest energy level of the excited state. This is the reason for the emission of a photon with a longer (less energetic) wavelength than the absorbed photon. Thus, the energy difference between the absorbed and emitted photon is referred to as Stokes shift. (Absorption and emission occur mostly from molecules with the lowest vibrational energy level).

Internal conversion (IC), and intersystem crossing (ISC) are two other processes of the conversion of an excited molecule into the ground state. IC is a non-radioactive decay where the energy of the excited molecule is released to the environment as heat. Molecules in the excited state can transit from a singlet to triplet state called intersystem crossing (ISC) that is forbidden. Emission form the triplet state, termed phosphorescence, occurs at much slower rate than fluorescence.

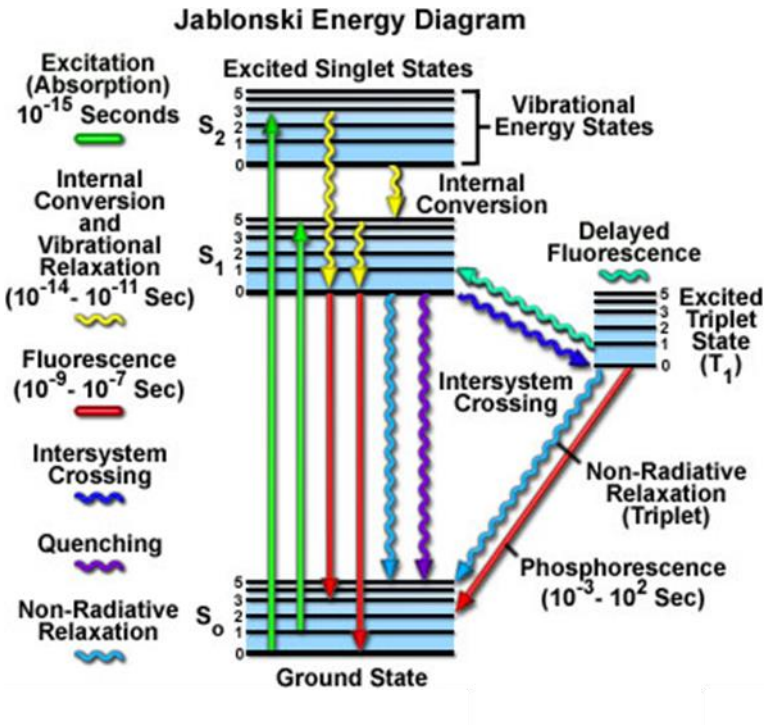


Figure 2.12 Jablonski diagram indicating the possible radiative and non-radiative transitions

Absorption by a molecule occurs when light with a suitable energy interacts with the molecule in the ground state (S_0). Excited photons return to the ground state by either radiative or non-radiative decay processes. Possible relaxation processes of an excited photon by the absorption are fluorescence, internal conversion and intersystem crossing into the triplet state. Straight lines represent radiative transitions while undulated lines non-radiative transitions. Available at https://www.olympus-lifescience.com/en/microscope-resource/primer/java/jablonski/jabintro/

2.8.1 Fluorescence/ Förster resonance energy transfer (FRET)

Fluorescence Resonance Energy Transfer or Förster Resonance Energy Transfer (FRET) is a well-established technique to estimate the distance between two molecules or different residues of a molecule in close proximity (over the range 1-10 nm). The radiationless energy transfer from an excited molecule to an acceptor molecule occurs via dipole-dipole interaction. So, the main requirement for a FRET experiment is to employ correct dye pairs that have an overlap between the fluorescence emission spectrum of the donor and the absorption spectrum of the acceptor. The rate of energy transfer is proportional to the inverse sixth power of the distance between the dyes, thus it is a very sensitive tool to measure small changes in distance.

The FRET efficiency (f_E) quantitively expressed by the following equation:

$$f_E = \frac{1}{1 + (r/R_0)^6}$$

where R_0 describes a characteristic distance where the FRET efficiency is 50% between dye pairs and the r is the distance between dyes.

In biology, FRET is a widely used technique and has many applications such as the detection of DNA hybridization and interaction of proteins both in *in vivo* and *in vitro*. Importantly, FRET can be used in protein folding studies since it provides information about the conformational changes between dye pairs attached to specific amino acid residues.

2.8.2 Fluorescence-based single-molecule studies in protein folding

Over the last 20 years, the development of single-molecule techniques provided us with the ability to study the properties of a single individual molecule with high resolution in heterogenic mixtures. The most widely used tool for detection at the single-molecule level is fluorescence. Individual fluorescently labelled molecule can be detected in a very small volume. Therefore, fluorescence based single-molecule studies revealed interesting information about the dynamics and kinetics of proteins.

Classical biochemical and biophysical experiments involve billions of analysed molecules and the properties of the molecules are averaged (Mashaghi et al., 2014a). Despite having high signal-to-noise ratio, much information is lost due to averaging (Schuler, 2007). Averaging of molecules prevents the analysis of individual molecules. Thus, the individual behavior of a single-molecule in a heterogeneous mixture can be investigated by performing single-molecule experiments. The challenging low signal-to-noise ratio in single-molecule experiments has been improved in recent decades by technical advances in microscopy setups.

Single-molecule experiments are particularly important in protein folding for two aspects. First, the conformation of folding intermediates, chaperone bound states and the native state can be detected in the folding reaction by performing single-molecule experiments. Second, single-molecule experiments are performed at highly diluted concentrations in solution (typically <100 pM). Under these diluted concentrations, the role of chaperones can be tested in the absence of aggregation that can otherwise lead to a bias in folding experiments. Single-molecule experiments perform either with the freely diffusing labelled molecules or with molecules immobilized on a

surface. Although several protein folding studies have been performed using surface immobilized proteins, protein activity and folding pathway can be negatively affected by the surface attachment. Thus, analysing freely diffusing labelled proteins is of great advantage in protein folding studies. Two major classes of fluorescence spectroscopy techniques are fluorescence correlation-spectroscopy (FCS) and single-molecule FRET (smFRET).

FCS is a well- established single-molecule technique for the sensitive detection of diffusion time (Elson, 2011). In contrast to other fluorescence techniques, the intensity of fluorescence emission is not the parameter of interest in FCS. Instead, FCS obtains information from the fluctuations in fluorescence intensity. Fluorescence intensity fluctuations are analysed as fluorescently labelled particles diffuse in or out of a small confocal volume. In order to perform FCS, fluorescently labelled proteins have to be prepared. The most common labelling methods are the labelling of the N-terminal amino group of proteins with N-hydroxysuccinimide-ester (NHS-ester) and cysteine with maleimide dyes. In protein folding studies, FCS is used to measure the change of diffusion time between the native and not-yet folded populations. Due to small differences in the diffusion time between these proteins, large molecules, such as GroEL, can bind to not-yet-folded but not the native state. Therefore, the rate of folding can be measured by distinguishing two populations via their diffusion time (Gupta et al., 2014). A variant of FCS is dual-colour fluorescence cross-correlation spectroscopy (dcFCCS), used for studying molecular interactions. Two spectrally different labelled proteins are prepared in this method to analyse protein-protein interactions. dcFCCS provides a particularly sensitive measure of aggregation during protein folding.

SmFRET enables us to monitor the conformational dynamics of individual protein molecules which cannot be revealed by classical ensemble experiments. To extend the capability of smFRET, such measurements can be performed using pulsed interleaved excitation (PIE), which is the technique of alternating laser excitation on the nanosecond timescales (Müller et al., 2005). To perform smFRET, cysteine residues (either intrinsic or introduced by mutation) at known positions should be labelled with spectrally suitable dye pairs. The power of smFRET experiments is to distinguish different subpopulations of molecules in heterogeneous mixtures (Lerner et al., 2018). For example, different conformations of proteins occur during the folding as proteins explore rugged energy landscapes. In such reactions, some subpopulations can accumulate as kinetically trapped states, while others may be chaperone-bound or have reached the native

state. Thus, all these three different subpopulations can be detected in the heterogeneous folding reaction.

Taken together, single-molecule studies in solution represent a valuable tool for studying the heterogeneous process of protein folding (Figure 2.13).

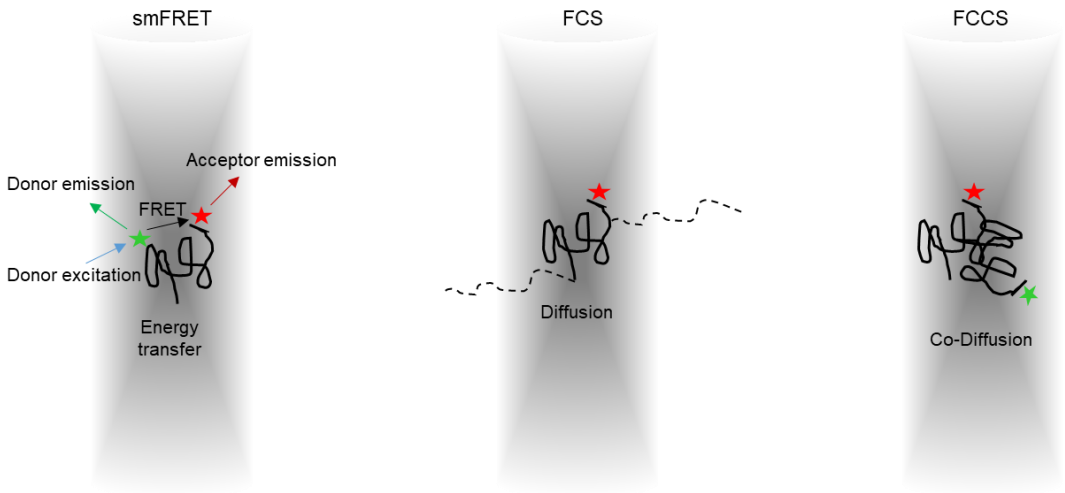


Figure 2.13 Jablonski diagram indicating the possible radiative and non-radiative transitions
In smFRET, the distance between two fluorophores is measured at very low concentrations (<100 pM) to ensure the presence of at most one double-labelled molecule in the confocal volume. SmFRET is performed with PIE setup to eliminate mono-labelled or the same fluorophore-labelled species. FCS is used for measuring the diffusion time of a labelled molecule. In FCCS, co-diffusion of differently labelled molecules is analysed.

2.9 Firefly luciferase (FLuc), a Hsp70 model substrate

Some organisms use bioluminescence for various purposes such as defense, camouflage, feeding and mating. Luciferase catalyzes a bioluminescent reaction in the presence of luciferin as substrate and ATP. *Photinus pyralis* firefly luciferase (FLuc) has been used extensively as a model multi-domain protein substrate of the chaperone DnaK. FLuc (~60 kDa) consists of a large N-domain (residues 1-440) and a smaller C-domain (residues 441-550) (Conti et al., 1996) (Figure 2.14). The N-domain can be divided into two subunits which are a small (N_S; residues 1-190) and a large (N_L; residues 191-440) subdomains (Frydman et al., 1999; Frydman et al., 1994; Scholl et al., 2014). FLuc uses ATP to catalyze the oxidation of D-Luciferin (luciferin), so that a photon is released as excited oxyluciferin returns to the ground state (Baldwin, 1996). The active

site pocket is located at the large N-terminal domain where ATP and luciferin bind, with the C-domain functioning as a lid over the active site pocket. DnaK-assisted, spontaneous and co-translational folding of FLuc have been extensively studied taking advantage of its sensitive bioluminescence assay. Although eukaryotic and bacterial Hsp70 can efficiently refold both thermally and chemically denatured FLuc (Agashe et al., 2004; Frydman et al., 1994; Schröder et al., 1993b; Szabo et al., 1994), the refolding of FLuc in the absence of chaperones is inefficient and slow (Herbst et al., 1998; Herbst et al., 1997). Spontaneous folding of FLuc was shown by two studies of Robert Seckler's group in 1997 and 1998, indicating that FLuc refolds spontaneously with high yield at temperatures below 30°C) and low protein concentrations (<30 nM) to minimize aggregation (Herbst et al., 1998; Herbst et al., 1997). Interestingly, they observed that the addition of a hydrophilic detergent (0.2% Tween20), increased the folding yield by preventing binding of FLuc to tube walls (Herbst et al., 1998; Herbst et al., 1997). They found that FLuc refolds spontaneously with a half time of 1.5 h at 20°C and 50 min at 30°C with the yield of ~80% in optimized conditions (Herbst et al., 1998; Herbst et al., 1997). Further studies by other groups showed that partially unfolded FLuc with the N-terminal domain folded can refold spontaneously (Scholl et al., 2014, 2017). Consistent with these reports, FLuc folds co-translationally with the N_S domain folding as it emerges from the ribosome, followed by the rapid formation of native FLuc when the full-length polypeptide is released from the ribosome (Frydman et al., 1999). Thus, vectorial domain folding avoids intramolecular misfolding and promotes the native state. Additionally, folding assisted by the Hsp70 system occurs with high efficiency with a half time of 5-8 min (Moran Luengo et al., 2018; Schröder et al., 1993b; Szabo et al., 1994). It remains to be determined whether Hsp70 physically separates the domains of FLuc during binding and release cycles, mimicking the effect of co-translational folding in avoiding inter-domain misfolding.

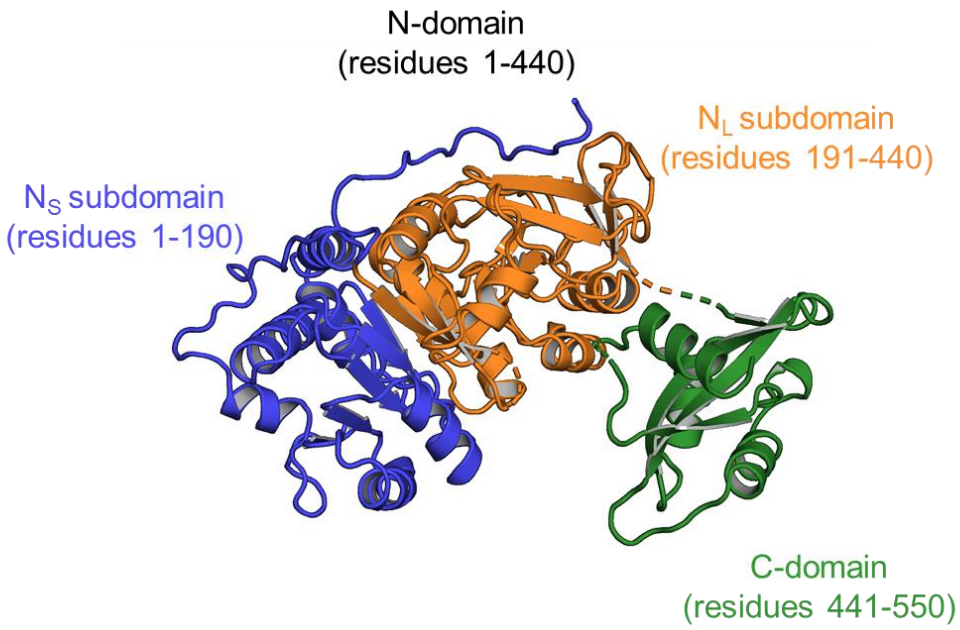


Figure 2.14 Structure of FLuc
The FLuc structure with domains is indicated (PDB: 1LCI). Blue, orange, green region is the N_S (1-190), N_L (191-440) and C- domains (441-550), respectively.

2.10 Aim of this study

Although Hsp70 is one of the most widely studied chaperone systems, with a wealth of information being available on the structure and allosteric regulation of Hsp70, there are still many unanswered questions regarding the mechanism of Hsp70 mediated protein folding. Specifically, it remains controversial whether successive cycles of Hsp70 binding and release merely prevent protein aggregation during folding or can actively promote the folding reaction to occur at an accelerated rate. Indeed, some recent studies suggested that the Hsp70 system unfolds misfolded protein species and can stabilize the native state out of equilibrium by using the energy of ATP, thereby shifting the folding equilibrium. Here, we investigated to what extent the Hsp70 system can modulate the energy landscape of folding reactions independent of preventing protein aggregation. The key question was whether Hsp70 not only increases the folding yield but also the folding rate. Using firefly luciferase as a model substrate and single-molecule FRET experiments we could demonstrate that that the Hsp70 chaperone system profoundly accelerates the folding rate in the absence of aggregation. These results allow us to develop a model for the mechanism underlying the Hsp70 assisted folding of multi-domain proteins.

3 Materials and Methods

3.1 Materials

3.1.1 Chemicals

Unless indicated otherwise, reagents and chemicals were purchased from Sigma-Aldrich (Steinheim, Germany).

Table 3.1 Chemicals

Suppliers	Chemical
Agilent Technologies (Santa Clara, California, USA)	QuikChange Multi Site-Directed Mutagenesis Kit
Atto-Tec GmbH (Siegen, Germany)	Atto532 maleimide Atto647N maleimide Atto655 maleimide
Carl Roth (Karlsruhe, Germany)	Isopropyl β-D-1-thiogalactopyranoside (IPTG)
Difco (Heidelberg, Germany)	Bacto Agar Bacto peptone Bacto Tryptone Bacto Yeast Extract Bacto Yeast Nitrogen Base
GE Healthcare (München, Germany)	Chloramphenicol
New England BioLabs (Ipswich, Massachusetts, USA)	Q5 Site-Directed Mutagenesis Kit
Merck (Darmstadt, Germany)	β-Mercaptoethanol Methanol Sodium bicarbonate
Metabion (Martinsried, Germany)	dNTP Oligonucleotides (Primers)

Pierce (Rockford, USA)	BSA Standard
Promega (Mannheim, Germany)	Luciferase assay system
Qiagen (Hilden, Germany)	Qiagen Plasmid Mini Kit
Roche (Basel, Switzerland)	Adenosine 5′-triphosphate disodium salt tri-hydrate (ATP) 1,4-Dithiothreitol (DTT) EDTA-free Complete protease inhibitor cocktail Aprotinin PepA Leupeptin
Roth (Karlsruhe, Germany)	Ampicillin Glycerol 4-(2-hydroxyethyl)-1-piperazineethanesul-fonic acid (HEPES) Imidazole Isopropyl β-D-1-thiogalactopyranoside (IPTG) Potasium chloride (KCl) Sodium chloride (NaCl)
Serva (Heidelberg, Germany)	Acrylamide-Bis solution 37.5:1 (30% w/v) Coomassie Blue R250 Lysozyme Phenylmethanesulfonylfluoride (PMSF) Sodium dodecyl sulfate (SDS)
Sigma-Aldrich (St. Louis, Missouri, USA)	Acetic acid Adenosin 5′-diphosphate sodium salt (ADP) 2-Amino-2-(hydroxymethyl)-1,3-propane-diol (Tris base) Ammonium persulfate (APS) Brij58

	Chloramphenicol Dimethylsulphoxide (DMSO) Ethanol Ethylenediaminetetraacetic acid (EDTA) Hydrochloric acid (37%) Magnesium oxalate 2-Phenylbenzothiazole (PBT) Phospho(enol)pyruvic acid monopotassium salt (PEP) Sodium dodecyl sulfate (SDS) Spermidine N,N,N',N'-Tetramethyl-ethane-1,2-diamine (TEMED) Tween-20
Thermo Fisher Scientific (Waltham, USA)	Guanidine-HCl Solution (8 M in H_2O) Alexa532 maleimide Alexa532 succinimidyl Alexa647 maleimide Alexa647 succinimidyl Dnase I TCEP·HCl Tris(2-Carboxyethyl) phosphine Hydrochloride

3.1.2 Proteins, enzymes

Table 3.2 Proteins, enzymes

Suppliers	Proteins, enzymes
Department of Cellular Biochemistry, MPI (Martinsried, Germany)	NR-Peptide NR-Peptide-Alexa532
Department of Cellular Biochemistry, MPI Prof. Dr. F. U. Hartl group (Martinsried, Germany)	GroEL GroEL (D87K)

Invitrogen (Karlsruhe, Germany)	Protein marker for SDS-PAGE
Promega (Mannheim, Germany)	QuantiLum® Recombinant Luciferase *Pfu* polymerase

3.1.3 Strains

Table 3.3 Strains

Suppliers	Chemical
Novagen (Darmstadt, Germany)	*E. coli* DH5α
Stratagene (Heidelberg, Germany)	*E. coli* BL21 (DE3)

3.1.4 Instruments

Table 3.4 Instruments

Suppliers	Chemical
Anton Paar (Graz, Austria)	DMA 5000 densitometer
Beckman Coulter (Pasadena, USA)	Benchtop centrifuge GS-6 High capacity centrifuge J6-MI Optima XL-I analytical centrifuge Ultracentrifuge Optima L-90K Ultracentrifuge rotor type 70 Ti Ultracentrifuge rotor type 45 Ti
Berthold Technologies (Oak Ridge, USA)	Lumat LB9508 luminometer
Eppendorf (Hamburg, Germany)	Benchtop centrifuges 5415D and 5417R Research plus pipette (2.5 µL, 10 µL, 20 µL, 100 µL, 200 µL, 1 mL) Protein LoBinding Tubes (0.5 mL, 1.5 mL)
Fujifilm (Tokio, Japan)	FLA-2000 Fluorescence Imager
GE Healthcare (München, Germany)	Äkta FPLC Äkta Purifier/Explorer Äkta Ettan

	Electrophoresis Power Supply – EPS600 Ettan LC Pre-packed chromatography columns: DEAE, MonoQ, Source 30Q, Sephacryl S200, Hiprep Heparin Sepharose FastFlow FF16/10, HiPrep 26/10 desalting, hydroxylapatit column, Enrich Q, HiPrep16/60 Sephacryl-S100, HisTrap^{TM}HP HiTrap Chelating HP, HiTrap Blue HP, HiPrep 26/10 Desalting, HiLoad 16/60 Superdex 200 prep grade, MonoQ HR 10/16, Mono S HR 10/10, NAP-5 Sephadex G25 desalting, NAP-10 Sephadex G25 desalting, NAP-25 Sephadex G25 desalting
Horiba Yvon	FluoroLog 3 spectrofluorometer
Ibidi (Martinsried, Germany)	µ-slide 8 well chambered microscope coverslip
Jasco (Gross-Umstadt, Germany)	V-560 Spectrophotometer
New Brunswick Scientific	Innova 44 incubator/shaker
Milipore (Bedford, USA)	Amicon centrifuge filter units, steritop filter units Centriprep concentrators (3000, 10.000, 30.000 and 50.000 Da MWCO) Steritop Filtration System (pore size 0.22 µm) Thermomixer comfort
Misonix (Farmingdale, USA)	Sonicator 3000
PicoQuant (Berlin, Germany)	MicroTime 200 time resolved, confocal fluorescence microscope
Roth (Karlsruhe, Germany)	ZelluTrans dialysis membrane

Scientific Industries, Inc. (Bohemia/NY, USA)	Vortex Mixer Genie 2
Thermo Fisher Scientific (Waltham, USA)	NanoDrop 1000
Wissenschaftlich Technische Werkstätten (Weilheim, Germany)	pH meter

3.1.5 Media and buffers

All buffers and media used for experiments were filtered by using SteriTop filter units (0.2 µm) and degassed by sonication, and subsequently autoclaved when indicated in the methods section.

Table 3.5 Media and buffers

Media and buffers	Reagent
LB Medium	10 g/L Bacto peptone 5 g/L Bacto Yeast Extract 10 g/L NaCl (15 g/L agar for solid medium)
Coomassie staining solution	40% ethanol 8% acetic acid 0.1% (w/v) Serva Coomassie Blue R-250
Coomassie destaining solution	10 % (v/v) Ethanol 7 % (v/v) Acetic acid
2 x SDS loading buffer	0.25 M Tric-HCl (pH 6.8) 0.5 M DTT 10% SDS 50% glycerol 0.5% bromphenol blue
10 x SDS-PAGE running buffer	250 mM Tris-HCl (pH 8.8) 1920 mM glycine 1% SDS

3.2 Molecular biology methods

3.2.1 Plasmid transformation of competent *E. coli* cells

Approximately 50 ng of plasmid DNA was added to 100 µL of competent cells, followed by incubation on ice for 10 min. Cells were heat-treated at 42°C for 45 seconds for efficient transformation and kept on ice for 2 extra min, followed by addition of 400 µL LB medium and incubation in a shaker at 37°C for 1 h. Aliquots of culture were plated on pre-heated LB agar plates containing suitable antibiotics. Plates were then incubated at 37°C for 16 h.

3.2.2 Site-directed mutagenesis with polymerase chain reaction (PCR)

Full-length FLuc from *Photinus pyralis* was cloned into plasmid pET3a, carrying an isopropyl-1-thio-β-D-galactopyranoside (IPTG) inducible promoter and site-directed mutagenesis was performed by PCR. Point mutations for site-specific fluorophore labelling with maleimide chemistry dyes were introduced using the QuikChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies). A stop codon was inserted using the Q5 Site-Directed Mutagenesis Kit (New England BioLabs) to generate FLuc190. The NEBaseChanger™ tool (New England BioLabs) were used to design the degenerated primers. The complete protocol, thermocycling reaction and the primer list is shown in Tables 3.6, 3.7, 3.8. After PCR reactions, 1 µL *Dpn*I was added to 25 µL PCR sample and incubated at 37°C for 1 h to digest the template DNA. Finally, 2 µL of the sample was transformed into *E. coli* DH5α. All isolated plasmids were verified by sequencing (Eurofins). Plasmids were stored at –20°C.

Table 3.6: Primers used in this study

FLuc Mutant	Primer	Sequence (5'-3')
FLuc (D19C)	Forward	ATCCTCTAGAGTGTGGAACCGCTGGAG
	Reverse	CTCCAGCGGTTCCACACTCTAGAGGAT
FLuc (S170C)	Forward	CGTTCGTCACATGTCATCTACCTCC
	Reverse	GGAGGTAGATGACATGTGACGAACG
FLuc (E428C)	Forward	GCTTACTGGGACTGTGACGAACACTTC
	Reverse	GAAGTGTTCGTCACAGTCCCAGTAAGC
FLuc (D476C)	Forward	TCCCGACGATTGCGCCGGTGAACTTCC
	Reverse	GGAAGTTCACCGGCGCAATCGTCGGGA

FLuc(S504C)	Forward	GATTACGTCGCCTGTCAAGTAACAACC
	Reverse	GGTTGTTACTTGACAGGCGACGTAATC
FLuc190 (truncated)	Forward	CCACCACTAGACAATTGCACTGA-TAATGAATTC
	Reverse	TGATGATGATGTTTGTCACGATCAAAGGAC

Table 3.7: PCR reaction

Reaction component	Volume
10x QuikChange Multi reaction buffer	2.5 µL
ddH_2O	X µL to final volume of 25 µL
QuikSolution	0.5 µL
Template plasmid	100 ng
Primers (forward or reverse) for each mutation	100 ng
dNTP mix	1 µL
QuikChange Multi enzyme blend	1 µL
Total volume	25 µL

Table 3.8: PCR amplification protocol

Seg-ment	Process	Temperature	Duration	Cycles
1	Initial denaturation	95°C	1 min	1
2	Cycle denaturation	95°C	1 min	30
3	Primer annealing	55°C	1 min	
4	Primer extension	65°C	2 min/kb	
5	Final extension	70°C	8 min	1
6	Cooling	4°C	∞	1

3.2.3 Plasmid DNA and DNA fragment purification

Amplification of plasmids was performed in *E. coli* DH5α cells. The transformed cells were grown overnight in LB medium in the presence of suitable antibiotics as a selective marker.

Plasmids were isolated using the Qiagen Plasmid Mini Kit and isolated plasmids were stored at -20°C.

3.3 Protein biochemical and biophysical methods

3.3.1 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Induction of protein expression and all subsequent steps of purification were analysed by SDS-PAGE. Gels were prepared as described in Table 3.9 and the sample was mixed with 5 x SDS loading buffer (0.25 M Tris-HCl (pH 6.8), 0.5 M DTT, 10% SDS, 50% glycerol 0.5% bromo-phenol blue). Before loading into the gel, the protein sample was completely denatured at 95°C for 5 min. Gel electrophoresis was performed in the SDS-PAGE running buffer (25 mM Tris-HCl (pH 8.8), 192 mM glycine, 0.1% SDS) at 200 V using Bio-Rad Mini Protean 2 electro-phoresis. For visualization of protein bands, the gels were incubated in Coomassie gel staining solution (40% ethanol, 8% acetic acid, 0.1% (w/v) Serva Coomassie Blue R-250). To remove the dye from protein-free regions, Coomassie destaining solution (10% ethanol, 7% acetic acid) was used.

Protein concentrations were measured either by Bradford assay or by measuring the absorption of the sample at 280 nm (A_{280}). Molar extinction coefficient at A_{280} were calculated with the ProtParam (http://web.expasy.org/protparam).

Table 3.9: Protocol of resolving and stacking gels for SDS-PAGE.

Resolving gel	12%	15%	Stacking gel	5%
Deionized water	4.900 mL	4.300 mL	Deionized water	4.100 mL
1.5 M Tris-HCl pH 8.8	3.800 mL	3.800 mL	1.0 M Tris-HCl pH 6.8	0.750 mL
30 % Acrylamide	6.000 mL	7.500 mL	30 % Acrylamide	1.000 mL
10 % (w/v) SDS	0.150 mL	0.150 mL	10 % (w/v) SDS	0.060 mL
10 % (w/v) APS	0.150 mL	0.150 mL	10 % (w/v) APS	0.060 mL
TEMED	0.006 mL	0.006 mL	TEMED	0.006 mL

3.3.2 Expression and purification of DnaK

To overexpress DnaK, plasmid pET11d_DnaK was transformed into BL21(DE3) cells. Cells were grown at 37°C until OD_{600}=0.6-0.8 was reached. DnaK expression was induced with 0.5

mM IPTG at 30°C. After 4-5 h of induction, cells were harvested by centrifugation at 4200 rpm for 30 min. Cells were resuspended in buffer A (20 mM Tris-HCl (pH 7.4), 1 mM EDTA) and subsequently frozen in liquid nitrogen and stored at -80°C. The resuspended pellet was thawed in a 23°C water bath. Cells were lysed both chemically and physically. Chemical lysis was achieved by lysozyme, PMSF, Aprotinin, PepA, Leupeptin, Dnase I and protease inhibitor (Complete^{TM}, EDTA free, Roche). Physical lysis of the cells was performed in an Emulsiflex apparatus. Then the sample was centrifuged for 45 min at 4°C and 4000 rpm in a Beckman 45 Ti rotor. The supernatant was loaded onto a Source 30Q (GE Healthcare, 50 mL) column which is pre-equilibrated with buffer A. After loading the sample, the column was washed with buffer A to remove unspecific binding protein. Elution was performed with buffer B (20 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1000 mM NaCl). The fractions containing DnaK were pooled and loaded onto a HiPrep 26/10 desalting (GE Healthcare, 56 mL) column and buffer was exchanged to 1% HMK_{1000} buffer (20 mM Hepes/KOH (pH 7.4), 5 mM $Mg(OAc_2)$, 1000 mM KCl; HM buffer: 20 mM Hepes/KOH (pH 7.4), 5 mM $Mg(OAc_2)$). Then, the sample was loaded onto a Hiprep Heparin Sepharose FastFlow FF16/10 (GE Healthcare; 20 ml) column which was pre-equilibrated with 1% HMK_{1000}. Elution was performed with a gradient of 1 - 50% HMK_{1000}. The fractions containing DnaK were pooled and buffer exchanged to 1% HMK_{1000} for the third chromatographic step on MonoQ 10/100 (GE Healthcare, 8 mL). Elution was performed with a gradient of 1 - 50% HMK_{1000} and the fractions containing DnaK were pooled and buffer exchanged to 10% HMK_{1000}. As a last chromatographic step, DnaK was purified by gel filtration using Superdex 200 16/60, (110 mL). The fractions were analysed on 12% SDS-PAGE and DnaK containing fractions were pooled. The concentration of DnaK was measured by the nanodrop method before snap-freezing in liquid nitrogen and storage at -80°C.

3.3.3 Expression and purification of DnaJ

DnaJ was purified as described previously (Zylicz et al., 1985). To overexpress DnaJ, plasmid pET11d_DnaJ was transformed into BL21(DE3) cells. Cells were grown at 37°C until OD_{600} =0.6-0.8 was reached. DnaJ expression was induced with 0.5 mM IPTG at 30°C. After 4-5 h of induction, cells were harvested by centrifugation at 4200 rpm for 30 min. Cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 10% (w/v) sucrose, 10 mM DTT, 10 mM EDTA, 0.6% (w/v) Brij58 and 1 mg/mL lysozyme) and subsequently frozen in liquid nitrogen

and stored at -80°C. Thawed cells were lysed by ultrasonication and the lysate was then centrifuged at60000 g for 1 h at 4°C. The pellet was resuspended in buffer A (2M urea, 50 mM Tris-HCl pH 7.8, 100 µg/mL PMSF, 5 mM DTT, 5 mM EDTA, 10% (w/v) sucrose, and 0.1% (v/v) Triton X-100) and mixed for 1 h at 4°C. The extract was then centrifuged again at 60000 g for 1.5 h at 4°C. Subsequently, the supernatant was loaded onto a DEAE column pre-equilibrated with buffer A. Elution was performed with buffer B1 (40 mM potassium phosphate pH 6.8, 5 mM DTT, 0.1 M KCl, 10% (v/v) glycerol, and 0.05% (w/v) Brij58). Fractions of interest were pooled and precipitated with 0.28 g/mL ammonium sulfate, followed by centrifugation at 20000 g for 30 min at 4°C. The pellet was washed with 1/10 volumes of buffer B1 containing 0.28 g/mL ammonium sulfate and subsequently centrifuged at 20000 g for 30 min at 4°C. The pellet was resuspended in 5 mL buffer B1 and loaded onto HiPrep 26/10 desalting column pre-equilibrated with buffer B1. For 2nd chromatographic step, the sample was loaded onto hydroxylapatit column (30 mL) pre-equilibrated with buffer B1. The sample was eluted with buffer B2 (500 mM potassium phosphate (pH 6.8), 5 mM DTT, 0.1 M KCl, 10% (v/v) glycerol, and 0.05% (w/v) Brij58). The fractions were analysed on 12% SDS-PAGE. DnaJ containing fractions were pooled and concentrated with 10 kDa cut-off membrane. The concentration of DnaJ was measured by the Bradford assay before snap-freezing in liquid nitrogen and storage at -80°C.

3.3.4 Expression and purification of GrpE

To overexpress GrpE, plasmid pET11d_GrpE was transformed into BL21 (DE3) cells. Cells were grown at 37°C until OD_{600}=0.6-0.8 was reached. GrpE expression was induced with 0.5 mM IPTG at 30°C. After 4-5 h of induction, cells were harvested by centrifugation at 4200 rpm for 30 min. Cells were resuspended in a minimal volume of buffer A (50 mM Tris/HCl (pH 8.0), 10% sucrose (w/v)) and subsequently frozen in liquid nitrogen and stored at -80°C. Thawed cells were lysed by incubation with 10 mL of a lysozyme solution containing 46 µg/mL spermidine, 50 mM EDTA, 50 mM DTT, 0.2 g/mL ammonium sulfate and 440 µL of a 100 mg/mL solution of lysozyme. Dilution in buffer A was performed by bringing the volume to final 74 mL and was incubated on ice for 45 min. It was then transferred to 37°C for 5 min with shaking which is followed by cooling on ice for 5 min. The lysate was clarified by centrifugation at 66000 g for 30 min at 4°C. Then, the supernatant was taken and ammonium

sulfate was gently added to a final concentration of 0.35 g/mL under stirring for 20 min. Following centrifugation at 66000 g for 30 min, the pellet was resuspended in 10 mL of buffer B (25 mM Hepes/KOH (pH 8.0), 10 mM 2-mercaptoethanol, 1 mM EDTA, 10% (v/v) glycerol) and loaded then onto an Enrich Q column (BioRad) which is pre-equilibrated with buffer B. Elution was performed with buffer B1 (25 mM Hepes/KOH (pH 8), 2 M KCl, 2-mercaptoethanol, 1 mM EDTA, 10% (v/v) glycerol). Fractions containing GrpE were buffer exchanged into buffer B and then second round of ion exchange chromatography was performed using a MonoQ (GE Healthcare). As a final chromatographic step, the pooled protein was further purified by gel filtration using HiPrep16/60 Sephacryl-S100 (GE Healthcare) which is pre-equilibrated with buffer C (25 mM Hepes/KOH (pH 8.0), 50 mM KCl, 10 mM 2-mercaptoethanol, 0.1 mM EDTA, 20% (v/v) glycerol). The fractions containing pure GrpE were analysed on 12% SDS-PAGE and pooled. The concentration of protein was measured by the Bradford assay before snap-freezing in liquid nitrogen and storage at -80°C.

3.3.5 Expression and purification of FLuc mutants

Wild-type FLuc was purchased from Promega, or expressed and purified as described below (Imamoglu et al., 2020). Plasmid pET3a Luc_FXa_Myc_6xHis was transformed into BL21 (DE3) cells. Cells were grown at 37°C until OD_{600}~0.4 was reached. The temperature was then shifted to 18°C and cells were incubated for 1 h more before the IPTG induction. FLuc expression was induced with 0.25 mM IPTG for 16 h at 18°C. Cells were harvested by centrifugation at 4200 rpm for 30 min. Pellets were resuspended in resuspension buffer (50 mM NaH_2PO_4 (pH 8.0), 10 mM imidazole, 300 mM NaCl) before freezing in liquid nitrogen and storage at -80°C. Cells were thawed in a 4°C water bath. Thawed cells were lysed both chemically and physically. Chemical lysis was achieved by lysozyme (1 mg/mL), PMSF, Aprotinin, PepA, Leupeptin, Dnase I and protease inhibitor (Complete™, EDTA free, Roche). Physical lysis of the cells was performed in the Misonix sonicator®3000 (Cole-Parmer) and the lysate was cleared by centrifugation at 17400 g for 30 min. The supernatant was then loaded onto a Ni-NTA column (HisTrap™HP 5 mL, GE Healthcare Life Sciences), which is pre-equilibrated with equilibration buffer (50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 20 mM imidazole). Protein was eluted with elution buffer (50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 500 mM imidazole) collecting 2 mL fractions. The fractions were assessed by 12% SDS-PAGE. Fractions

that contain pure FLuc were pooled and buffer exchanged into storage buffer (25 mM Tris-acetate (pH 7.8), 1 mM EDTA, 2 mM TCEP, 0.2 M Ammonium sulfate, 15% glycerol, 30% ethylene glycol). The concentration of FLuc was measured by the nanodrop method before snap-freezing in liquid nitrogen and storage at -80°C.

To overexpress FLuc190, plasmid pET3a FLuc (1-190) 6xHis was transformed into BL21(DE3) cells. Growth of the cells, the induction and the purification were performed as described for other FLuc purifications.

3.3.6 Luminescence assay of FLuc folding

The Lumat LB9508 luminometer (Berthold Technologies) was used to measure the activity of FLuc. To measure the luminescence activity, aliquots of folding reactions or native FLuc were diluted at least 25-fold in Luciferase Assay System (Promega) at indicated time points (Imamoglu et al., 2020). Luminescence activity was measured for 2 seconds at 25°C (Imamoglu et al., 2020).

To measure spontaneous and assisted folding, stock FLuc at -80°C was thawed on ice and then centrifuged at 20000 g for 30 min to remove any aggregates. FLuc was then chemically denatured in 5 M GuHCl and 10 mM DTT for 1 h at 25°C. For spontaneous folding, denatured FLuc at different concentrations was 100-fold diluted into reaction buffer (25 mM Hepes/KOH (pH 7.5), 100 mM KCl, 10 mM $Mg(OAc)_2$, 2 mM DTT, 0.05% Tween20). For assisted folding, 10 µM denatured FLuc was diluted 100-fold into reaction buffer (25 mM Hepes/KOH (pH 7.5), 100 mM KCl, 10 mM $Mg(OAc)_2$, 2 mM DTT, 0.05% Tween20) containing 3 µM DnaK, 1 µM DnaJ, 1.5 µM GrpE and 5 mM ATP, unless otherwise stated.

For the rescue experiment, 0.3 µM DnaK, 0.1 µM DnaJ, 0.5 µM GrpE and 5 mM ATP were added to spontaneous folding of 1 nM FLuc after 2 h, 4 h, 8 h.

For the cycle assay, denatured FLuc was diluted to 100 pM in folding buffer containing 0.3 µM DnaK, 0.1 µM DnaJ, 5 mM ATP and 125 µM DnaK-model peptide (H-GSGNRLLLTG-OH) (Zhu et al., 1996). After incubation for 1 min, the reaction was initiated by the addition of 0.5 µM GrpE, followed by the luminescence activity measurement at intervals as described

above. As controls, DnaK-model peptide was added to the buffer containing chaperone mixture prior to addition of FLuc, or after 5 min of assisted folding.

Folding reactions at 37°C were performed the same as described above, except that the folding was assayed at 10 pM, and assisted folding reaction was performed in the presence of 1 µM DnaK, 0.33 µM DnaJ, 1.5 µM GrpE and 5 mM ATP.

The yield was measured according to the activity of the same concentration of non-denatured (native) FLuc. Data fitting was performed in sigma plot 14. Folding kinetics were fitted with a single exponential equation.

3.3.7 Fluorophore labelling of proteins for smFRET

For smFRET experiments, double-cysteine mutant variants of FLuc were first labelled with ATTO532-maleimide (ATTO-TEC) and ATTO647N-maleimide (ATTO-TEC) at 1.5-fold molar excess of each fluorophore (Imamoglu et al., 2020). Before the labelling reaction, cysteine residues of FLuc were reduced in 2 mM tris(2-carboxyethyl)phosphine (TCEP) that was subsequently removed by using a Nap5 column (GE Healthcare). To perform the labelling reactions, FLuc was added to labelling buffer (25 mM Hepes/KOH (pH 7.5), 100 mM KCl, 10 mM $Mg(OAc)_2$, 100 µM PBT, 5 mM ATP), which is premixed with fluorophores and incubated for 2 h on ice. After the labelling reaction, free dye was removed using a Nap5 column (GE Healthcare) that was pre-equilibrated with the storage buffer (25 mM Tris-acetate (pH 7.8), 1 mM EDTA, 2 mM TCEP, 0.2 M ammonium sulfate, 15% glycerol, 30% ethylene glycol). The concentration of protein and the labelling efficiency (typically ~70%) were measured by the nanodrop method as described below before snap-freezing in liquid nitrogen and storage at -80°C.

Specific labelling of introduced cysteines was confirmed by the consistency of FRET efficiency distribution with the distance between labelling residues, which is calculated according to crystal structure (PDB: 1LCI).

3.3.8 Fluorophore labelling of proteins for fluorescence correlation and dual-colour cross-correlation spectroscopy (FCS and dcFCCS)

For FCS and dcFCCS experiments, wild-type FLuc was labelled with Alexa532 or Alexa647 N-hydroxysuccinimide (NHS) ester dye (succinimidyl ester) at N-terminus (Imamoglu et al., 2020). FLuc was incubated with a 5-fold excess of Alexa532 or Alexa647 NHS ester (Molecular Probes) at 4°C for 90 min in 0.1 M $NaHCO_3$ (pH 8.3). Non-reacted dyes were removed using a Nap5 column (GE Healthcare) pre-equilibrated with storage buffer (25 mM Tris-acetate (pH 7.8), 1 mM EDTA, 2 mM TCEP, 0.2 M ammonium sulfate, 15% glycerol, 30% ethylene glycol). The concentration of protein and the labelling efficiency (typically ~75%) were measured by the nanodrop method before snap-freezing in liquid nitrogen and storage at -80°C. As a control, double-labelled FLuc was prepared using FLucNC, which was labelled with Alexa532 and Alexa647 C5 maleimide (Molecular Probes) at a 1.5-fold molar excess of each fluorophore. The labelling reaction was performed as described for the smFRET constructs.

N-terminus of DnaJ was labelled with Alexa647 N-hydroxysuccinimide ester (Molecular Probes) at a 2-fold molar excess and labelling was performed at 4°C for 90 min in 0.1 M $NaHCO_3$ (pH 8.3). Free dye after the labelling was removed using a Nap5 column, which is pre-equilibrated with buffer containing 25 mM Hepes-KOH pH 7.5, 100 mM KCl, 10 mM $Mg(OAc)_2$, 2 mM DTT, 10% (v/v) glycerol). Labelling efficiencies were approximately ~80%.

For analytical ultracentrifugation experiments, labelling of FLucNC was performed with a 6-fold molar excess of Atto532-maleimide in labelling buffer containing 4 M GuHCl at 25°C for 1 h. Free dye was removed using a Nap5 column, which is pre-equilibrated with labelling buffer containing 4 M GuHCl.

3.3.9 Degree of labelling

The degree of labelling (DOL) was controlled by the nanodrop method, using the following extinction coefficients

(FLuc, ε_{280}=39310 M^{-1} cm^{-1}; ATTO532, ε_{532}=115000 M^{-1} cm^{-1}, cf_{280}=0.11; ATTO647N, ε_{647}=150000 M^{-1} cm^{-1}, cf_{280}=0.05; Alexa532, ε_{532}=81000 M^{-1} cm^{-1}, cf_{280}=0.09; Alexa647, ε_{647}=239000 M^{-1} cm^{-1}, cf_{280}=0.03) and the equation,

$$\mathrm{DOL} = \frac{A_{dye} \times \varepsilon_{dye}}{(-(A_{dye} \times cf_{280}) + A_{280}) \times \varepsilon_{protein}}$$

3.3.10 FCS and dcFCCS experiments

FCS and dcFCCS measurements were performed on a Microtime 200 inverse time-resolved fluorescence microscope (PicoQuant) using pulsed interleaved excitation (PIE) (Müller et al., 2005) as previously described (Gupta et al., 2014). Experiments were performed 20°C unless otherwise stated. Pulsed diode lasers at 530 nm (LDH-P-FA-530) and at 640 nm (LDH-PC-640B) were used for the excitation of Alexa532 and Alexa647, respectively. For FCS measurements, the laser power was set to 60 µW, whereas each laser had a power of 40 µW for dcFCCS measurements. The power of lasers was measured before the major dichroic filter and the lasers were pulsed with a rate of 26.7 mHz for experiments. The excitation light was directed into the sample cuvette (Ibidi), through a water immersion objective (60 x 1.2 NA, Olympus). The emitted light was separated from excitation light by a dichroic mirror (Z532/635RPC), which is guided through a pinhole (75 µm) and split by a beam splitter (600 DCXR) onto photon avalanche diodes (SPADs) (PDM series, MPD) according to wavelength. Emission bandpass filters (HQ 690/70 and HQ 580/70, Chromas) in front of the respective detector were employed to filter the emission light. Detection was performed with time-correlated single photon counting (TCSPC). For FCS measurements, fluorescence lifetime filters (Symphotime, PicoQuant) were used to remove the after-pulsing artifacts (Enderlein et al., 2005). The symphotime software 32 (PicoQuant) was used to fit the correlation plots with triplet diffusion equation. ATTO655 maleimide dye was used to calibrate the confocal volume (V_{eff}).

3.3.11 Inter-molecular association measured by dcFCCS

dcFCCS measurements were employed to analyse the presence of inter-molecular interactions during the spontaneous folding of FLuc (Imamoglu et al., 2020). We analysed the presence of transient oligomerization at different concentrations. To determine the monomeric concentration of FLuc during spontaneous folding reaction, Alexa532 succinimidyl labelled FLuc (FLuc-Alexa532) and Alexa647 succinimidyl labelled FLuc (FLuc-Alexa647) were denatured in 5 M GuHCl and 10 mM DTT for 1 h at room temperature. They were then diluted to different final concentrations in folding buffer and a 1:1 molar ratio of labelled proteins was maintained.

Where indicated, different concentrations of denatured, un-labelled FLuc was added to reactions. dcFCCS was recorded with PIE for 30 min at 25°C, or 1 h at 37°C. Reactions were performed at 25°C or 37°C using a temperature-controlled chamber (Ibidi Heating System). As a positive control, double-labelled FLucNC was used.

To test the presence of DnaJ in the chaperone-FLuc complex, complexes were prepared and analysed as described above, and dcFCCS was recorded for 30 min. Complexes with labelled DnaJ were prepared with 0.3 µM DnaK, 90 nM un-labelled DnaJ, 10 nM DnaJ-Alexa647, and 10 nM FLuc-Alexa532 (Imamoglu et al., 2020). To determine whether complexes contain multiple FLuc molecules, complexes were prepared with 0.3 µM DnaK, 0.1 µM DnaJ, 50 pM FLuc-Alexa532 and 50 pM FLuc-Alexa647. As a positive control, double-labelled FLucNC-Alexa532/647 at 50 pM was used.

3.3.12 Spontaneous and assisted folding of FLuc measured by FCS

FCS measurements were performed to measure the folding kinetics of spontaneous and assisted folding of 500 pM FLuc-Alexa647. FLuc-Alexa647 was denatured as described above. Spontaneous folding at 25°C was initiated by a 100-fold dilution of the denatured FLuc-Alexa647 to reaction buffer (25 mM Hepes/KOH (pH7.5), 100 mM KCl, 10 mM $Mg(OAc)_2$, 50 µM PBT, 2 mM DTT, 0.05% Tween20, 5 mM ATP). Folding reactions were stopped at indicated time points by the addition of 1 µM GroEL (D87K). GroEL (D87K) binds to not-yet folded FLuc but not the native FLuc. After stopping the reaction, FCS measurements were performed at 20°C for 30 min. From the average speed of diffusion, we deduced the fraction of folded protein. For assisted folding, the unfolded FLuc-Alexa647 was diluted into reaction buffer (25 mM Hepes/KOH (pH7.5), 100 mM KCl, 10 mM $Mg(OAc)_2$, 50 µM PBT, 2 mM DTT, 0.05% Tween20, 5 mM ATP) containing 0.3 µM DnaK, 0.1 µM DnaJ, 0.5 µM GrpE. Folding reactions were stopped and subsequently measured as described above for the spontaneous folding reaction. To measure the folding kinetics, the mean diffusion time, reflecting a shift of FLuc molecules from GroEL (D87K)-bound FLuc to free, were plotted against the folding time and fitted with a single exponential rate.

3.3.13 Single-molecule FRET (smFRET)

SmFRET experiments were generally performed at a constant temperature of 20°C on a Microtime 200 inverse time-resolved fluorescence microscope (PicoQuant) using PIE (Imamoglu et al., 2020; Müller et al., 2005; Sharma et al., 2008) as described above. Each laser had a laser power of 40 µW, which is measured before the major dichroic filter. Unless otherwise stated, we performed all smFRET experiments at 50 pM FLuc to ensure that the statistical probability of having more than one FLuc residing in the confocal observation volume is <1% at this concentration. At least 500 events were analysed for a significant measurement.

FRET was recorded with PIE for 15 min (chaperone-assisted folding at 20 °C), 30 min (chaperone-assisted folding at 37 °C) or 1 h (spontaneous folding), unless otherwise stated (Imamoglu et al., 2020). Assisted folding reactions contained 0.3 µM DnaK, 0.1 µM DnaJ and 0.5 µM GrpE unless otherwise stated. The smFRET experiments were performed in folding buffer with 5 mM ATP and 50 µM PBT and performed at least three times (Imamoglu et al., 2020).

The role of DnaK at 37°C was also analysed with the smFRET experiment (as described above) using FLucNC labelled with ATTO532 and ATTO647N. 100 pM of FLucNC was incubated at 37°C in the presence of 0.3 µM DnaK, 0.1 µM DnaJ for certain time. Then, 0.5 µM GrpE were added to the reaction to initiate fording reaction under the denaturing condition.

A burst intensity approach (Deniz et al., 1999; Zander et al., 1996) in Symphotime 32 (PicoQuant) was used to analyse data. A double-labelled FLuc that diffuse through the confocal volume leads to a burst in fluorescence intensity. When a burst contains more than 25 photons in a 1 ms time window, it was considered as a significant event (Gupta et al., 2014). The presence of an acceptor fluorophore was confirmed by a threshold of 15 photons following red excitation (Imamoglu et al., 2020). FRET efficiencies were calculated from fluorescence intensities of Donor I_D and Acceptor I_A fluorophore by the following equation:

$$E = \frac{I_A}{I_A + \gamma I_D}$$

where $\gamma = (\Phi_A \eta_A / \Phi_D \eta_D)$ is a correction factor for differences in quantum yields (Φ) and detection efficiencies (η) (Lee et al., 2005; Müller et al., 2005) and was found to be 0.9 in previous

studies with the same instrument (Gupta et al., 2014). Crosstalk and the direct excitation of acceptor by green laser were subtracted.

The FRET efficiency histograms were analysed with Origin (OriginLabs). To quantify the fraction of N-FLuc, the area of the histogram corresponding to N-FLuc was divided by the total area of the histogram (Imamoglu et al., 2020) fitting to a sum of Gaussian distributions. The fraction of N-FLuc was then plotted against folding time and fitted with a single exponential function (Gupta et al., 2014). All experiments were performed at least in triplicate to ensure reproducibility.

3.3.14 Analytical ultracentrifugation (AUC)

Optima XL-I analytical centrifuge (Beckman Inc., Indianapolis, In, U.S.A.) using an An-60 Ti rotor and double-sector 12 mm centrepieces were used to perform sedimentation velocity experiments (Imamoglu et al., 2020). For AUC experiment, the labelling of FLucNC was performed in 4 M GuHCl with labelling buffer without PBT to improve labelling efficiency. Before the labelling reactions, proteins were reduced in 2 mM TCEP that was subsequently removed by Nap5 column (GE Healthcare). FLuc was incubated for 1 h at room temperature with 6-fold excess ATTO532 maleimide (ATTO-TEC). Free dyes were removed using Nap5 column equilibrated with labelling buffer containing 4 M GuHCl in 25 mM Hepes/KOH (pH 7.5), 100 mM KCl, 10 mM $Mg(OAc)_2$. The complex was analysed in 25 mM Hepes/KOH (pH 7.5), 100 mM KCl, 10 mM $Mg(OAc)_2$, 2 mM TCEP, 5 mM ATP, 200 nM final concentration of FLuc containing 3 µM DnaK, 1 µM DnaJ. Buffer density was measured using a DMA 5000 densitometer (Anton Paar, Graz, Austria) and found as 1.02054 kg/L (Imamoglu et al., 2020). Protein concentration distribution was monitored by the optical absorption at 532 nm, using 184000 g rotor speed. The SEDFIT software package was used to compute time-derivative analysis (Schuck, 2000), leading to a c(s) distribution and an estimate for the molecular weight (Imamoglu et al., 2020).

3.3.15 Computational prediction of DnaK binding sites on FLuc

The LIMBO web server was used to predict DnaK binding sites on FLuc (Van Durme et al., 2009b). High confidence DnaK binding sites were considered for heptapeptide motifs with a score >10 on the sequence of *Photinus pyralis* luciferase (Imamoglu et al., 2020).

4 Results

4.1 Bacterial Hsp70 system accelerates FLuc folding

We used firefly luciferase (FLuc) as a model multi-domain substrate to investigate the mechanism of Hsp70-mediated folding, using the bacterial Hsp70 system, DnaK/DnaJ/GrpE (KJE). We analysed the spontaneous and KJE-assisted folding of FLuc by a luminescence assay. Consistent with other studies (Herbst et al., 1998; Herbst et al., 1997), we observed minimal spontaneous (~25% after 3 h) folding at 100 nM FLuc (Figure 4.1a). Next, we tested the KJE-assisted folding of FLuc by luminescence assay. Since previous studies showed that the efficiency of the refolding activity depends on the optimized concentration of KJE (Sarbeng et al., 2015), we first optimized the concentration of chaperones for efficient refolding, which was achieved at a molar ratio of 3:1:1.5 for K/J/E. We found that KJE-ATP rapidly refolded denatured FLuc with a $t_{1/2}$ of ~4 min and ~90% efficiency (Figure 4.1a). To understand the contribution of individual chaperones to the KJE-ATP assisted folding reaction, we diluted denatured FLuc into buffer where one chaperone or all chaperones were omitted. We added the omitted chaperone or chaperones after 1 h of folding. Minimal folding activity was observed in the absence of any single chaperone/co-chaperone (Figure 4.1b), indicating that omission of any chaperone abolished the reactivation of FLuc. However, the addition of DnaK or GrpE after 1 h resulted in ~80% or ~90% activity, respectively (Figure 4.1b), although the addition of DnaJ did not lead to efficient folding. Consistent with previous studies (Kellner et al., 2014; Szabo et al., 1994), these results demonstrate that DnaJ is required early in the reaction, for the complex formation between the DnaK and the substrate protein. DnaK cannot efficiently bind to substrate protein in the absence of DnaJ.

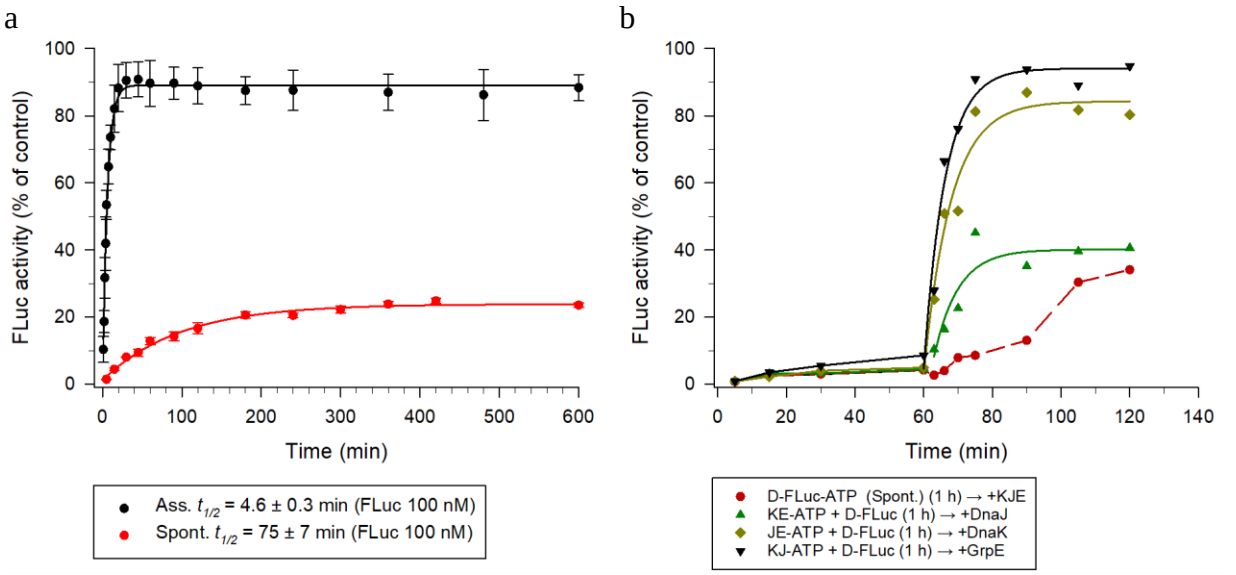


Figure 4.1 The KJE chaperone system assisted folding and the spontaneous folding of FLuc. (a) Spontaneous (red) and the KJE-assisted (black) folding of 100 nM FLuc was assayed by luminescence assay. KJE-assisted folding was performed with 3 µM DnaK, 1 µM DnaJ, 1.5 µM GrpE and 5 mM ATP. Error bars represent s.d. (n=10). (b) The chaperone dependency of the KJE-assisted folding reaction. Folding of FLuc was assayed in the presence of two chaperones and in the absence of one chaperone. The omitted chaperone (DnaJ, DnaK and GrpE omitted is green, brown and black, respectively.) was added to the folding reaction after 1 h. Alternatively, all chaperones were also added to spontaneous folding reaction after 1 h (red). Concentrations of KJE-ATP were the same as in (a). Ass., KJE-ATP-assisted; Spont., spontaneous folding; D-FLuc, denatured FLuc.

To investigate whether FLuc folds spontaneously in the absence of aggregation, we first measured folding at different concentrations of FLuc. The rate and the yield of spontaneous folding are independent of protein concentration in the absence of oligomerization, whereas aggregation impacts the rate and the yield of spontaneous folding. Importantly, transient oligomerization during spontaneous folding may change the rate but not the yield of folding. In light of this, we performed dual-colour fluorescence cross-correlation spectroscopy (dcFCCS) to establish a condition where (transient) aggregation is completely excluded. dcFCCS is a very sensitive method to detect the interaction of two spectrally distinct labelled proteins. Co-diffusion of differently labelled proteins through the very small confocal volume (~1 fL) gives a cross-correlation signal, and the intensity of the normalized signal amplitude is proportional to the concentration of oligomers. To test the formation of aggregation during the spontaneous folding reaction, we labelled

the two different populations of FLuc with either Alexa532 or Alexa647 succinimidyl ester. The denatured labelled proteins were mixed 1:1 over a wide concentration range (100 pM to 50 nM total protein). No cross-correlation signal was observed at concentrations below 10 nM, indicating that FLuc is monomeric below 10 nM during spontaneous folding (Figure 4.2a). FLuc simultaneously labelled with both fluorophores served as a positive control, mimicking the presence of dimeric aggregates. This species exhibited strong cross-correlation signal at 5 nM, demonstrating the sensitivity of the assay.

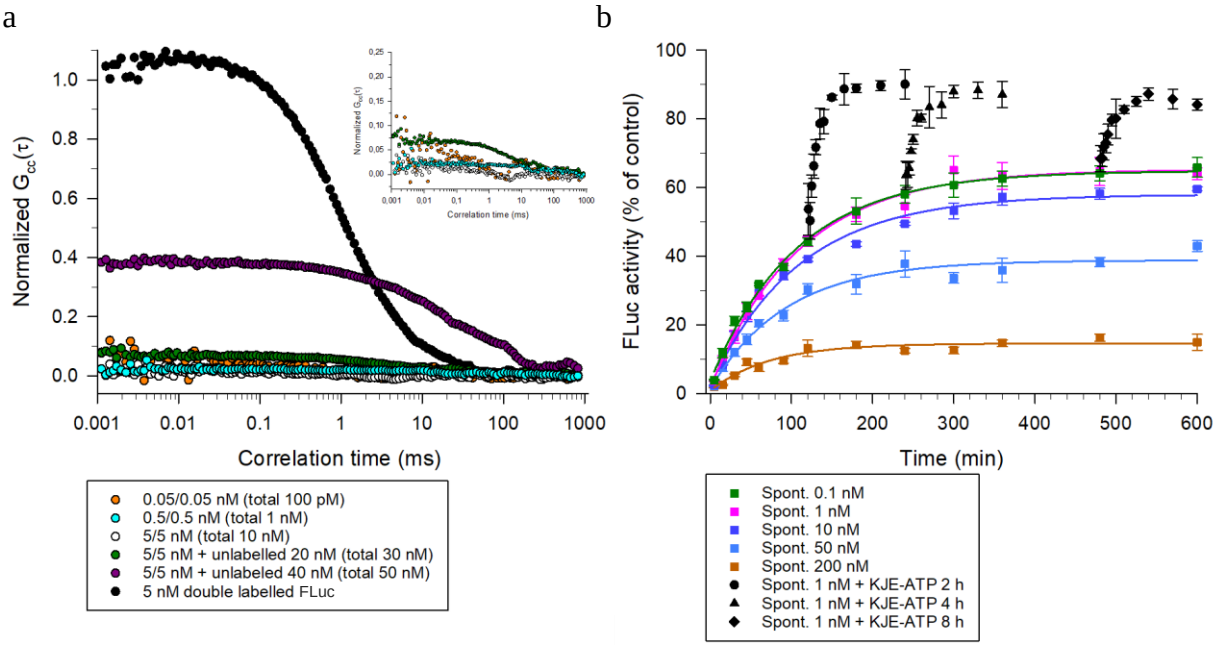


Figure 4.2 Spontaneous folding of FLuc in the absence of aggregation.
(a) Equimolar mixtures of 5 M GuHCl-denatured FLuc-Alexa532 or FLuc-Alexa647 were diluted to total concentrations of 100 pM to 10 nM in folding buffer and inter-molecular interactions monitored using dcFCCS. Double-labelled FLucNC (5 nM) was used (black) as a positive control mimicking dimeric aggregation. There is no cross-correlation signal below 10 nM. Additionally, FLuc aggregation was induced by increasing the unlabelled denatured FLuc. Samples were incubated at 25°C for 30 min before the analysis. Inset, the plot of dcFCCS was shown in a narrower scale corresponding to the y-axis. The dcFCCS data were recorded for 30 min. Representative data from three independent measurements are shown. (b) Concentration dependence of FLuc spontaneous folding was assayed using luminescence assay. GuHCl-denatured FLuc was diluted 100-fold into buffer, to final concentrations ranging from 0.1-200 nM (0.1 nM, 1 nM, 10 nM, 50 nM and 200 nM is green, pink, dark blue, blue, orange, respectively.). Additionally, rescue of spontaneous folding for FLuc (1 nM) was assayed by adding optimized concentration of KJE-ATP (black) (0.3 µM DnaK, 0.1 µM DnaJ, 0.5 µM GrpE and 5 mM ATP). Error bars represent s.d. (n=3). Spont., spontaneous folding.

After establishing the concentration at which FLuc remains monomeric during spontaneous folding, we assayed the spontaneous folding of FLuc at 25°C and at different concentrations (100 pM to 200 nM) by luminescence. As expected for a reaction that is limited by aggregation, we observed that reducing the concentration of FLuc during spontaneous folding increased the yield of folding up to ~60% (Figure 4.2b). Consistent with previous reports (Herbst et al., 1998; Herbst et al., 1997), the rate of spontaneous folding is slow with $t_{1/2}$ ~75 min in the absence of aggregation. Importantly, the rate of spontaneous folding remains unchanged below 10 nM, indicating that the cause of slow spontaneous folding is not transient aggregation, in support of the dcFCCS findings (Table 4.1, Figure 4.2a). The presence of ATP also did not affect the rate and yield of spontaneous folding (Figure 4.3a). Note that, to prevent FLuc adsorption to tube walls, we used an optimized amount of non-ionic detergent (0.05% v/v Tween 20) (Figure 4.3b). In conclusion, FLuc folds spontaneously at low concentrations, but with a very slow rate.

The limited yield of spontaneous folding (~60%) indicates that FLuc populates a kinetically trapped intermediate(s) during spontaneous folding that is not able to convert to the native state within 10 h. To test whether this kinetically trapped intermediate(s) of FLuc remains competent for folding by KJE in the absence of aggregation, we added KJE-ATP to spontaneous folding reactions at different time points, 2 h, 4 h or 8 h. Chaperone addition resulted in the rapid conversion of the kinetically trapped state(s) to the native state (Figure 4.2b).

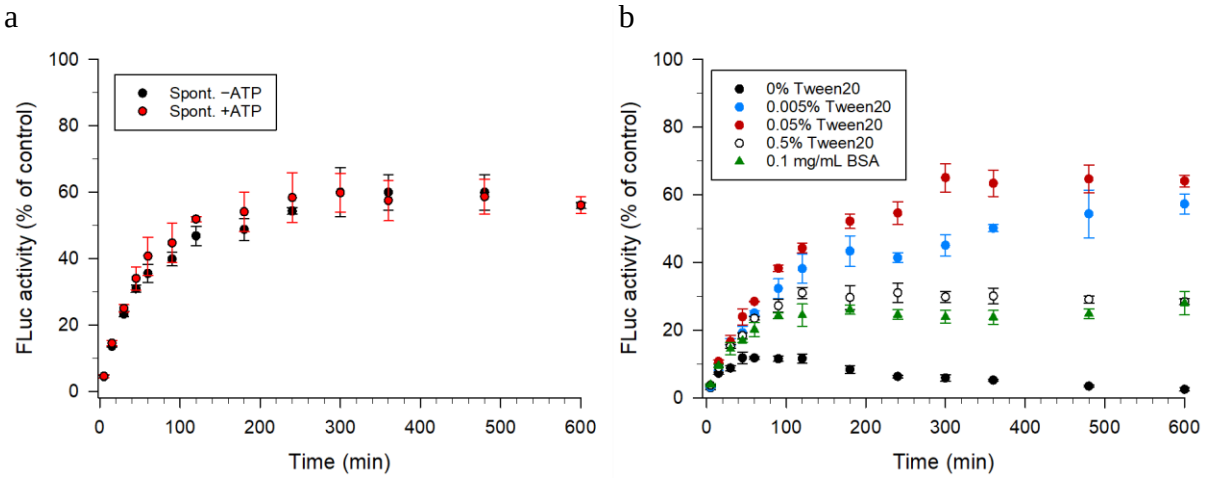


Figure 4.3 The effect of ATP and the non-ionic detergent (Tween 20) on FLuc spontaneous folding.

(a) ATP does not impact the rate or the yield of FLuc spontaneous folding. Refolding of 5 M GuHCl denatured FLuc (final concentration of 1 nM) was diluted into folding buffer. Refolding was assayed using the luminescence assay. (b) Spontaneous refolding of 5 M GuHCl denatured FLuc (final concentration of 1 nM) was assayed in a buffer supplemented with different amounts of Tween 20 (0-0.5% v/v) or BSA (0.1 mg/mL).

We next tested whether spontaneous folding could be improved by phenylbenzothiazole (PBT), a luciferin analogue that stabilizes native FLuc (Thompson et al., 1991). Since PBT inhibits FLuc activity, we used an orthogonal approach to assay folding in the absence of aggregation. To measure the kinetics of folding in the presence of PBT, we used fluorescence correlation spectroscopy (FCS), used for the determination of mobility-related parameters, such as concentration of the labelled molecules, diffusion time and diffusion coefficient. The GroEL chaperonin selectively binds non-native FLuc-Alexa647, confirmed by size-exclusion chromatography (Figure 4.4a), and significantly slows its diffusion as measured by FCS (Figure 4.4b). Non-native FLuc could therefore be quantified by addition of GroEL at different times after initiating folding, followed by measurement of the diffusion time of FLuc-Alexa647 by FCS. Addition of PBT improved the yield of spontaneous folding to ~75%, but did not significantly affect the folding rate (Figure 4.4c). In the absence of aggregation, FLuc therefore refolds spontaneously to almost full yield, but at least 20-fold slower than in the presence of DnaK and its co-chaperones. These data demonstrate that beyond its well-described role in preventing protein aggregation, the DnaK chaperone system can accelerate client protein folding.

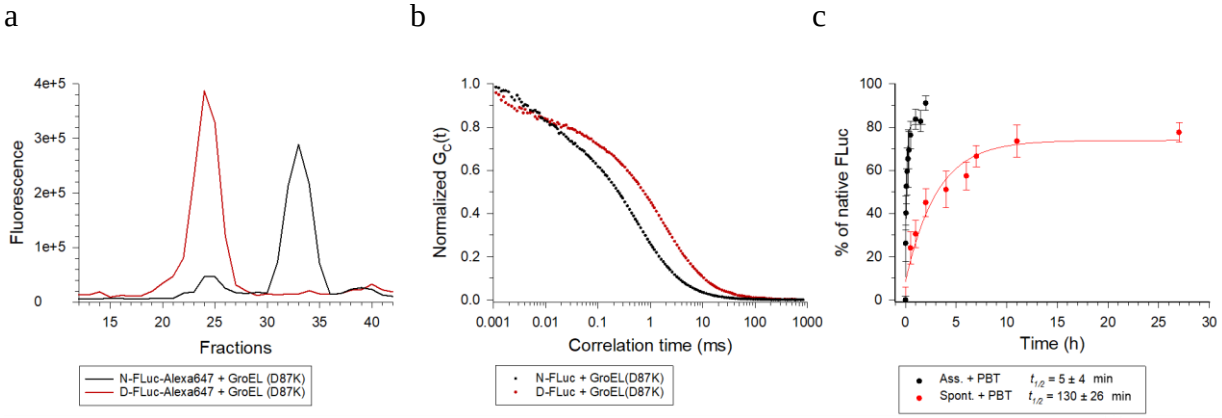


Figure 4.4 Spontaneous and KJE-assisted folding of FLuc using FCS.
(a and b) GroEL (D87K mutant) can bind to denatured FLuc but not native FLuc. We assayed GroEL binding to FLuc using (a) size-exclusion chromatography (SEC) or (b) FCS. In SEC, GroEL-bound FLuc was eluted at earlier fractions than that of native FLuc, which is independent

of the presence of GroEL. (b) Slower diffusion was observed for GroEL-bound denatured FLuc (D = 31.50 ± 1.25 μm^2/s), whereas the diffusion rate of the native FLuc (D = 57.87 ± 1.59 μm^2/s) was not affected by GroEL (D87K) (D = 57.35 ± 1.36 μm^2/s). Diffusion time of native FLuc and GroEL (D87K)-bound denatured FLuc were monitored by FCS. (c) Spontaneous and KJE-assisted refolding of FLuc labelled with Alexa647 were measured at the final FLuc concentration of 500 pM. Refolding was stopped by the addition of GroEL (D87K) (1 μM) at different time points. Refolding kinetics were calculated based on the difference in mean diffusion time between GroEL-bound FLuc and native FLuc. The mean diffusion time was calculated from the auto correlation data of labelled FLuc. Error bars represent standard deviation from at least three independent experiments.

We next analysed the role of ATP on the accelerated folding reaction. In the absence of ATP, KJE refolding occurred with a rate similar to that of spontaneous folding (Figure 4.5), whereas we observed slightly lowered yield (~40%), likely because of KJ binding to folding intermediates. The addition of ATP at different time points induced rapid FLuc folding, indicating that the KJE system uses the energy of ATP hydrolysis to shift the folding equilibrium towards to the native state, consistent with previous proposals (Goloubinoff et al., 2018).

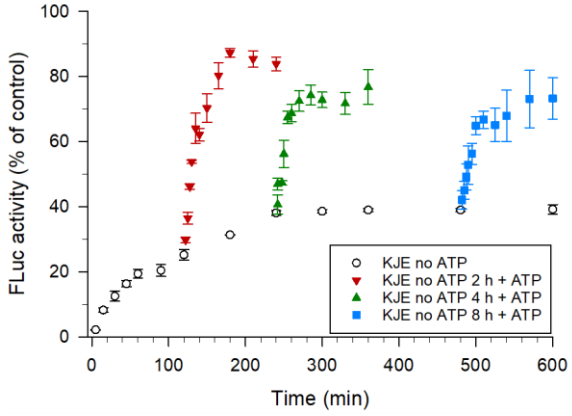


Figure 4.5 ATP dependence of chaperone-accelerated folding of FLuc.
KJE-assisted folding of FLuc (1 nM) was assayed as in Figure 4.2b in the presence of KJE, but in the absence of ATP. 5 mM ATP was added to the reaction after 2, 4 or 8 h. Error bars represent s.d (n=3).

Table 4.1 Rates and yields of KJE-assisted and spontaneous FLuc folding.

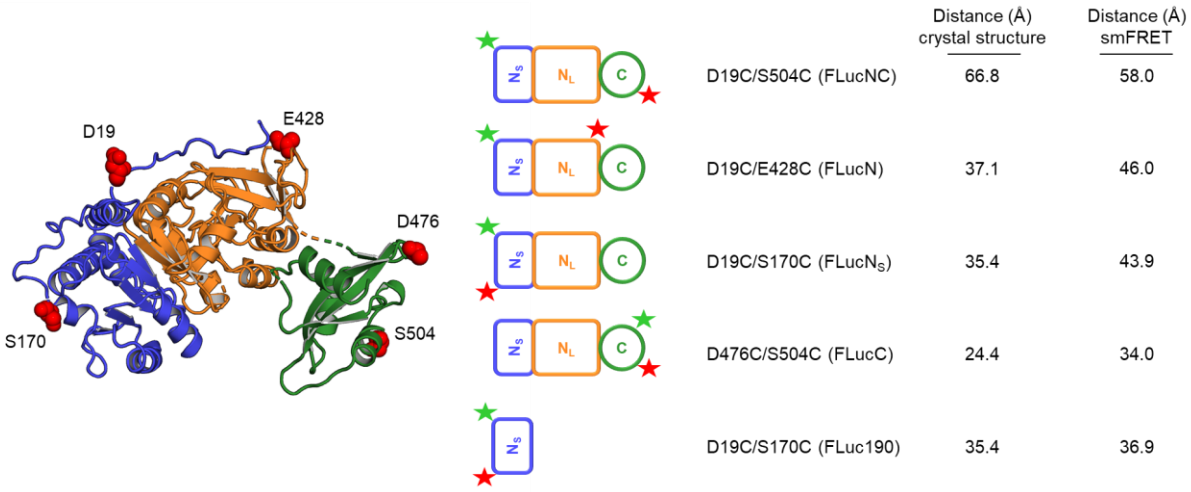
	[FLuc] (nM)	Temperature (T) °C	[DnaK] (µM)	[DnaJ] (µM)	[GrpE] (µM)	Rate (min)	Yield (% of control)
Ass. folding	100	25	3	1	1.5	4.6 ± 0.3	89 ± 1
	0.1	25	0.3	0.1	0.5	4.5 ± 0.3	87 ± 2
	0.01	37	1	0.1	0.5	2.9 ± 0.3	86 ± 6
Spont. folding	200	25	-	-	-	50 ± 9	14 ± 1
	100	25	-	-	-	75 ± 7	23 ± 0.8
	50	25	-	-	-	65 ± 8	38 ± 2
	10	25	-	-	-	75 ± 6	56 ± 2
	1	25	-	-	-	75 ± 4	64 ± 2
	0.1	25	-	-	-	74 ± 4	61 ± 1
	0.01	37	-	-	-	38 ± 3	15 ± 0.4

Taken together, these findings clearly demonstrate that the KJE chaperone system can accelerate folding by modulating the folding energy landscape beyond aggregation prevention.

4.2 Inter-domain misfolding of FLuc causes slow spontaneous folding

To analyse the cause of slow spontaneous folding and understand the conformation of kinetically trapped intermediate(s), we performed single-molecule fluorescence resonance energy transfer (smFRET) combined with pulsed interleaved excitation (PIE), which is based on sequential excitation of both fluorophores. A key strength of smFRET is to provide information about heterogeneity in folding reactions, thus distinguishing different protein conformational states present in a reaction. Hence, smFRET is a powerful tool to simultaneously monitor chaperone-bound, native and kinetically trapped state(s) during folding reactions. To probe the different regions of FLuc, we designed four double-cysteine mutants and fluorescently labelled them with Atto532 and Atto647N as FRET donor and acceptor, respectively (R_0~46 Å). SmFRET experiments were performed at 50 pM concentration of FLuc to ensure the presence of, at the majority of bursts, one labelled species in the small confocal volume (~1 fL) at any one time.

FLucNC (D19C/S504C), consisting of one fluorophore in the N-terminal domain and one in the C-terminal domain, provides information on overall FLuc folding. FLucN (D19C/E428C) with both fluorophores in the N-domain, FLucN_S (D19C/S170C) with both fluorophores in the N-subdomain, and FLucC (D476C/S504C) with both fluorophores in the C-domain, allow the folding of the N-domain, N-subdomain and C-domain to be analysed, respectively (Figure 4.6). Each construct exhibits a well-defined and homogenous distribution of FRET efficiencies (f_E) (Figure 4.7a-d) in the native state, which are consistent with the expected distance between fluorophore pairs according to the crystal structure of FLuc. Note that fluorophore labelling did not influence FLuc activity and foldability (Figure 4.7e-h).



	Distance (Å) crystal structure	Distance (Å) smFRET
D19C/S504C (FLucNC)	66.8	58.0
D19C/E428C (FLucN)	37.1	46.0
D19C/S170C (FLucN_S)	35.4	43.9
D476C/S504C (FLucC)	24.4	34.0
D19C/S170C (FLuc190)	35.4	36.9

Figure 4.6 Structure of FLuc showing the labelling positions.
Indicated positions are mutated to cysteine for site-specific maleimide labelling with Atto532 and Atto647N. The estimated distances between dye pairs are calculated from the crystal structure of FLuc (PDB: 1LCI) and from the smFRET experiments.

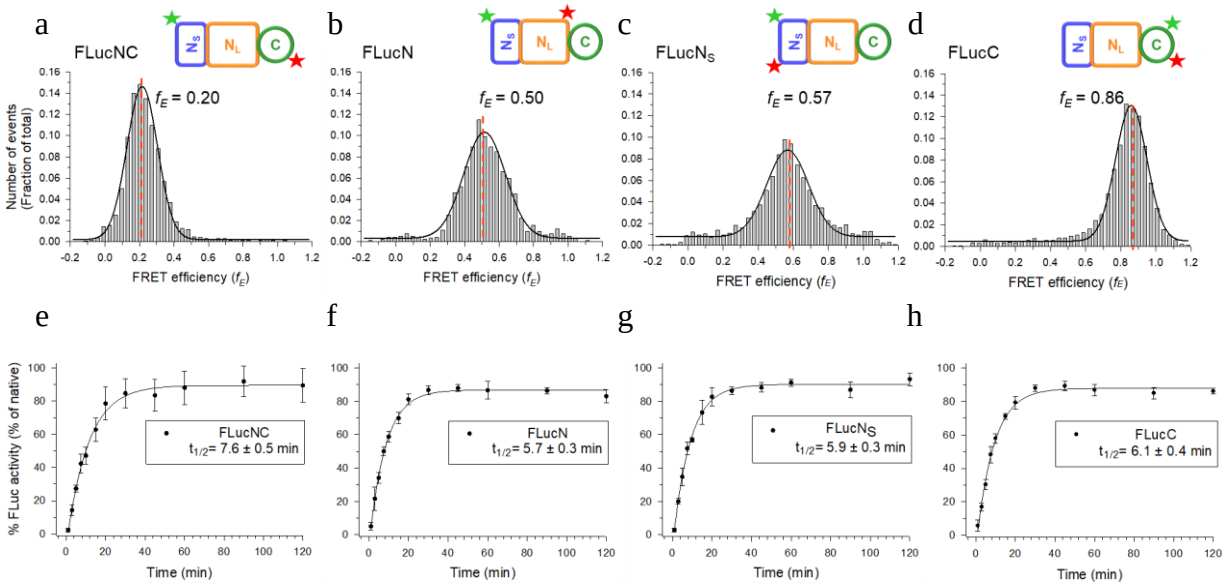


Figure 4.7 smFRET efficiency (f_E) histograms of native double-labelled FLuc constructs. (a-d), Measurements were performed with 50 pM FLuc in the presence of 50 µM PBT and 5 mM ATP. (a) FLucNC, (b) FLucN, (c) FLucN$_S$, and (d) FLucC. A dashed vertical red line represents the f_E corresponding to the native state for each construct. SmFRET data for each plot were recorded for 1 h and a minimum of 1000 events were collected. The plots are representative examples of three independent experiments. (e-h), Labelled FLuc proteins are competent to fold. (e) FLucNC, (f) FLucN, (g) FLucN$_S$ and (h) FLucC were diluted to 1 nM from 5 M GuHCl into folding buffer containing 0.3 µM DnaK, 0.1 µM DnaJ, 0.5 µM GrpE and 5 mM ATP. Folding at 25°C was monitored by luminescence assay. Error bars represent s.d. (n=4).

At early times of spontaneous folding, upon dilution from denaturant (0 h time point), FLucNC, FLucN, and FLucN$_S$ populated broad FRET distributions with high f_E values (Figure 4.8a-d), indicating the presence of intermediate states that are conformationally more compact than the native state. These kinetically trapped intermediates likely contain non-native interactions, consistent with misfolded conformations detected previously for FLuc in optical tweezer experiments (Mashaghi et al., 2014b). At later time points of spontaneous folding, the high f_E distribution slowly shifted towards the f_E distribution of the native state. Additionally, we tested the rate and the yield of folding according to simulated data (Gaussian) of the smFRET distributions and found a $t_{1/2}$ of ~70–85 min (Figure 4.8e-h), consistent with the results of the luminescence based folding assay. Note that a slightly higher yield than in the luminescence assay was observed, because the native state of FLuc was stabilized using the substrate analogue of luciferin, PBT.

FLucC folded about 10-times faster than the other constructs without populating any detectable intermediates (Figure 4.8d), suggesting that the folding of C-domain is independent of the N-domain. Taken together, these data suggest that a compact intermediate state(s) arises from a non-native arrangement within the N-domain, whereas the C-domain folds rapidly upon dilution from denaturant.

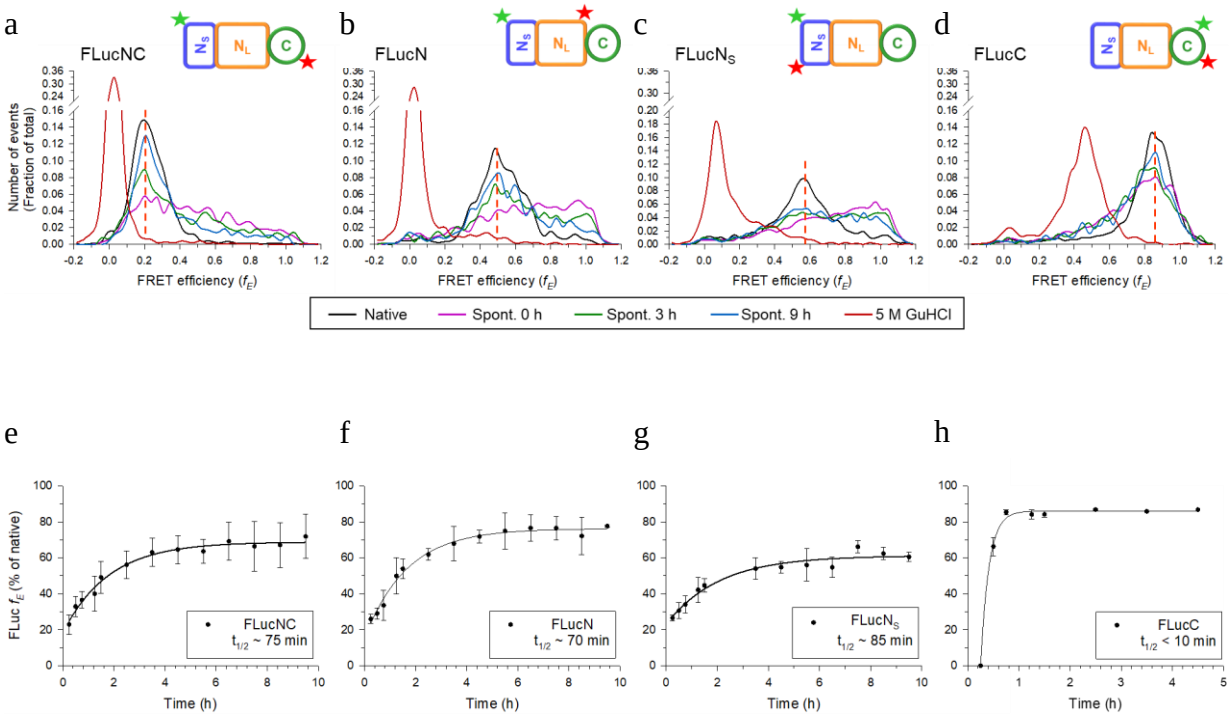


Figure 4.8 Slow spontaneous folding of FLuc due to inter-domain misfolding.
(a-d), Spontaneous folding of 5 M GuHCl-denatured double-labelled FLuc (50 pM) was measured using smFRET. Reactions contained 50 µM PBT and 5 mM ATP. SmFRET was recorded for 1 h, either immediately upon dilution from denaturant (0 h) or at the indicated folding time points. A dashed vertical red line represents the f_E corresponding to the native state for each construct. The plots are representative examples of three independent experiments. A minimum of 1000 events were collected. (e-h), Kinetics of spontaneous FLuc folding derived from smFRET experiments. Conversion of the f_E distribution of denatured (e) FLucNC, (f) FLucN, (g) FlucNS, and (h) FLucC towards the f_E of the native state was calculated by dividing the area of the f_E histogram corresponding to the native protein by the total area of the f_E histogram at each folding time point (0-10 h). Time points correspond to the midpoint of each recording interval. Error bars represent s.d. (n=3).

To understand whether the N_S-domain misfolds itself or non-natively interacts with the N_L-domain, we prepared isolated N_S-domain FLuc (FLuc190) (Figure 4.9a). Interestingly, FLuc190 folds rapidly in the absence of the N_L-domain and C-domain (Figure 4.9b), indicating that the misfolded state arises from non-native interactions between the N_S-domain and N_L-domain. This result is consistent with the ability of the N_S-domain to fold co-translationally (Frydman et al., 1999). Taken together, these data reveal that misfolding between N_S- and N_L-domains is the rate-limiting step of slow spontaneous folding.

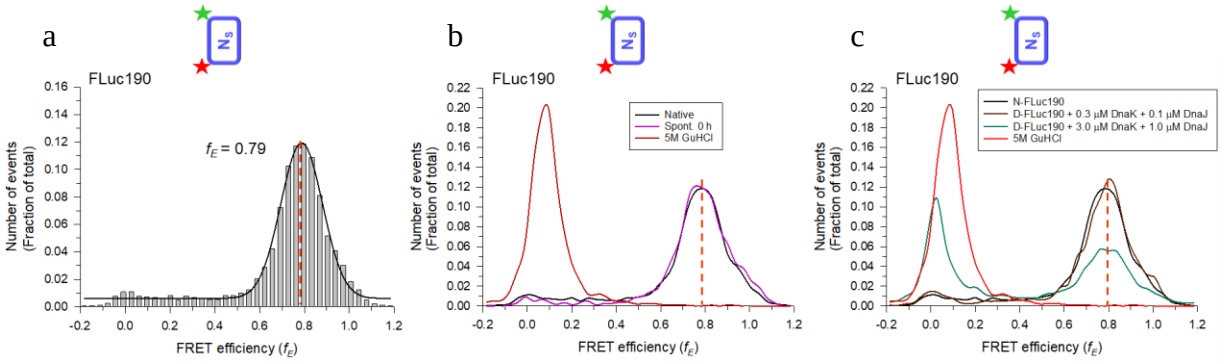


Figure 4.9 Efficient folding of isolated FLuc190.
(a) smFRET efficiency (f_E) histogram of the N_S-subdomain (FLuc190). (b) Spontaneous folding of GuHCl-denatured FLuc190 (50 pM). Data were recorded for 30 min. (c) Interaction of FLuc190 with KJ. FLuc190 was denatured in 5 M GuHCl, then diluted to 50 pM in folding buffer containing 0.3 µM DnaK/0.1 µM DnaJ or 3 µM DnaK/1 µM DnaJ, and analysed by smFRET. All reactions contained 5 mM ATP and 50 µM PBT. A dashed vertical red line represents the f_E corresponding to the native state of FLuc190. The plots are the representative of three independent experiments. SmFRET data for each plot were recorded for 30 min or 1 h and a minimum of 1000 events were collected.

4.3 Efficient folding of FLuc by KJE-ATP involves unfolding of misfolded states

Our luminescence-based assays explicitly showed that the KJE chaperone system dramatically accelerates FLuc folding. To understand the underlying mechanism of this reaction, we performed a smFRET analysis of FLuc folding reactions in the presence of chaperones. In the established model for the chaperone function of KJE-ATP (Figure 4.1b), DnaK forms a stable complex with the substrate in the presence of ATP and DnaJ. GrpE triggers the release of bound substrate and thus initiates folding (Figure 4.1b). Accordingly, to observe the DnaK-bound conformation of FLuc, we first diluted completely denatured FLuc into folding buffer containing DnaK, DnaJ (KJ) and ATP but not GrpE. Interestingly, DnaK-bound FLucNC exhibited a very

low f_E (Figure 4.10a) similar to the f_E values of the denatured proteins, indicating that DnaK-binding results in a highly expanded conformation. As expected, this action of DnaK required the function of DnaJ and ATP, as DnaK cannot bind to FLuc in the absence of DnaJ or ATP (Figure 4.10a). Note that binding of DnaJ alone does not lead to the formation of the expanded conformation (Figure 4.10a) and KJ-ATP do not bind native FLucNC (N-FLucNC) (Figure 4.10b). Similar to DnaK-bound FLucNC, DnaK binding to FLucN and FLucN$_S$ led to a very low f_E. In contrast, DnaK binding to FLucC did not result in the expanded FLucC, consistent with its rapid spontaneous folding. Interestingly, isolated SubN (FLuc190) acquires its native conformation rapidly upon dilution from denaturant without exhibiting any expanded conformation, suggesting that FLuc190 folds faster than DnaK binding (Figure 4.9c). Only the presence of excess amount of KJ-ATP can bind FLuc190 and form an expanded conformation.

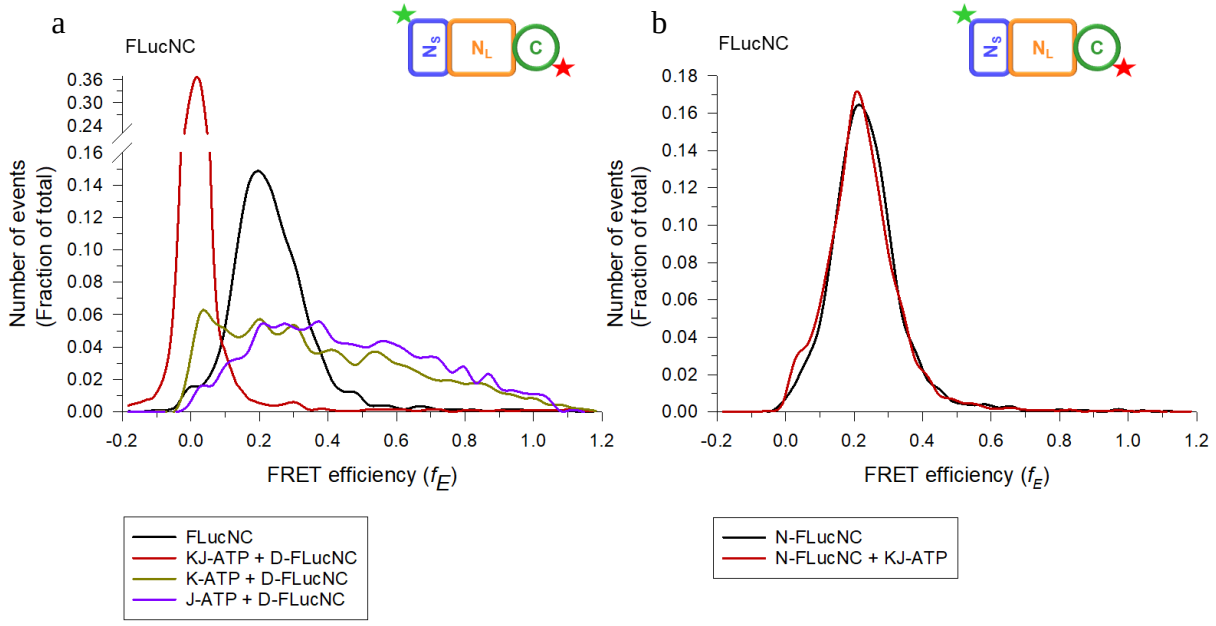


Figure 4.10 Requirement of both DnaK and DnaJ for chaperone-mediated unfolding.
(a) GuHCl-denatured FLucNC (D-FLucNC) was diluted to 50 pM in folding buffer containing either 0.3 µM DnaK, 0.1 µM DnaJ or a mixture of both chaperones. ATP (5 mM) was always present. SmFRET was recorded for 30 min. Native FLucNC (N-FLucNC) is also shown for comparison. (b) Incubation of 50 pM native FLucNC in a buffer containing both 0.3 µM DnaK, 0.1 µM DnaJ and 5 mM ATP. A minimum of 1000 events were collected. One representative measurement of three independent repeats is shown.

Addition of GrpE to the DnaK-bound, expanded FLuc (FLucNC, FLucN and FLucN$_S$) results in the formation of three different smFRET distributions at very early time points (0-15 min), suggesting the presence of at least three different populations (Figure 4.11a-c). We observed a minor population of DnaK-bound expanded FLuc and a major population of native FLuc. Interestingly, we also saw a minor population of misfolded FLuc similar to that at early time points of spontaneous folding. The reaction reached almost full yield of native state within 1 h after the addition of GrpE, suggesting that non-native FLuc is being recycled by the KJE chaperone system for proper folding.

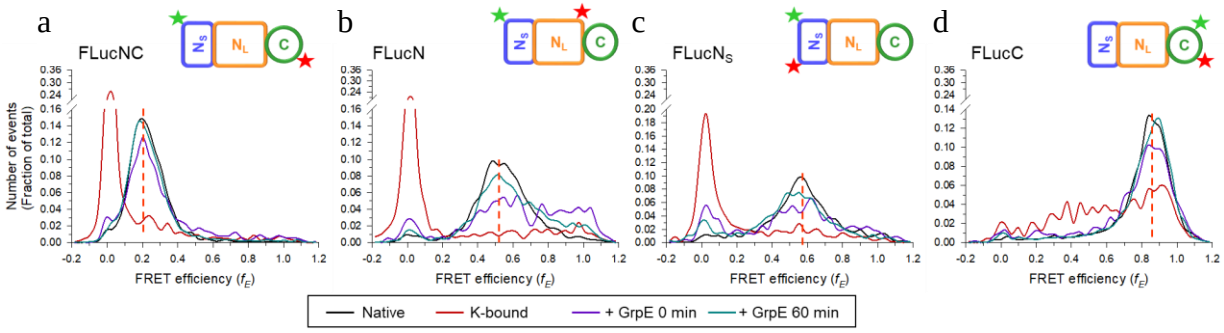


Figure 4.11 Assisted folding of FLuc analysed by smFRET.
(a-d), smFRET efficiency histograms measured during chaperone-assisted refolding of FLuc, for FRET pairs probing (a) FLucNC, (b) FLucN, (c) FLucN$_S$, (d) FLucC. 5 M GuHCl-denatured FLuc was diluted to 50 pM in folding buffer containing 0.3 µM DnaK, 0.1 µM DnaJ, 5 mM ATP and 50 µM PBT, and folding reactions were initiated with 0.5 µM GrpE. At each folding time point, smFRET was recorded for 15 min and a minimum of 500 events were collected. A dashed vertical red line represents the f_E corresponding to the native state for each construct. The plots are the representative of three independent experiments.

Next, to understand the fate of misfolded intermediates formed at the early time point of the assisted folding reaction, we tested whether the KJE chaperone system can recognized these states. We first generated misfolded states of FLucNC by diluting denatured FLucNC into buffer without chaperones, as we did to measure spontaneous folding (Figure 4.12). After 30 min, we added KJ. We observed a major amount of DnaK-bound expanded FLuc and a minor amount of native FLuc formed during the 30 min of spontaneous folding. This population did not shift as KJ does not bind native FLuc (Figure 4.12). The addition of GrpE resulted in rapid conversion of DnaK-bound, expanded FLuc to the native state (Figure 4.12). DnaK therefore accelerates the folding of FLuc via formation of an expanded state, in a mechanism that involves unfolding of kinetically trapped misfolded states. Taken together, the misfolded state, arising from non-native

interaction within the N domain, requires conformational expansion or unfolding by DnaK for efficient folding.

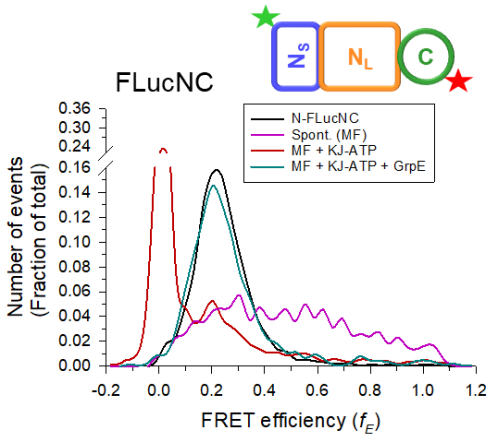
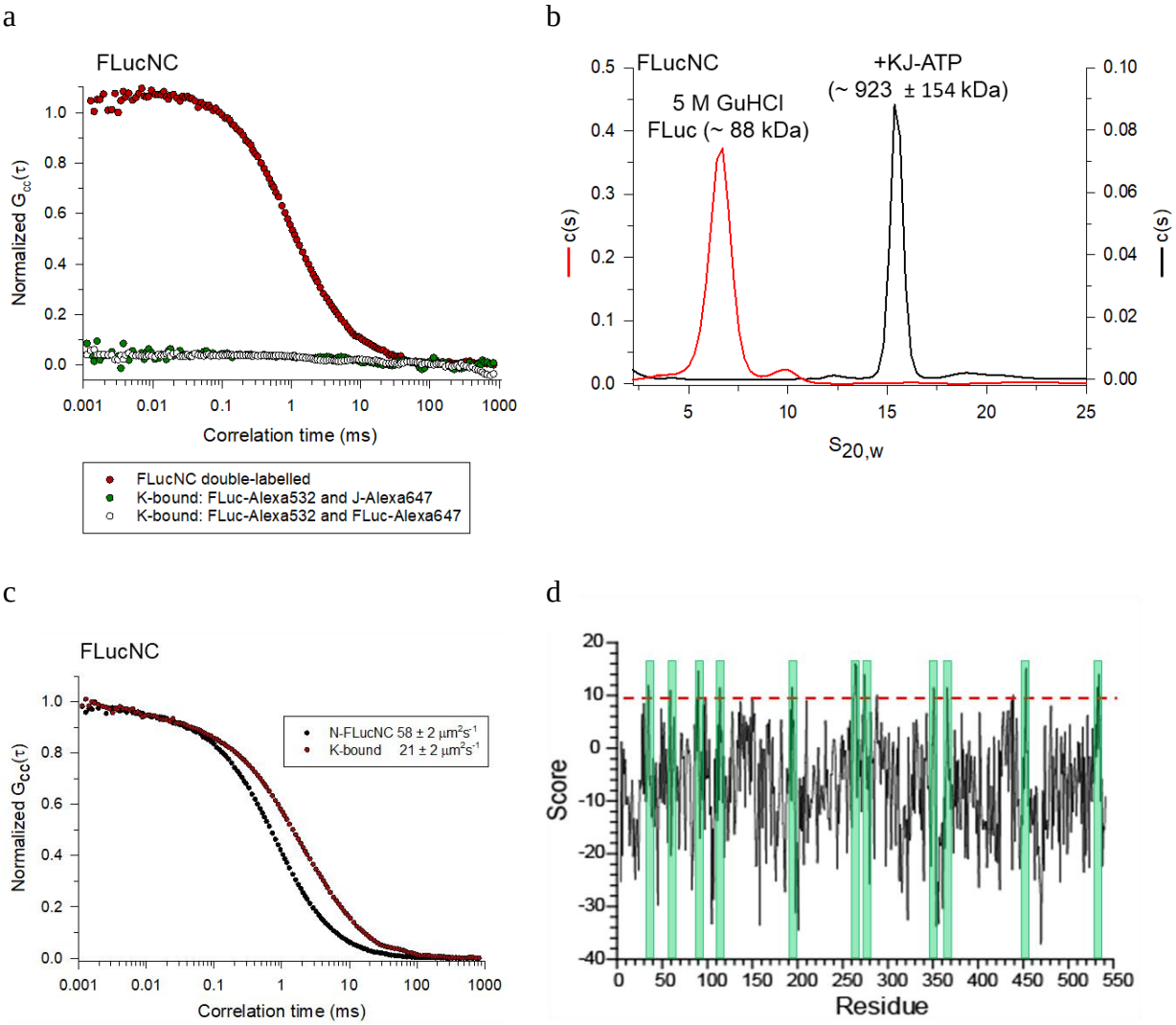


Figure 4.12 Rescue of the misfolded states by the KJE chaperone system.
GuHCl-denatured FLuc was diluted to 50 pM in buffer containing 5 mM ATP and 50 µM PBT as in Figure 4.8, and the N-C interdomain distance was probed by smFRET as in other smFRET experiments. DnaK (0.3 µM) and DnaJ (0.1 µM) were added after 1 h, followed by GrpE (0.5 µM) to initiate the chaperone-assisted folding reaction. SmFRET data for each plot were recorded for 30 min and a minimum of 500 events were collected. One representative measurement of three independent repeats is shown. MF, misfolded states; Spont. Spontaneous folding.

4.4 Multiple DnaK bind per FLuc in the expanded conformation

After defining the DnaK-bound expanded conformation of FLuc, we next aimed at answering the following two questions: (1) Does the DnaK-FLuc complex contain more than one FLuc or DnaJ? and (2) What is the size of the complex? We tested whether there is more than one FLuc in the complex using dcFCCS. Using an equimolar mixture of FLuc-Alexa532 and FLuc-Alexa647, we did not observe a cross-correlation signal (Figure 4.12a), indicating that the DnaK-FLuc complex contains only FLuc. Additionally, to detect the presence of DnaJ in the complex, we performed dcFCCS with FLuc-Alexa532 and DnaJ-Alexa647. Again, no significant cross-correlation signal was observed, indicating that the DnaK-FLuc complex does not include DnaJ (<5% of complexes) (Figure 4.12a). We determined a mass of ~900 kDa for the DnaK-FLuc complex by analytical ultracentrifugation (AUC) (Figure 4.12b). We further confirmed the size of the complex using FCS (Figure 4.12c). The diffusion coefficient measured by FCS can be converted into a hydrodynamic radius using the Stokes-Einstein equation. We found a diffusion

coefficient of 21 $\mu m^2 s^{-1}$ by FCS, from which we calculated that the size of the complex is ~1000 kDa (Figure 4.12c). The AUC and FCS results are highly consistent, and indicate that ~12 DnaK can be bound per FLuc. Additionally, we used a specific prediction algorithm for identifying the possible DnaK binding sites and, consistent with our experimental data, the number of high confidence DnaK binding motifs was determined as 13 (Van Durme et al., 2009a) (Figure 4.13d-e).



e

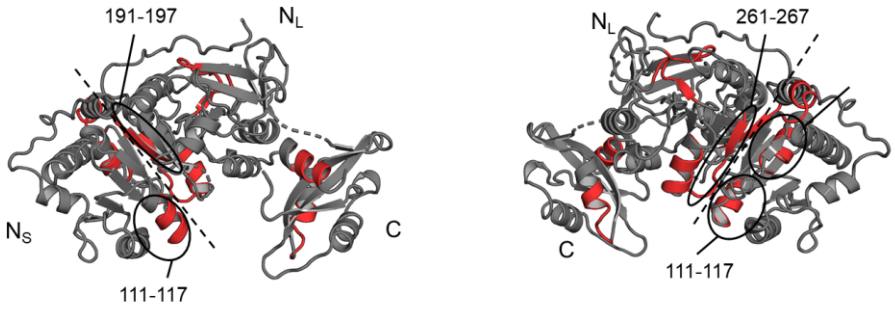


Figure 4.13 Up to 12 DnaK bind per FLuc to form the DnaK-FLuc complex
(a) Chaperone-bound FLuc contains only one molecule of FLuc and does not contain DnaJ. Wild-type FLuc N-terminally labelled with Alexa53 was diluted to 10 nM in a buffer containing 300 nM DnaK and 90 nM DnaJ, 10 nM DnaJ-Alexa647 and 5 mM ATP. Alternatively, 50 pM denatured FLuc-Alexa532 was mixed with 50 pM denatured FLuc-Alexa647 and diluted into folding buffer containing 0.3 µM DnaK, 0.1 µM DnaJ and 5 mM ATP. Double-labelled FLucNC (5 nM) was used as a positive control. One representative measurement of three independent repeats is shown. (b) Analytical ultracentrifugation was used to analyse the size of DnaK-bound FLucNC labelled with Atto532. FLucNC was unfolded in 5 M GuHCl. As a control, the monomeric size of FLucNC was analysed in 5 M GuHCl. To determine the size of DnaK-bound FLucNC, 5 M GuHCl denatured FLucNC was diluted into buffer containing 3 µM DnaK, 1 µM DnaJ. The final concentration of FLucNC was 200 nM. The sedimentation profile (absorption at 532 nm) is shown in black and red for denatured FLucNC and DnaK-bound FLucNC complex, respectively. The X-axis indicates the sedimentation coefficient and the y-axis the sedimentation coefficient distribution. The apparent mass of the denatured-FLucNC and the DnaK-bound FLuc complex are indicated. Representative measurements of 3 independent experiments are shown. (c) The diffusion coefficients of native FLucNC and the DnaK-bound FLucNC were measured using FCS. Experiments were performed with 50 pM FLucNC-Atto532. The complex was formed by dilution of FLucNC to a final concentration of 50 pM in buffer containing 0.3 µM DnaK and 0.1 µM DnaJ and 5 mM ATP. The size of the confocal volume was measured by testing free Atto655 dye. The diffusion coefficients of native and DnaK-bound FLuc are indicated. Representative measurements of 3 independent repeats are shown (d) Predicted DnaK binding sites on FLuc were analysed by LIMBO. Peptide regions with a score above 10 (red dashed line) are considered to be high confidence DnaK binders (green lines) with a false discovery rate <1 %. (e) The structure of FLuc with predicted DnaK interaction sites from (d) coloured red. The interface between the N_S and N_L subdomains is indicated with a dashed line.

4.5 DnaK commits a fraction of bound FLuc to fast folding trajectory

In our smFRET experiment we observed that DnaK accelerates folding by unfolding misfolded FLuc through cycles of binding and release. Next, we tested whether DnaK commits expanded FLuc to a different folding pathway than that of spontaneous folding. To address this question, we inhibited the rebinding of FLuc to DnaK after GrpE-stimulated release using the substrate-mimicking model peptide GSGNRLLLTG (NR) (Mayer et al., 2000a; Pierpaoli et al., 1998; Zhu et al., 1996). First, we tested whether excess NR can completely block DnaK. To answer this question, we added excess NR to KJ-ATP prior to the addition of denatured FLuc. Folding occurred at the slow rate of spontaneous folding, indicating that binding to DnaK was effectively blocked (Figure 4.14). Note that the slightly reduced folding yield (~40%) as compared to spontaneous folding occurred because of the presence of DnaJ, which is not blocked by NR (Kellner et al., 2014). Additionally, to assess the peptide-inhibition of DnaK in another control experiment, we added excess NR to an ongoing reaction of KJE-assisted folding after 5 min, which resulted in an immediate shift from fast to slow folding (Figure 4.14). Therefore, these experiments indicate that the optimized model DnaK peptide, NR, completely blocks the (re)binding of substrate to DnaK. We next asked whether FLuc folding is accelerated in a single round of chaperone action. To address this question, we first formed the DnaK-FLuc complex in the absence of GrpE and NR. To prevent rebinding of FLuc to DnaK and to initiate FLuc release by GrpE, we added NR and GrpE, respectively. Interestingly, in this case, the folding kinetics are biphasic, with a fast phase ($t_{1/2}$ = 4 ± 1 min) similar to the rate of chaperone-assisted folding, and a slow phase ($t_{1/2}$ = 49 ± 8 min) corresponding to spontaneous folding (Figure 4.14). Additionally, we observed approximately 15% of released FLuc reached the native state after a single round of chaperone action, whereas about 3% obtained the native structure in the equivalent time of spontaneous folding, indicating that ~10-fold more native FLuc is produced in a single cycle of chaperone action than folding from the GuHCl-denatured state. These results show that the KJE-ATP system provides access to a fast folding trajectory for a fraction of FLuc, with a single round of DnaK-driven conformational expansion being sufficient to accelerate FLuc folding. We suggest that fast folding may be achieved by a mechanism of step-wise release of bound DnaK molecules from FLuc, by preventing the non-native interactions between N_S- and N_L- domains that are rate-limiting for spontaneous folding (Imamoglu et al., 2020).

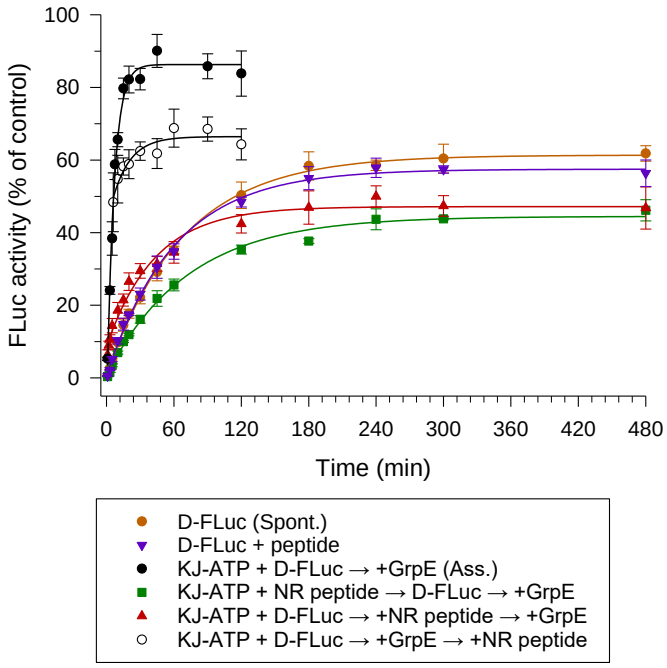


Figure 4.14 The KJE chaperone system directs the folding pathway of FLuc towards rapid folding.
Single round chaperone action was analysed using DnaK-binding peptide. FLuc folding was followed by luminescence assay. Rebinding of FLuc to DnaK was inhibited using 125 µM peptide NR (sequence: N-GSGNRLLLTG-C) which binds the substrate-binding groove of DnaK. GuHCl-denatured FLuc was diluted to 100 pM in folding buffer containing 0.3 µM DnaK, 0.1 µM DnaJ and 5 mM ATP. Peptide NR was then added, before initiating the folding reaction with 0.5 µM GrpE. As controls, peptide NR was added to KJ-ATP before adding FLuc, or after 5 min of chaperone-assisted folding. Spontaneous folding was measured in the absence or in the presence of the same concentration of peptide NR. As additional controls, NR was either omitted (black circles) or added to KJ-ATP before FLuc (green squares), or after 5 min of chaperone-assisted folding (white circles). Error bars represent s.d. (n = 3).

4.6 KJE-ATP shifts the folding equilibrium of proteins towards the native state

Previous studies suggested that the Hsp70 system assists folding through a passive mechanism in which multiple cycles of substrate binding and release will prevent aggregation (Hartl and Hayer-Hartl, 2009). However, our results showed that the Hsp70 system can accelerate folding, indicating an active mechanism. We hypothesized that KJE-mediated acceleration of protein folding, but not unfolding, would shift the folding equilibrium of FLuc towards the native state. To test this hypothesis, we analysed folding under denaturing conditions, where the native state

is less stable than the misfolded states. We destabilized the native state of FLuc by raising the temperature to 37°C, and measured the luminescence activity in the absence or presence of chaperones (Figure 4.15a). Note that folding was monitored at a very low concentration of FLuc (10 pM) to exclude aggregation (Figure 4.15b), as demonstrated by dcFCCS. In the absence of chaperones, incubation at 37°C caused a loss of luminescence activity over time ($t_{1/2}$ = 51 ± 6 min) (Figure 4.15a). Strikingly, in the presence of KJE-ATP, native luminescence activity was maintained for at least 3 h at 37°C (Figure 4.15a). Spontaneous folding at 37°C was inefficient with a yield of only ~20% (Figure 4.15a). Addition of KJE to the spontaneous folding reaction after 1 h rapidly converted FLuc to the native state with almost full yield (Figure 4.15a), similar to the rescue experiments at 25°C (Figure 4.2b). In contrast to inefficient spontaneous folding, KJE-assisted folding reached a yield of ~85% ($t_{1/2}$ = 2.9 ± 0.3 min) (Figure 4.15a).

To further monitor the role of KJE-ATP under denaturing conditions, we performed smFRET with FLucNC. When we added KJ to a spontaneous folding reaction of FLuc at 37°C, we observed the formation of expanded FLuc conformations, which converted to the native state upon addition of GrpE, even under denaturing conditions (4.15c). SmFRET also revealed that native FLucNC populated a broad conformational distribution at 37°C (Figure 4.15d), similar to the misfolded state observed during spontaneous folding from the chemically-denatured state (Figure 4.12). When native FLucNC was incubated with KJ and incubated at 37°C, the f_E distribution first shifted to that of the expanded DnaK-bound state, before rapidly converting to the native state upon addition of GrpE (Fig. 4.15d). Taken together, and consistent with recent studies (Goloubinoff et al., 2018; Zhao et al., 2019a), these results demonstrate that the KJE chaperone system shifts the folding equilibrium of FLuc towards the native state under denaturing conditions by using the energy of ATP hydrolysis.

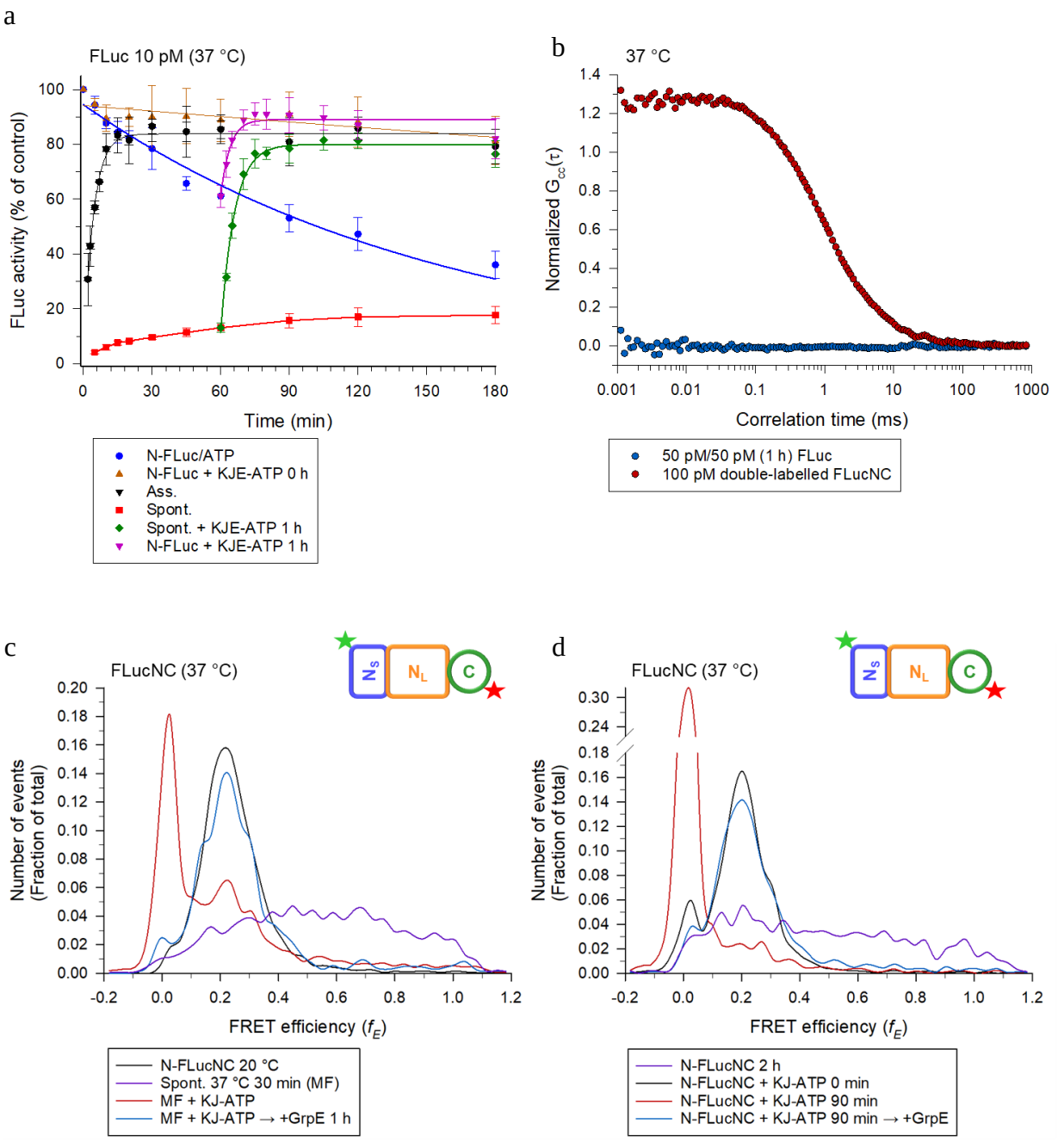


Figure 4.15 The native state of FLuc maintained by KJE- accelerated folding.
(a) Luminescence activity of FLuc (10 pM) was assayed at 37°C, with or without KJE-ATP. FLuc activity was monitored over time in the absence (blue circles) or presence (orange circles) of the KJE system (1 µM DnaK, 0.33 µM DnaJ and 1.5 µM GrpE). Black squares and red squares indicate the assisted and spontaneous folding at 37°C, respectively. In other reactions KJE (1 µM DnaK, 0.33 µM DnaJ and 1.5 µM GrpE) was added either to native FLuc (purple diamonds) or to FLuc after 1 h of spontaneous folding (green triangles). Ass., KJE-ATP assisted folding; Spont., spontaneous folding; N-FLuc, native FLuc. Data were fitted to a single exponential function. Error bars represent s.d. (n=5). All reactions contain 5 mM ATP. Ass., KJE-ATP assisted

folding. (b) FLuc remains monomeric during spontaneous folding at 37°C. An equimolar mixture of GuHCl-denatured FLuc labelled at the N-terminus with Alexa532 or Alexa647 was diluted to a total concentration of 100 pM in folding buffer at 37°C, and inter-molecular association monitored using dcFCCS after 1 h of incubation. As a positive control mimicking a dimeric protein, 100 pM double-labelled FLucNC was analysed. Representative measurements of 3 independent experiments. (c) The KJE system rescues compact folding intermediates generated during spontaneous folding at 37°C. GuHCl-denatured FLucNC was diluted to 50 pM in folding buffer containing ATP (5 mM) and PBT (50 µM), followed by recording of smFRET data for 30 min. DnaK (0.3 µM) and DnaJ (0.1 µM) were added after 30 min to generate DnaK-bound FLuc (red), followed by GrpE (0.5 µM) to initiate folding (blue) for 1 h. N-FLucNC was shown for comparison. SmFRET data were recorded as in Fig. 4e. Representative measurements of 3 independent repeats are shown. (d) N-FLucNC (10 pM) in folding buffer, containing PBT and ATP, was shifted from 25°C to 37°C in the presence or absence of KJ. SmFRET was recorded for 30 min after incubation for 2 h without chaperones. Upon transfer to 37°C in the presence of KJ-ATP, smFRET was recorded either immediately (0 min) or after 90 min. GrpE was added after 2 h to initiate refolding of FLuc from the DnaK-bound state. One representative measurement of three independent repeats is shown.

5 Discussion

In this study, we investigated how the Hsp70 chaperone system promotes the folding of substrate proteins. We found that Hsp70 actively accelerates the folding of the model substrate, firefly luciferase (FLuc). Accelerated folding is achieved by two complementary mechanisms: unfolding of misfolded states and by directing a fraction of molecules towards a kinetically more efficient folding pathway. The ability to accelerate folding allows the Hsp70 system to shift the folding equilibrium towards the native state in an ATP-dependent manner.

5.1 The DnaK/DnaJ/GrpE system catalyzes protein folding

The canonical model of Hsp70 function proposes that successive cycles of substrate binding and release by Hsp70 inhibit aggregation of the client protein without influencing its folding rate. Here, we showed that Hsp70 accelerates folding of a model multi-domain protein, FLuc, and thus has an active role in promoting protein folding beyond preventing aggregation. FLuc has been extensively used as an Hsp70-dependent multi-domain model protein with high aggregation propensity *in vitro* and *in vivo* (Agashe et al., 2004; Frydman et al., 1994; Schröder et al., 1993a; Szabo et al., 1994). Several studies showed that Hsp70-assisted folding of FLuc occurs with high yield at an apparenthal-time of 6-8 min (Kityk et al., 2015; Moran Luengo et al., 2018; Sharma et al., 2010), whereas significant spontaneous folding was only observed in one study using low FLuc concentrations (Herbst et al., 1998). In this prior work, the spontaneous folding of FLuc was found to be exceedingly slow, reaching equilibrium after ~72 h with a yield of ~80% at low temperature and low protein concentration (~30 nM). These conditions were chosen to minimize aggregation, which is typically, though not always, irreversible. Unlike irreversible aggregation, reversible (or transient) aggregation would be expected to reduce the rate rather than the yield of folding (Georgescauld et al., 2014; Gupta et al., 2014). The GroEL/ES chaperonin was previously thought to act only to inhibit aggregation, and observations of accelerated folding were attributed to prevention of reversible aggregation during folding. However, later work definitively demonstrated that GroEL can accelerate folding, even in the absence of aggregation (Georgescauld et al., 2014; Gupta et al., 2014).

To explore the effect of the Hsp70 system on folding in the absence of aggregation, we first used dcFCCS to establish conditions in which reversible aggregation does not occur during spontaneous folding. We found that above a concentration of 10 nM, FLuc oligomerized during the spontaneous folding reaction. Additionally, we demonstrated with luminescence-based activity assays that the rate of folding below 10 nM was concentration independent, indicating that transient aggregation does not explain slow folding in this concentration regime. Analysis of spontaneous and chaperone-assisted folding under these conditions definitively revealed that the Hsp70 system can both accelerate folding of FLuc and rapidly convert kinetically trapped states to the native state. Both the bacterial Hsp70 system and the chaperonin GroEL/ES can therefore be considered as folding catalysts.

5.2 DnaK unfolds kinetically trapped states

Ever since the pioneering experiments of Christian Anfinsen, demonstrating the spontaneous folding of small single-domain proteins (Anfinsen, 1973b) folding studies by numerous researchers have mainly focused on relatively small proteins around 100 amino acids in length. The folding properties of larger, multi-domain proteins remained under-studied. When we analysed the cause of slow spontaneous folding of FLuc by smFRET, we observed that a kinetically trapped state arises from non-native interactions between the N_S and N_L subdomains (Figure 5.1a). Interestingly, we found that this misfolded state is more compact than the native state. Consistent with our results, prior work using atomic force microscopy (AFM) had provided evidence for inefficient spontaneous folding of FLuc (Scholl et al., 2014). While the spontaneous folding from the fully extended state resulted in FLuc being trapped in a kinetically stable, non-native configuration, partially unfolded FLuc with a folded N-terminal domain could fold spontaneously, suggesting that misfolding either occurred within the large N-domain or due to inappropriate interactions between the unfolded N-terminal and unfolded C-terminal domain (Scholl et al., 2014). We hypothesized that acceleration of folding by Hsp70 would require the misfolded state between the N_S and N_L subdomains to be resolved. Interestingly, folding of FLuc upon synthesis is fast and efficient, as N_S folds co-translationally and folding is rapidly completed upon release of the full-length protein from the ribosome (Frydman et al., 1999). It seemed plausible to assume that DnaK might mimic the effect of co-translational folding, by physically separating the FLuc domains. This possibility was tested using smFRET.

Earlier research suggested that DnaK-binding to a client protein results in the formation of expanded, DnaK-bound conformation (Kellner et al., 2014; Rosenzweig et al., 2017; Sekhar et al., 2017), but the significance of this phenomenon remained unclear. Kellner et al. probed the conformation of rhodanese with smFRET, and demonstrated that DnaK leads to expansion of the bound protein (Kellner et al., 2014). However, since rhodanese is not a DnaK substrate, addition of GrpE to the expanded DnaK-rhodanese complex did not lead to the formation of the native state, thus the significance of the conformational expansion for folding remained obscure. Additionally, an NMR study with human telomere repeat binding factor (hTRF1), a small 53-residue polypeptide, demonstrated that DnaK binding results in expansion, removing long-range interactions (Rosenzweig et al., 2017). However, similarly to rhodanese, hTRF1 does not fold using DnaK.

In our study, probing of each domain of FLuc in solution by smFRET indicated that DnaK binding catalyzes escape from a kinetic trap by unfolding the substrate protein to an expanded state (Figure 5.1a). These misfolded states are aggregation-prone but could be studied at single-molecule concentration by smFRET in a non-aggregated state. Due to the presence relatively stable non-native interactions, FLuc is trapped in a local energy minimum and must cross substantial kinetic energy barriers to reach the native state. Multiple DnaKs (up to ~ 12 per FLuc) bind to the misfolded protein and convert it from a compact to an expanded state. Notably, multiple DnaK binding sites are predicted to be present on FLuc at the N_S:N_L interface where misfolding initiates. Additionally, hydrogen/deuterium exchange coupled to mass spectrometry (H/DX-MS) experiments used in our study strongly supported our interpretation that the misfolded N_S:N_L interface is targeted by DnaK during chaperone-assisted folding (Imamoglu et al., 2021). Note that DnaK cannot bind to the native state due to the inaccessibility of hydrophobic regions, but can bind to the misfolded state, indicating that this state exposes hydrophobic regions in the N-domain. Taken together, these results demonstrate that DnaK-mediated substrate expansion is functionally significant in catalyzing escape from compact kinetically-trapped states.

5.3 FLuc avoids misfolding upon release from DnaK

It has been proposed that proteins fold spontaneously to their native state upon release from DnaK (Sharma et al., 2010). In this model, both release from DnaK and dilution from the denaturant should result in the same efficiency of FLuc folding as long as aggregation is prevented. To test this hypothesis, we devised a novel experimental approach, using a substrate-mimicking peptide

that completely blocks substrate binding to DnaK. Interestingly, ~5-fold more native FLuc is produced in a single round of chaperone action than during the equivalent time period of spontaneous folding. Therefore, the DnaK system not only catalyses escape from kinetic traps, but also promotes a fast folding trajectory, thereby avoiding misfolding for a fraction of the released substrate molecules. We conclude that the DnaK system accelerates folding by two complementary activities (Figure 5.1a): (i) DnaK binds and unfolds misfolded FLuc, thereby rescuing it from stable local minima, and (ii) allows refolding to the native state along a fast folding trajectory that avoids misfolding (Imamoglu et al., 2020). Upon GrpE-mediated release from DnaK, an expanded FLuc is driven towards an intermediate conformation (I_C), which is committed to a fast folding trajectory. Thus, DnaK effectively lowers the kinetic energy barrier toward the native state. A single round of chaperone action has a limited efficiency, with ~15% of molecules being enabled to fold rapidly. The remainder (~85%) of FLuc populates misfolded states after GrpE-mediated release, followed by another cycle of unfolding and release. We suggest that sequential release of multiple DnaK molecules from FLuc may mimic co-translational folding by physically separates the N_S-domain from N_L-domain, thereby preventing interdomain misfolding, albeit with an efficiency limited to ~15%.

We calculate that ~4 cycles of DnaK binding and release are required for half-maximal folding. At ~12 DnaK binding per FLuc, each cycle consumes ~12 molecules of ATP. Thus, on average ~50 ATP would be required for folding per molecule of FLuc. However, whether in each cycle FLuc is returned to the fully extended state remains to be determined.

5.4 DnaK uses ATP to drive substrate folding out of equilibrium

Several chaperones are ATPases and use ATP binding and hydrolysis to regulate their affinity for substrate protein (Balchin et al., 2016). Interestingly, a recent study provided evidence that DnaK uses the energy of ATP to drive substrates toward the native state (Goloubinoff et al., 2018). To test this concept, we analysed the role of DnaK under conditions of elevated temperature, which destabilizes the native protein. At elevated temperature, FLuc preferentially populates misfolded states, indicating that the native state is energetically disfavored. The presence of DnaK under these conditions promoted folding to the native state. Our data suggest that kinetic folding assistance by the Hsp70 system underlies this remarkable effect, as the forward reaction of folding is accelerated relative to the unfolding rate. In support of the physiological significance of this capacity, a recent study showed that overexpression of DnaK/DnaJ/GrpE stabilized

numerous proteins against thermal denaturation *in vivo* (Zhao et al., 2019b). We suggest that accelerated folding by the Hsp70 system has evolved to stabilize proteins in their functionally active states under stress conditions (Figure 5.1b).

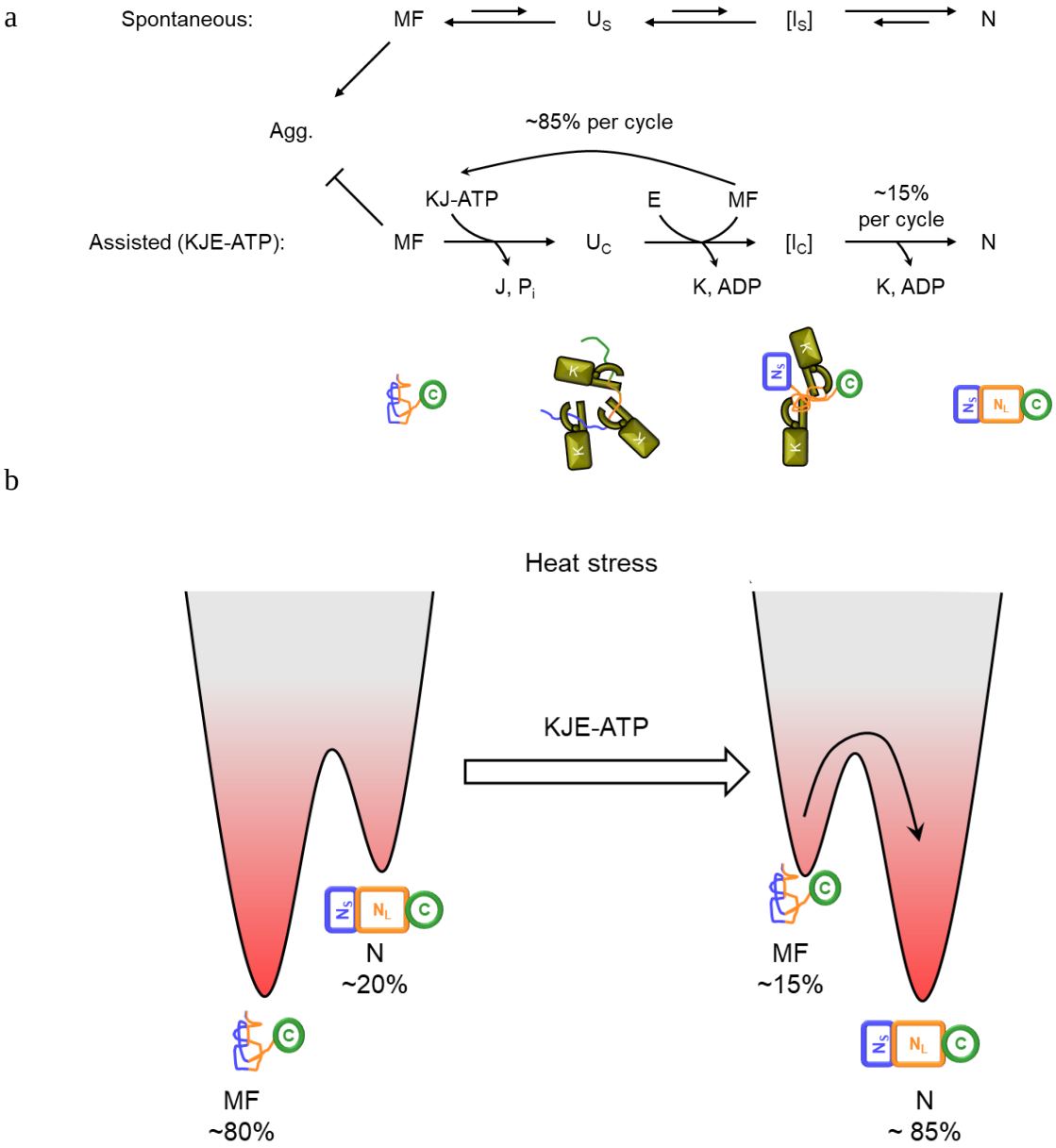


Figure 5.1 Model of spontanoue folding and KJE-ATP assisted folding.
(a) Agg., aggregated protein; MF, misfolded compact intermediate; U_S, unfolded state in the absence of chaperones; I_S, folding intermediate of spontaneous folding; U_C, unfolded state in complex with DnaK; I_C, intermediate commited to fast folding; N, native state. Each KJE cycle results in 15% native FLuc via intermediated commited to fast folding. (b) Simplified energy diagram represent the role of KJE system under heat stress. KJE-ATP stabilizes the native state of FLuc by accelerated folding of MF. Modified from (Imamoglu et al., 2020)

6 Conclusions

We have shown that the bacterial Hsp70 chaperone system, consisting of DnaK, DnaJ and GrpE (KJE), actively promotes the folding of the multi-domain model protein FLuc in an ATP-dependent manner. KJE-ATP accelerates folding by two complementary mechanisms: unfolding of misfolded states and direction of unfolded protein towards a fast folding trajectory.

Our results also define the category of folding problem that is resolved by the DnaK system. KJ-mediated conformational expansion rescues compact misfolded states with non-native inter-domain contacts. This type of folding problem is common in large multi-domain proteins (Borgia et al., 2011; Chiti and Dobson, 2017; Scholl et al., 2017), and might occur during *de novo* folding, or recovery from heat stress. Indeed, FLuc populates a similar collapsed conformational ensemble whether diluted from denaturant or exposed to elevated temperature. Notably, co-translational folding has been observed to facilitate rapid biogenesis of FLuc in eukaryotes by enforcing sequential folding of the N_S and N_L-domains (Frydman et al., 1999). Our data show that the DnaK system provides an analogous post-translational solution to inter-domain misfolding. This solution would be operative when the translation rate does not synchronize with domain folding at the ribosome (Netzer and Hartl, 1997; Pechmann and Frydman, 2013), or during recovery from proteotoxic stress.

Additionally, we found that the Hsp70 system leverages its capacity to accelerate folding to drive FLuc toward the native state, even under denaturing conditions. These results have broad implications for the role of Hsp70 chaperones in protein biogenesis, and conformational maintenance under stress conditions.

7 Reference

Agashe, V.R., Guha, S., Chang, H.C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.U., and Barral, J.M. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. Cell *117*, 199-209.

Anfinsen, C.B. (1973a). Principles that govern the folding of protein chains. Science *181*, 223-230.

Anfinsen, C.B. (1973b). Principles that Govern the Folding of Protein Chains. Science *181*, 223-230.

Anfinsen, C.B., and Haber, E. (1961). Studies on the reduction and re-formation of protein disulfide bonds. J Biol Chem *236*, 1361-1363.

Anfinsen, C.B., Haber, E., Sela, M., and White, F.H., Jr. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A *47*, 1309-1314.

Anson, M.L., and Mirsky, A.E. (1931). The Reversibility of Protein Coagulation. The Journal of Physical Chemistry *35*, 185-193.

Balchin, D., Hayer-Hartl, M., and Hartl, F.U. (2016). In vivo aspects of protein folding and quality control. Science *353*, aac4354.

Balchin, D., Miličić, G., Strauss, M., Hayer-Hartl, M., and Hartl, F.U. (2018). Pathway of Actin Folding Directed by the Eukaryotic Chaperonin TRiC. Cell *174*, 1507-1521.e1516.

Baldwin, T.O. (1996). Firefly luciferase: the structure is known, but the mystery remains. Structure *4*, 223-228.

Barouch, W., Prasad, K., Greene, L., and Eisenberg, E. (1997). Auxilin-induced interaction of the molecular chaperone Hsc70 with clathrin baskets. Biochemistry *36*, 4303-4308.

Barraclough, R., and Ellis, R.J. (1980). Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. Biochim Biophys Acta *608*, 19-31.

Bauer, D., Meinhold, S., Jakob, R.P., Stigler, J., Merkel, U., Maier, T., Rief, M., and Zoldak, G. (2018). A folding nucleus and minimal ATP binding domain of Hsp70 identified by single-molecule force spectroscopy. Proc Natl Acad Sci U S A *115*, 4666-4671.

Bertelsen, E.B., Chang, L., Gestwicki, J.E., and Zuiderweg, E.R. (2009). Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. Proc Natl Acad Sci U S A *106*, 8471-8476.

Bhattacharya, A., Kurochkin, A.V., Yip, G.N.B., Zhang, Y., Bertelsen, E.B., and Zuiderweg, E.R.P. (2009). Allostery in Hsp70 chaperones is transduced by subdomain rotations. Journal of molecular biology *388*, 475-490.

Borgia, M.B., Borgia, A., Best, R.B., Steward, A., Nettels, D., Wunderlich, B., Schuler, B., and Clarke, J. (2011). Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. Nature *474*, 662-665.

Brinker, A., Pfeifer, G., Kerner, M.J., Naylor, D.J., Hartl, F.U., and Hayer-Hartl, M. (2001). Dual function of protein confinement in chaperonin-assisted protein folding. Cell *107*, 223-233.

Calloni, G., Chen, T., Schermann, S.M., Chang, H.C., Genevaux, P., Agostini, F., Tartaglia, G.G., Hayer-Hartl, M., and Hartl, F.U. (2012). DnaK functions as a central hub in the E. coli chaperone network. Cell Rep *1*, 251-264.

Chakraborty, K., Chatila, M., Sinha, J., Shi, Q., Poschner, B.C., Sikor, M., Jiang, G., Lamb, D.C., Hartl, F.U., and Hayer-Hartl, M. (2010). Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. Cell *142*, 112-122.

Cheng, M.Y., Hartl, F.U., Martin, J., Pollock, R.A., Kalousek, F., Neupert, W., Hallberg, E.M., Hallberg, R.L., and Horwich, A.L. (1989). Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. Nature *337*, 620-625.

Chiti, F., and Dobson, C.M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. Annu Rev Biochem *86*, 27-68.

Conti, E., Franks, N.P., and Brick, P. (1996). Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure *4*, 287-298.

Dahiya, V., and Buchner, J. (2019). Functional principles and regulation of molecular chaperones. Adv Protein Chem Struct Biol *114*, 1-60.

Daugaard, M., Rohde, M., and Jaattela, M. (2007). The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. FEBS Lett *581*, 3702-3710.

De Los Rios, P., Ben-Zvi, A., Slutsky, O., Azem, A., and Goloubinoff, P. (2006). Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. Proc Natl Acad Sci U S A *103*, 6166-6171.

Deniz, A.A., Dahan, M., Grunwell, J.R., Ha, T., Faulhaber, A.E., Chemla, D.S., Weiss, S., and Schultz, P.G. (1999). Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. Proceedings of the National Academy of Sciences *96*, 3670-3675.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature *400*, 693-696.

Diamant, S., Ben-Zvi, A.P., Bukau, B., and Goloubinoff, P. (2000). Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. J Biol Chem *275*, 21107-21113.

Dill, K.A., and Chan, H.S. (1997). From Levinthal to pathways to funnels. Nat Struct Biol *4*, 10-19.

Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychkova, V.E., Semisotnov, G.V., Venyaminov, S., and Ptitsyn, O.B. (1981). Alpha-Lactalbumin: compact state with fluctuating tertiary structure? FEBS Lett *136*, 311-315.

Eisenberg, E., and Greene, L.E. (2007). Multiple roles of auxilin and hsc70 in clathrin-mediated endocytosis. Traffic *8*, 640-646.

Ellis, J. (1987). Proteins as molecular chaperones. Nature *328*, 378-379.

Ellis, R.J. (1994). Molecular chaperones. Opening and closing the Anfinsen cage. Curr Biol *4*, 633-635.

Ellis, R.J. (2001). Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci *26*, 597-604.

Ellis, R.J., and Hartl, F.U. (1996). Protein folding in the cell: competing models of chaperonin function. Faseb j *10*, 20-26.

Elson, E.L. (2011). Fluorescence correlation spectroscopy: past, present, future. Biophys J *101*, 2855-2870.

Enderlein, J., Gregor, I., Patra, D., Dertinger, T., and Kaupp, U.B. (2005). Performance of fluorescence correlation spectroscopy for measuring diffusion and concentration. Chemphyschem *6*, 2324-2336.

Fan, C.Y., Lee, S., Ren, H.Y., and Cyr, D.M. (2004). Exchangeable chaperone modules contribute to specification of type I and type II Hsp40 cellular function. Mol Biol Cell *15*, 761-773.

Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. (1989). The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. J Bacteriol *171*, 1379-1385.

Flaherty, K.M., DeLuca-Flaherty, C., and McKay, D.B. (1990). Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature *346*, 623-628.

Frydman, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F.U. (1999). Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. Nat Struct Biol *6*, 697-705.

Frydman, J., Nimmesgern, E., Ohtsuka, K., and Hartl, F.U. (1994). Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. Nature *370*, 111-117.

Georgescauld, F., Popova, K., Gupta, A.J., Bracher, A., Engen, J.R., Hayer-Hartl, M., and Hartl, F.U. (2014). GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. Cell *157*, 922-934.

Georgopoulos, C. (2006). Toothpicks, serendipity and the emergence of the Escherichia coli DnaK (Hsp70) and GroEL (Hsp60) chaperone machines. Genetics *174*, 1699-1707.

Gillis, J., Schipper-Krom, S., Juenemann, K., Gruber, A., Coolen, S., van den Nieuwendijk, R., van Veen, H., Overkleeft, H., Goedhart, J., Kampinga, H.H., *et al.* (2013). The DNAJB6 and DNAJB8 protein chaperones prevent intracellular aggregation of polyglutamine peptides. J Biol Chem *288*, 17225-17237.

Goloubinoff, P. (2017). Editorial: The HSP70 Molecular Chaperone Machines. Front Mol Biosci *4*, 1.

Goloubinoff, P., Christeller, J.T., Gatenby, A.A., and Lorimer, G.H. (1989a). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfoleded state depends on two chaperonin proteins and Mg-ATP. Nature *342*, 884-889.

Goloubinoff, P., and De Los Rios, P. (2007). The mechanism of Hsp70 chaperones: (entropic) pulling the models together. Trends Biochem Sci *32*, 372-380.

Goloubinoff, P., Gatenby, A.A., and Lorimer, G.H. (1989b). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli. Nature *337*, 44-47.

Goloubinoff, P., Sassi, A.S., Fauvet, B., Barducci, A., and De Los Rios, P. (2018). Chaperones convert the energy from ATP into the nonequilibrium stabilization of native proteins. Nat Chem Biol *14*, 388-395.

Gupta, A.J., Haldar, S., Milicic, G., Hartl, F.U., and Hayer-Hartl, M. (2014). Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. J Mol Biol *426*, 2739-2754.

Han, W., and Christen, P. (2003). Mechanism of the targeting action of DnaJ in the DnaK molecular chaperone system. J Biol Chem *278*, 19038-19043.

Harrison, C.J., Hayer-Hartl, M., Di Liberto, M., Hartl, F., and Kuriyan, J. (1997). Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science *276*, 431-435.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature *475*, 324-332.

Hartl, F.U., and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol *16*, 574-581.

Hayer-Hartl, M., Bracher, A., and Hartl, F.U. (2016). The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding. Trends Biochem Sci *41*, 62-76.

Herbst, R., Gast, K., and Seckler, R. (1998). Folding of firefly (Photinus pyralis) luciferase: Aggregation and reactivation of unfolding intermediates. Biochemistry *37*, 6586-6597.

Herbst, R., Schafer, U., and Seckler, R. (1997). Equilibrium intermediates in the reversible unfolding of firefly (Photinus pyralis) luciferase. J Biol Chem *272*, 7099-7105.

Hipp, M.S., Kasturi, P., and Hartl, F.U. (2019). The proteostasis network and its decline in ageing. Nat Rev Mol Cell Biol *20*, 421-435.

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. Biochim Biophys Acta *1803*, 650-661.

Horwich, A.L., Apetri, A.C., and Fenton, W.A. (2009). The GroEL/GroES cis cavity as a passive anti-aggregation device. FEBS Lett *583*, 2654-2662.

Horwich, A.L., Fenton, W.A., Chapman, E., and Farr, G.W. (2007). Two families of chaperonin: physiology and mechanism. Annu Rev Cell Dev Biol *23*, 115-145.

Imamoglu, R., Balchin, D., Hayer-Hartl, M., and Hartl, F.U. (2020). Bacterial Hsp70 resolves misfolded states and accelerates productive folding of a multi-domain protein. Nat Commun *11*, 365.

Itzhaki, L.S., Otzen, D.E., and Fersht, A.R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol *254*, 260-288.

Jackson, S.E., and Fersht, A.R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry *30*, 10428-10435.

Jiang, J., Maes, E.G., Taylor, A.B., Wang, L., Hinck, A.P., Lafer, E.M., and Sousa, R. (2007). Structural basis of J cochaperone binding and regulation of Hsp70. Mol Cell *28*, 422-433.

Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco, I., Jr., and Bustamante, C. (2011). The ribosome modulates nascent protein folding. Science *334*, 1723-1727.

Kampinga, H.H., and Craig, E.A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol *11*, 579-592.

Karplus, M., and Weaver, D.L. (1976). Protein-folding dynamics. Nature *260*, 404-406.

Karzai, A.W., and McMacken, R. (1996). A bipartite signaling mechanism involved in DnaJ-mediated activation of the Escherichia coli DnaK protein. J Biol Chem *271*, 11236-11246.

Kellner, R., Hofmann, H., Barducci, A., Wunderlich, B., Nettels, D., and Schuler, B. (2014). Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein. Proc Natl Acad Sci U S A *111*, 13355-13360.

Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013). Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem *82*, 323-355.

Kirschke, E., Goswami, D., Southworth, D., Griffin, P.R., and Agard, D.A. (2014). Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. Cell *157*, 1685-1697.

Kitadai, N., and Maruyama, S. (2018). Origins of building blocks of life: A review. Geoscience Frontiers *9*, 1117-1153.

Kityk, R., Kopp, J., and Mayer, M.P. (2018). Molecular Mechanism of J-Domain-Triggered ATP Hydrolysis by Hsp70 Chaperones. Mol Cell *69*, 227-237 e224.

Kityk, R., Vogel, M., Schlecht, R., Bukau, B., and Mayer, M.P. (2015). Pathways of allosteric regulation in Hsp70 chaperones. Nat Commun *6*, 8308.

Klaips, C.L., Jayaraj, G.G., and Hartl, F.U. (2018). Pathways of cellular proteostasis in aging and disease. J Cell Biol *217*, 51-63.

Koplin, A., Preissler, S., Ilina, Y., Koch, M., Scior, A., Erhardt, M., and Deuerling, E. (2010). A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. J Cell Biol *189*, 57-68.

Kramer, G., Shiber, A., and Bukau, B. (2019). Mechanisms of Cotranslational Maturation of Newly Synthesized Proteins. Annu Rev Biochem *88*, 337-364.

Kulak, N.A., Geyer, P.E., and Mann, M. (2017). Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. Molecular & Cellular Proteomics *16*, 694-705.

Kumar, D.P., Vorvis, C., Sarbeng, E.B., Cabra Ledesma, V.C., Willis, J.E., and Liu, Q. (2011). The four hydrophobic residues on the Hsp70 inter-domain linker have two distinct roles. J Mol Biol *411*, 1099-1113.

Labbadia, J., and Morimoto, R.I. (2015). The biology of proteostasis in aging and disease. Annu Rev Biochem *84*, 435-464.

Laskey, R.A., Honda, B.M., Mills, A.D., and Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature *275*, 416-420.

Laufen, T., Mayer, M.P., Beisel, C., Klostermeier, D., Mogk, A., Reinstein, J., and Bukau, B. (1999). Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. Proceedings of the National Academy of Sciences *96*, 5452-5457.

Lee, N.K., Kapanidis, A.N., Wang, Y., Michalet, X., Mukhopadhyay, J., Ebright, R.H., and Weiss, S. (2005). Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. Biophys J *88*, 2939-2953.

Lerner, E., Cordes, T., Ingargiola, A., Alhadid, Y., Chung, S., Michalet, X., and Weiss, S. (2018). Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. Science *359*.

Levinthal, C. (1968). Are there pathways for protein folding? J Chim Phys *65*, 44-45.

Lin, Z., Madan, D., and Rye, H.S. (2008). GroEL stimulates protein folding through forced unfolding. Nature structural & molecular biology *15*, 303-311.

Lindquist, S. (1986). The heat-shock response. Annu Rev Biochem *55*, 1151-1191.

Markesich, D.C., Gajewski, K.M., Nazimiec, M.E., and Beckingham, K. (2000). bicaudal encodes the Drosophila beta NAC homolog, a component of the ribosomal translational machinery. Development *127*, 559-572.

Mashaghi, A., Kramer, G., Bechtluft, P., Zachmann-Brand, B., Driessen, A.J., Bukau, B., and Tans, S.J. (2013). Reshaping of the conformational search of a protein by the chaperone trigger factor. Nature *500*, 98-101.

Mashaghi, A., Kramer, G., Lamb, D.C., Mayer, M.P., and Tans, S.J. (2014a). Chaperone Action at the Single-Molecule Level. Chem Rev *114*, 660-676.

Mashaghi, A., Mashaghi, S., and Tans, S.J. (2014b). Misfolding of luciferase at the single-molecule level. Angew Chem Int Ed Engl *53*, 10390-10393.

Mattoo, R.U., Sharma, S.K., Priya, S., Finka, A., and Goloubinoff, P. (2013). Hsp110 is a bona fide chaperone using ATP to unfold stable misfolded polypeptides and reciprocally collaborate with Hsp70 to solubilize protein aggregates. J Biol Chem *288*, 21399-21411.

Mayer, M.P. (2013). Hsp70 chaperone dynamics and molecular mechanism. Trends Biochem Sci *38*, 507-514.

Mayer, M.P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci *62*, 670-684.

Mayer, M.P., and Gierasch, L.M. (2019). Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. J Biol Chem *294*, 2085-2097.

Mayer, M.P., Rudiger, S., and Bukau, B. (2000a). Molecular basis for interactions of the DnaK chaperone with substrates. Biol Chem *381*, 877-885.

Mayer, M.P., Schroder, H., Rudiger, S., Paal, K., Laufen, T., and Bukau, B. (2000b). Multistep mechanism of substrate binding determines chaperone activity of Hsp70. Nat Struct Biol *7*, 586-593.

Melero, R., Moro, F., Perez-Calvo, M.A., Perales-Calvo, J., Quintana-Gallardo, L., Llorca, O., Muga, A., and Valpuesta, J.M. (2015). Modulation of the chaperone DnaK allosterism by the nucleotide exchange factor GrpE. J Biol Chem *290*, 10083-10092.

Misselwitz, B., Staeck, O., and Rapoport, T.A. (1998). J proteins catalytically activate Hsp70 molecules to trap a wide range of peptide sequences. Mol Cell *2*, 593-603.

Mogk, A., Kummer, E., and Bukau, B. (2015). Cooperation of Hsp70 and Hsp100 chaperone machines in protein disaggregation. Frontiers in Molecular Biosciences *2*.

Moran Luengo, T., Kityk, R., Mayer, M.P., and Rudiger, S.G.D. (2018). Hsp90 Breaks the Deadlock of the Hsp70 Chaperone System. Mol Cell *70*, 545-552 e549.

Morimoto, R.I. (1993). Cells in stress: transcriptional activation of heat shock genes. Science *259*, 1409-1410.

Müller, B.K., Zaychikov, E., Bräuchle, C., and Lamb, D.C. (2005). Pulsed interleaved excitation. Biophys J *89*, 3508-3522.

Nakamura, A., Takumi, K., and Miki, K. (2010). Crystal structure of a thermophilic GrpE protein: insight into thermosensing function for the DnaK chaperone system. J Mol Biol *396*, 1000-1011.

Netzer, W.J., and Hartl, F.U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature *388*, 343-349.

Netzer, W.J., and Hartl, F.U. (1998). Protein folding in the cytosol: chaperonin-dependent and -independent mechanisms. Trends Biochem Sci *23*, 68-73.

Nickson, A.A., and Clarke, J. (2010). What lessons can be learned from studying the folding of homologous proteins? Methods *52*, 38-50.

Nillegoda, N.B., and Bukau, B. (2015). Metazoan Hsp70-based protein disaggregases: emergence and mechanisms. Frontiers in Molecular Biosciences *2*.

O'Brien, E.P., Christodoulou, J., Vendruscolo, M., and Dobson, C.M. (2012). Trigger Factor Slows Co-translational Folding through Kinetic Trapping while Sterically Protecting the Nascent Chain from Aberrant Cytosolic Interactions. Journal of the American Chemical Society *134*, 10920-10932.

Ostermann, J., Horwich, A.L., Neupert, W., and Hartl, F.U. (1989). Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. Nature *341*, 125-130.

Packschies, L., Theyssen, H., Buchberger, A., Bukau, B., Goody, R.S., and Reinstein, J. (1997). GrpE accelerates nucleotide exchange of the molecular chaperone DnaK with an associative displacement mechanism. Biochemistry *36*, 3417-3422.

Palleros, D.R., Shi, L., Reid, K.L., and Fink, A.L. (1994). hsp70-protein complexes. Complex stability and conformation of bound substrate protein. J Biol Chem *269*, 13107-13114.

Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol *20*, 237-243.

Pelham, H.R. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell *46*, 959-961.

Pierpaoli, E.V., Gisler, S.M., and Christen, P. (1998). Sequence-specific rates of interaction of target peptides with the molecular chaperones DnaK and DnaJ. Biochemistry *37*, 16741-16748.

Preissler, S., and Deuerling, E. (2012). Ribosome-associated chaperones as key players in proteostasis. Trends Biochem Sci *37*, 274-283.

Rampelt, H., Kirstein-Miles, J., Nillegoda, N.B., Chi, K., Scholz, S.R., Morimoto, R.I., and Bukau, B. (2012). Metazoan Hsp70 machines use Hsp110 to power protein disaggregation. Embo j *31*, 4221-4235.

Rapoport, I., Boll, W., Yu, A., Böcking, T., and Kirchhausen, T. (2008). A motif in the clathrin heavy chain required for the Hsc70/auxilin uncoating reaction. Mol Biol Cell *19*, 405-413.

Ravindran, M.S., Bagchi, P., Inoue, T., and Tsai, B. (2015). A Non-enveloped Virus Hijacks Host Disaggregation Machinery to Translocate across the Endoplasmic Reticulum Membrane. PLoS Pathog *11*, e1005086.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in drosophila. Experientia *18*, 571-573.

Rodriguez, F., Arsene-Ploetze, F., Rist, W., Rudiger, S., Schneider-Mergener, J., Mayer, M.P., and Bukau, B. (2008). Molecular basis for regulation of the heat shock transcription factor sigma32 by the DnaK and DnaJ chaperones. Mol Cell *32*, 347-358.

Rosam, M., Krader, D., Nickels, C., Hochmair, J., Back, K.C., Agam, G., Barth, A., Zeymer, C., Hendrix, J., Schneider, M., *et al.* (2018). Bap (Sil1) regulates the molecular chaperone BiP by coupling release of nucleotide and substrate. Nat Struct Mol Biol *25*, 90-100.

Rosenzweig, R., Nillegoda, N.B., Mayer, M.P., and Bukau, B. (2019). The Hsp70 chaperone network. Nat Rev Mol Cell Biol *20*, 665-680.

Rosenzweig, R., Sekhar, A., Nagesh, J., and Kay, L.E. (2017). Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles. eLife *6*, e28030.

Rudiger, S., Germeroth, L., Schneider-Mergener, J., and Bukau, B. (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. Embo j *16*, 1501-1507.

Rudiger, S., Schneider-Mergener, J., and Bukau, B. (2001). Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. Embo j *20*, 1042-1050.

Sarbeng, E.B., Liu, Q., Tian, X., Yang, J., Li, H., Wong, J.L., Zhou, L., and Liu, Q. (2015). A functional DnaK dimer is essential for the efficient interaction with Hsp40 heat shock protein. J Biol Chem *290*, 8849-8862.

Schmid, D., Baici, A., Gehring, H., and Christen, P. (1994). Kinetics of molecular chaperone action. Science *263*, 971-973.

Scholl, Z.N., Yang, W., and Marszalek, P.E. (2014). Chaperones rescue luciferase folding by separating its domains. J Biol Chem *289*, 28607-28618.

Scholl, Z.N., Yang, W., and Marszalek, P.E. (2017). Competing Pathways and Multiple Folding Nuclei in a Large Multidomain Protein, Luciferase. Biophys J *112*, 1829-1840.

Schröder, H., Langer, T., Hartl, F.U., and Bukau, B. (1993a). DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. The EMBO journal *12*, 4137-4144.

Schröder, H., Langer, T., Hartl, F.U., and Bukau, B. (1993b). DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. Embo j *12*, 4137-4144.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J *78*, 1606-1619.

Schuler, B. (2007). Application of single molecule Förster resonance energy transfer to protein folding. Methods Mol Biol *350*, 115-138.

Sekhar, A., Nagesh, J., Rosenzweig, R., and Kay, L.E. (2017). Conformational heterogeneity in the Hsp70 chaperone-substrate ensemble identified from analysis of NMR-detected titration data. Protein Science *26*, 2207-2220.

Sekhar, A., Rosenzweig, R., Bouvignies, G., and Kay, L.E. (2015). Mapping the conformation of a client protein through the Hsp70 functional cycle. Proceedings of the National Academy of Sciences *112*, 10395-10400.

Sekhar, A., Rosenzweig, R., Bouvignies, G., and Kay, L.E. (2016). Hsp70 biases the folding pathways of client proteins. Proceedings of the National Academy of Sciences *113*, E2794-E2801.

Sela, M., White, F.H., Jr., and Anfinsen, C.B. (1957). Reductive cleavage of disulfide bridges in ribonuclease. Science *125*, 691-692.

Sharma, S., Chakraborty, K., Müller, B.K., Astola, N., Tang, Y.C., Lamb, D.C., Hayer-Hartl, M., and Hartl, F.U. (2008). Monitoring protein conformation along the pathway of chaperonin-assisted folding. Cell *133*, 142-153.

Sharma, S.K., De los Rios, P., Christen, P., Lustig, A., and Goloubinoff, P. (2010). The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. Nat Chem Biol *6*, 914-920.

Silberg, J.J., Tapley, T.L., Hoff, K.G., and Vickery, L.E. (2004). Regulation of the HscA ATPase reaction cycle by the co-chaperone HscB and the iron-sulfur cluster assembly protein IscU. J Biol Chem *279*, 53924-53931.

Singh, A.K., Balchin, D., Imamoglu, R., Hayer-Hartl, M., and Hartl, F.U. (2020). Efficient Catalysis of Protein Folding by GroEL/ES of the Obligate Chaperonin Substrate MetF. Journal of Molecular Biology.

Smock, R.G., Blackburn, M.E., and Gierasch, L.M. (2011). Conserved, disordered C terminus of DnaK enhances cellular survival upon stress and DnaK in vitro chaperone activity. J Biol Chem *286*, 31821-31829.

Sousa, R., and Lafer, E.M. (2015). The role of molecular chaperones in clathrin mediated vesicular trafficking. Frontiers in Molecular Biosciences *2*.

Sousa, R., Liao, H.S., Cuéllar, J., Jin, S., Valpuesta, J.M., Jin, A.J., and Lafer, E.M. (2016). Clathrin-coat disassembly illuminates the mechanisms of Hsp70 force generation. Nat Struct Mol Biol *23*, 821-829.

Suh, W.C., Burkholder, W.F., Lu, C.Z., Zhao, X., Gottesman, M.E., and Gross, C.A. (1998). Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. Proc Natl Acad Sci U S A *95*, 15223-15228.

Swain, J.F., Dinler, G., Sivendran, R., Montgomery, D.L., Stotz, M., and Gierasch, L.M. (2007). Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. Mol Cell *26*, 27-39.

Szabo, A., Langer, T., Schroder, H., Flanagan, J., Bukau, B., and Hartl, F.U. (1994). The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE. Proc Natl Acad Sci U S A *91*, 10345-10349.

Tang, Y.C., Chang, H.C., Chakraborty, K., Hartl, F.U., and Hayer-Hartl, M. (2008). Essential role of the chaperonin folding compartment in vivo. Embo j *27*, 1458-1468.

Tang, Y.C., Chang, H.C., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M.J., Hartl, F.U., and Hayer-Hartl, M. (2006). Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. Cell *125*, 903-914.

Terada, K., and Oike, Y. (2010). Multiple molecules of Hsc70 and a dimer of DjA1 independently bind to an unfolded protein. J Biol Chem *285*, 16789-16797.

Teter, S.A., Houry, W.A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C., and Hartl, F.U. (1999). Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell *97*, 755-765.

Thirumalai, D., and Lorimer, G.H. (2001). Chaperonin-mediated protein folding. Annu Rev Biophys Biomol Struct *30*, 245-269.

Thompson, J.F., Hayes, L.S., and Lloyd, D.B. (1991). Modulation of firefly luciferase stability and impact on studies of gene regulation. Gene *103*, 171-177.

Tissières, A., Mitchell, H.K., and Tracy, U.M. (1974). Protein synthesis in salivary glands of Drosophila melanogaster: relation to chromosome puffs. J Mol Biol *84*, 389-398.

Ungewickell, E., Ungewickell, H., Holstein, S.E., Lindner, R., Prasad, K., Barouch, W., Martin, B., Greene, L.E., and Eisenberg, E. (1995). Role of auxilin in uncoating clathrin-coated vesicles. Nature *378*, 632-635.

Van Durme, J., Maurer-Stroh, S., Gallardo, R., Wilkinson, H., Rousseau, F., and Schymkowitz, J. (2009a). Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. PLoS Comput Biol *5*, e1000475.

Van Durme, J., Maurer-Stroh, S., Gallardo, R., Wilkinson, H., Rousseau, F., and Schymkowitz, J. (2009b). Accurate Prediction of DnaK-Peptide Binding via Homology Modelling and Experimental Data. PLOS Computational Biology *5*, e1000475.

van Heijenoort, J. (2001). Formation of the glycan chains in the synthesis of bacterial peptidoglycan. Glycobiology *11*, 25r-36r.

Weaver, J., Jiang, M., Roth, A., Puchalla, J., Zhang, J., and Rye, H.S. (2017). GroEL actively stimulates folding of the endogenous substrate protein PepQ. Nat Commun *8*, 15934.

Wetlaufer, D.B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. Proc Natl Acad Sci U S A *70*, 697-701.

Weyer, F.A., Gumiero, A., Gese, G.V., Lapouge, K., and Sinning, I. (2017). Structural insights into a unique Hsp70-Hsp40 interaction in the eukaryotic ribosome-associated complex. Nat Struct Mol Biol *24*, 144-151.

White, F.H., Jr. (1961). Regeneration of native secondary and tertiary structures by air oxidation of reduced ribonuclease. J Biol Chem *236*, 1353-1360.

Wu, C.C., Naveen, V., Chien, C.H., Chang, Y.W., and Hsiao, C.D. (2012). Crystal structure of DnaK protein complexed with nucleotide exchange factor GrpE in DnaK chaperone system: insight into intermolecular communication. J Biol Chem *287*, 21461-21470.

Yamagishi, N., Yokota, M., Yasuda, K., Saito, Y., Nagata, K., and Hatayama, T. (2011). Characterization of stress sensitivity and chaperone activity of Hsp105 in mammalian cells. Biochem Biophys Res Commun *409*, 90-95.

Yang, D., Ye, X., and Lorimer, G.H. (2013). Symmetric GroEL:GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine. Proc Natl Acad Sci U S A *110*, E4298-4305.

Yu, H.Y., Ziegelhoffer, T., Osipiuk, J., Ciesielski, S.J., Baranowski, M., Zhou, M., Joachimiak, A., and Craig, E.A. (2015). Roles of intramolecular and intermolecular interactions in functional regulation of the Hsp70 J-protein co-chaperone Sis1. J Mol Biol *427*, 1632-1643.

Zander, C., Sauer, M., Drexhage, K.H., Ko, D.S., Schulz, A., Wolfrum, J., Brand, L., Eggeling, C., and Seidel, C.A.M. (1996). Detection and characterization of single molecules in aqueous solution. Applied Physics B *63*, 517-523.

Zhao, L., Vecchi, G., Vendruscolo, M., Korner, R., Hayer-Hartl, M., and Hartl, F.U. (2019a). The Hsp70 Chaperone System Stabilizes a Thermo-sensitive Subproteome in E. coli. Cell Rep *28*, 1335-1345.e1336.

Zhao, L., Vecchi, G., Vendruscolo, M., Körner, R., Hayer-Hartl, M., and Hartl, F.U. (2019b). The Hsp70 Chaperone System Stabilizes a Thermo-sensitive Subproteome in E. coli. Cell Reports *28*, 1335-1345.e1336.

Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E., and Hendrickson, W.A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. Science *272*, 1606-1614.

Zimmerman, S.B., and Trach, S.O. (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. J Mol Biol *222*, 599-620.

Zylicz, M., Ang, D., Liberek, K., and Georgopoulos, C. (1989). Initiation of lambda DNA replication with purified host- and bacteriophage-encoded proteins: the role of the dnaK, dnaJ and grpE heat shock proteins. Embo j *8*, 1601-1608.

Zylicz, M., Yamamoto, T., McKittrick, N., Sell, S., and Georgopoulos, C. (1985). Purification and properties of the dnaJ replication protein of Escherichia coli. J Biol Chem *260*, 7591-7598.