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

GroEL/ES modulates the mechanism and accelerates the rate of TIM-barrel domain folding

Kristina Popova

aus

Raevskii, Russische Föderation

2014

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn F.-U. Hartl betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 30.10.2014

.....
(Unterschrift des Autors / der Autorin)

Dissertation eingereicht am 10.11.2014

1. Gutachter: Prof. F.-U. Hartl
2. Gutachter: Prof. R. Beckmann

Mündliche Prüfung am 10.12.2014

ACKNOWLEDGEMENTS

I would first and foremost like to express my deepest gratitude to my supervisor Dr. Manajit Hayer-Hartl who provided me with an opportunity to work on this exciting project and constantly supported my work. I learned a lot during the years of my PhD and grew on both personal and professional level.

I am very thankful to Prof. Ulrich Hartl for his valuable advice and support to this project. It is the great ideas and encouragement from Prof. Hartl and Dr. Hayer-Hartl that made this work possible.

Very special thanks go to Dr. Florian Georgescauld who taught me everything and was a close collaborator to me during all years of my PhD. He was my true mentor and an enormous support all through.

Great thanks to Amit Gupta and other members of the GroEL office whose valuable input and insightful discussions were so helpful to this project.

Thanks to Romy, Nadine, Ana, Peter, Emmanuel, Elizabeth, Evelyn and of course Darija, for it is due to their hard work and support that this project always ran smoothly.

To all hartl-fighters, for numerous evenings of laughter and fun, their helpful suggestions and the true atmosphere of science, which I will always remember.

And finally, my deepest thanks to my friends and family who always encouraged me and helped me at every step of my life.

CONTENTS

1 Summary

The interactome of GroEL/ES has been characterized extensively in several studies and substrates of the chaperonin have been classified (Kerner et al., 2005; Fijiwara et al., 2010). However, the question of what makes some proteins GroEL-dependent and how exactly the chaperonin system promotes their folding remained unresolved. Moreover, it has been unclear how the chaperonin acts on its substrates and whether the protein folding pathway is modified inside the cage as compared to free solution. The aim of this study, therefore, was to characterise and compare the spontaneous and chaperonin-assisted refolding pathway of an obligate substrate of GroEL/ES, in order to elucidate the mechanism of GroEL/ES action.

This study presents evidence that encapsulation in the GroEL/ES-cage accelerates the rate and modulates the mechanism of folding of its obligate TIM-barrel substrate, dihydrodipicolinate synthase. We found that the spontaneous refolding of DAPA is slow due to high cooperativity of the process, as it initiates from an ensemble of unstructured intermediates. We demonstrated that the confining environment of the chaperonin cage promotes formation of the TIM-barrel structure in a segmental manner, lowering the entropic component of the activation barrier and accelerating the rate of DAPA folding. Moreover, the spontaneous refolding pathway of a GroEL-independent homolog of DAPA, MsNANA, closely resembles that of DAPA inside the chaperonin cage. Thus, we conclude that GroEL/ES is a powerful folding catalyst for the substrates that otherwise fail to effectively reach their native state.

2 Introduction

> The word “protein” that I propose to you… I would wish to derive from *proteios* as it appears to be the primitive or principal substance of animal nutrition that plants prepare for the herbovores, and which the latter then furnish to the carnovores.
>
> *J. J. Berzelius,* letter to G. J. Mulder, 1838.

Proteins are linear polymers of amino acids linked together in a specific sequence. They are major macromolecular constituents of all cells, forming key structural elements (cytoskeleton, outer and intra-cellular membrane components etc.) and participating in nearly all cellular activities. For instance, enzymes (catalytic proteins) mediate biochemical reactions, membrane transport proteins (ion channels, specific transporters and pumps) regulate the flux of molecules through the cell membrane, regulatory proteins (kinases, DNA-binding proteins, receptors) control the cell cycle, signal transduction and gene expression, and the superfamily of immunoglobulins are a central part of the immune system.

In order to perform their functions, proteins must adopt a unique three-dimensional structure. The process of acquiring the native structure is called protein folding and is the major focus of the current work.

2.1 Protein folding

2.1.1 Protein structure

Proteins are comprised of 20 different amino acid types, bound to each other in a chain through covalent peptidic bonds. Recently, two additional aminoacids – selenocysteine and pyrrolysine – were discovered to occur naturally in proteins, however, these are modifications of the standard cysteine and lysine residues.

Amino acid sequence is genetically determined and unique for each protein. It is often called the *primary structure* of a protein. Once synthesized, proteins begin to fold, forming local *secondary structure* elements, such as α-helices and β-sheets (Pauling and Corey, 1951a). These secondary elements are then packed together through hydrophobic, polar and ionic interactions, forming compact units called domains (*tertiary structure*). Domains are considered to be independent structural units and different domains within a protein can be performing separate functions (Doolittle et al., 1995). Polypeptide chains can assemble into oligomers and their spacial arrangement is referred to as the *quaternary structure* of a protein.

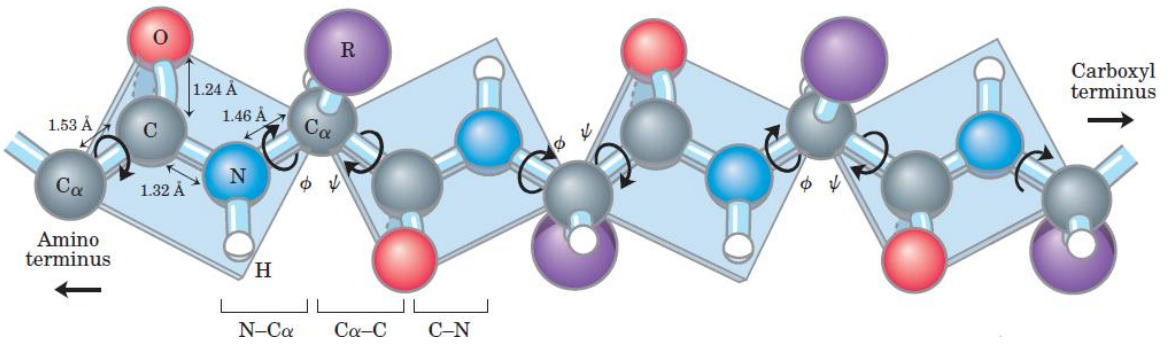


Figure 2.1 The planar peptide group and rotation about bonds in a polypeptide.

Three bonds separate sequential α-carbons in the polypeptide chain. Each peptide bond is planar and cannot rotate. N – Cα and Cα – C bonds can rotate, with bond angles assigned φ and ψ, respectively (Lehninger et al., 2005).

The α carbon atoms of adjacent amino acid residues in the primary structure are separated by three covalent bonds, arranged as Cα – C – N – Cα. The peptide bond has considerable double-bond character and is essentially planar (Figure 2.1). In contrast with the peptide bond, the bonds N – Cα and Cα – C bonds are pure single bonds. The two adjacent rigid peptide units may rotate about these bonds, taking on various orientations. This freedom of rotation about two bonds of each amino acid allows proteins to fold in many different ways. The rotations about these bonds can be specified by dihedral or torsion angles: phi (φ) (the angle of rotation about the bond N – Cα) and psi (ψ) (the angle of rotation about the bond Cα – C). A clockwise rotation about either bond as viewed from the front of the back group corresponds to a positive value. The φ and ψ angles determine the path of the polypeptide chain.

G. N. Ramachandran recognized that many combinations of φ and ψ angles are forbidden because of steric collisions between atoms (Ramachandran and Sasisekharan, 1968). The allowed values can be visualized on a two-dimensional plot called a Ramachandran diagram (Figure 2.2). Considering the energy contained in various pairs of ψ and φ angles, Ramachandran

and his colleagues found two most stable pairs, the so called α and β conformations. These two pairs of angles are found to almost exclusively occur in folded proteins.

In terms of thermodynamics, the favourable entropy associated with the large number of conformations in the unfolded form opposes folding and must be overcome by interactions favouring the folded form. Thus, highly flexible polymers with a large number of possible conformations do not fold into unique structures. The rigidity of the peptide unit and the restricted set of allowed φ and ψ angles limits the number of structures accessible to the unfolded form sufficiently to allow protein folding to occur (Berg et al., 2002).

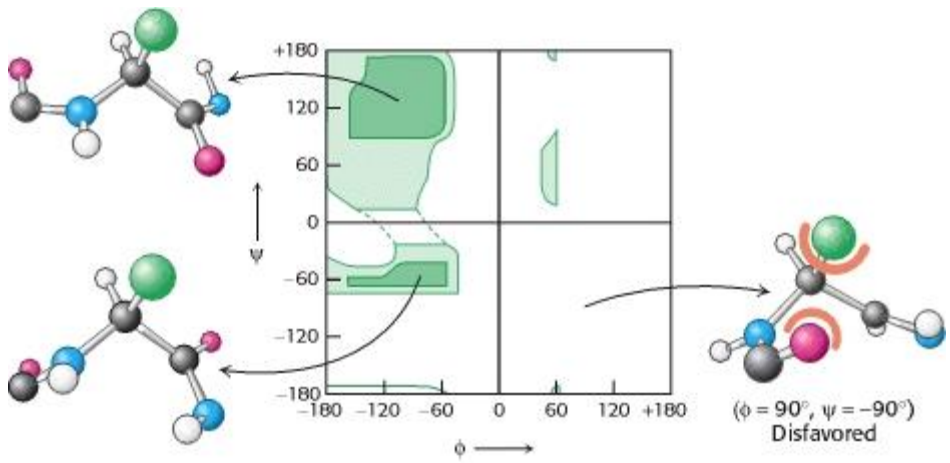


Figure 2.2 A Ramachandran Diagram Showing the Values of φ and ψ for L-alanine.

The most favourable regions are shown in dark green; borderline regions are shown in light green. The structure on the right is disfavoured because of steric clashes. The most favourable regions are shown in dark green; borderline regions are shown in light green. The structure on the right is disfavoured because of steric clashes. (adapted from Berg JM, Tymoczko JL, Stryer L., 2002)

2.1.2 The complexity of protein folding

Proteins have been evolutionarily designed to perform a specific biological function. Essential to this function for many proteins appears to be a well-defined conformational structure (native

conformation) under biological conditions. Thus, part of the protein evolution process must involve pressure to fold to a defined structure.

The question of how the protein chain reaches its native conformation remains a major question in the field since the time of the pioneering experiments performed by Christian Anfinsen. In the late 1950s, he found that purified, denatured Ribonuclease A regains its enzymatic activity upon removal of the denaturant and therefore, demonstrated that all the information required to fold the protein into its native, three-dimensional structure is contained in the amino acid sequence of the polypeptide chain (Haber and Anfinsen, 1962; Anfinsen, 1973).

The famous experiment allowed Anfinsen to postulate that folding of a protein is pathway independent and only depends on protein sequence and the external conditions, such as temperature or composition of the solution. However, this theory was soon opposed by Cyrus Levinthal, who in 1968 made an argument that there would be too many protein conformations to be sampled by random searching before reaching the native state.

Levinthal suggested that a protein containing 100 amino acid residues would result in 10100 different possible conformations, assuming the spontaneous folding process is a random process in which a polypeptide chain tries out all possible conformations with each amino acid residue having on average 10 different conformations,. Knowing that the interconversion between conformations needs ~10^{-13} seconds, the 100-residue polypeptide would take about ~ 1000 years to explore its conformational space, which is beyond the time range of any biological process (Levinthal et al., 1968).

However, it is known that in *E. coli* a functional protein containing 100 amino acid residues is synthesized in about 5 seconds at 37°C. Each *E.coli* cell divides approximately every 20 minutes,

which leads to the conclusion that protein folding cannot be a random, trial and error process. This argument is now known as Levinthal's paradox. Levinthal concluded that proteins must fold to their native conformation by specific folding pathways (Levinthal et al., 1968).

This paradox can be solved if we consider that reaching the global energy minimum (acquiring the native structure) and doing it fast (kinetic control) are not mutually exclusive.

The reversible *in vitro* folding of a single protein means that the protein in the native state is thermodynamically stable, and therefore that the native state has the global minimum free energy of all kinetically accessible structures (Levinthal et al., 1968). Moreover, the folded structure must then have the lowest internal energy of all kinetically accessible conformational structures. Therefore, we should expect the shape of the landscape for a protein to have funnel topography, at least in the vicinity of the native structure (Figure 2.3) (Plotkin and Onuchic, 2002). A consequence of funnelled landscape topography is that the native structure is kinetically accessible at the temperatures where it is thermodynamically stable. A funnelled landscape will also be robust to environmental perturbations as well as sequence mutations, because potentially competing low-energy states are still similar in structure.

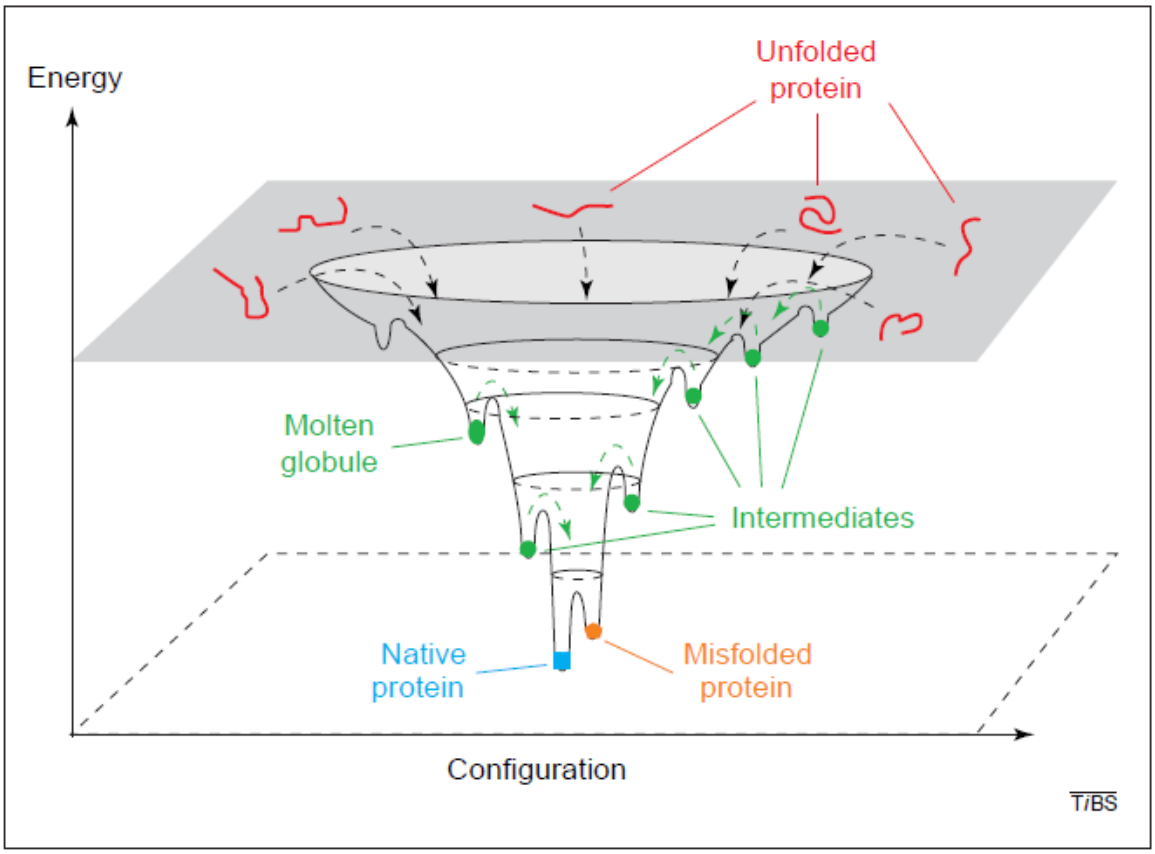


Figure 2.3 Funnel topography of energy landscape

The depth of the funnel represents the free energy of a conformational state, whereas the width is a measure of the configurational entropy. Different states have different energies. (adapted from Radford et al., 2000)

2.1.3 Protein folding mechanism

A funnelled folding mechanism is a well-defined physical solution to the Levinthal problem, however other generic ordering processes may operate in parallel with and possibly accelerate the folding rate. Such transitions have been experimentally observed and their timescale is often shorter than the overall protein timescale. (Ballew et al. 1996; Munoz et al. 1998; Kuwata et al. 2001).

Generic collapse due to net hydrophobicity in the protein increases the packing density, and results in an extensive amount of entropy reduction. Nucleation of this state from the coil is analogous to condensation from the gas phase, although the barriers for each process may be quite different. Many observations of molten globule intermediates may be simply generically collapsed states with minimal tertiary order (Onuchic, 1997). Additionally, appearance of non-native contacts in some of these intermediates was observed, supporting the notion that the tertiary structure of the intermediate is not yet well-defined (Hamada et al. 1996).

Another mechanism of generic entropy reduction is the formation of transient helical structure in the unfolded state. Various studies have observed secondary structure in the unfolded state (Miranker et al. 1991; Radford et al. 1992). Whether native or non-native, helical structure is likely to be transient in the high entropy unfolded state, rather than fixed and rigid. Partial formation of secondary structure in a protein reduces the conformational entropy by reducing the number of the statistical segments on the protein backbone (Saven & Wolynes, 1996).

Helices present in an unfolded protein tend to align generically, similar to nematic or cholesteric liquid crystal order, which assists folding since most helices in the folded state tend to be aligned. Aligned helices gain steric entropy relative to non-aligned helices, in that their excluded volume is reduced so that aligned helical residues suffer essentially no steric entropy loss upon collapse of the polymer (Saven & Wolynes, 1996). Aligned helices may grow while reducing the steric entropy loss and as a helix grows it gains an extra energy in hydrogen bonds proportional to its length. In contrast, the analogous process of stacking of β-sheets was only observed in aggregates and therefore plays an important role in protein misfolding rather than the formation of native protein structure (Plotkin and Onuchic, 2002).

It is also very important in folding that the protein sequence is composed of both hydrophilic and hydrophobic amino acids, which phase separate in water, analogous to micellar formation. This process may be thought of as microphase separation (Camacho & Thirumalai, 1993; Pande et al. 1994). Local concentrations of non-polar residues along the sequence have been suggested to be nucleation sites for folding. In smaller proteins this process may remove enough entropy to leave only a small ensemble of nearly native states. In larger proteins, the inside and outside behave as polymer melts with constraints on the interface, and there is a significant amount of entropy left.

In spite of high cooperativity and fast rates of protein folding, most proteins larger than 100 residues tend to populate intermediates early in their folding. Intermediates may vary in their conformational properties and stabilities and be highly native like (Lorch, 1999), or contain native-like structure in regions corresponding to a domain or subdomain of the native protein (Cavagnero et al., 1999). Sometimes intermediates may contain highly non-native structures (Kuwajima et al., 1996). Some intermediates are on-pathway and can fold to the native state without undergoing substantial unfolding steps. They might play a role in folding by limiting the conformational search to the native state as these intermediates would significantly reduce the number of possible conformations during folding and thus allow protein folding to take place on a biologically relevant time scale (Brockwell et al., 2000). Some proteins may form misfolded or trapped intermediates that cannot fold to the native state without rearrangement of the elements These non-native species tend to accumulate and can either be rescued by molecular chaperones (Shtilerman et al., 1999) or can lead to aggregation (Fink 1999). Populated intermediates on the way from unfolded to the native state are local minima in the energy landscape (Radford et al., 2000; Troullier et al., 2000). If an intermediate cannot escape a local minimum, it becomes kinetically trapped and results in the so-called misfolded protein.

2.1.4 Methods for studying protein folding *in vitro*

To achieve a detailed description of protein folding pathways has been a fascinating problem and a challenge for researchers since the times of the first study by Anfinsen. During the past decade, major advances have been made in the available methodology. New methods were developed that allow monitoring rapid transitions between structurally dynamic ensembles and advances were made in theoretical approaches that allow these complex phenomena to be modelled (Dobson et al., 1999). Combining the results of a vast array of individual experiments on different protein systems, trends and patterns in the folding mechanisms of the proteins are beginning to emerge, alongside several successful attempts of predicting protein structure entirely from its amino acid sequence (Alm et al., 1999). A range of experimental techniques that have recently been developed and applied in the studies of protein folding is summarized in Table 2.1. Applied alone or in combination, they allow researchers to investigate the structure, dynamics, energetics and mechanistic properties of the denatured ensemble and collapsed species, as well as partially folded intermediates.

Table 2.1. Experimental techniques used to study protein folding

Technique	Timescale	Description (what is measured?)
Fluorescence		
Intrinsic fluorescence	ms	Environment around Trp and Tyr residues
ANS binding	ms	Burial of hydrophobic area
FRET	ms	Inter-residue distance
Substrate/inhibitor binding		Formation of native contacts
Anisotropy		Correlation time/mobility
Circular dichroism		
Far UV	ms	Formation of secondary structure
Near UV	ms	Formation of tertiary structure
Small-angle X-ray scattering (SAXS)	ms	Polypeptide shape and dimension
Absorbance (near UV)		Environment around aromatic residues
Hydrogen exchange		
Native exchange		Global protein stability
Pulsed hydrogen exchange NMR		Rate of hydrogen protection of backbone and amino acid side chains
Pulsed hydrogen exchange ESI MS		Rate of hydrogen protection on folding populations
Force spectroscopy using the AFM		Folding and unfolding rates
Solution-state NMR		Environment of protein side chains

Table adapted with modifications from Brockwell et al., 2000 and Radford et al., 2000.

Abbreviations: ANS: 1-anilino naphthalene sulphonic acid; ESI MS: electrospray ionization mass spectrometry; FRET: fluorescence resonance energy transfer; FITR: Fourier transform intra-red

Mapping the folding pathway of a protein requires identification of the intermediate species. However, for most proteins, the mechanism by which intermediates form is poorly understood, because the rate of formation of these species is too rapid for conventional stopped-flow experiments. Fast measurements of folding using the techniques described above will be required to elucidate these details.

Protein engineering can provide information about the role of individual side chains in stabilizing populated intermediates and transient high-energy transition states. One of such methods, called Φ value analysis, has been developed by Fersht and his colleagues. In this approach, an amino acid side chain is removed from the protein of interest and the effect of the mutation on the stability of the native protein is determined by equilibrium denaturation. The effect of mutation on the intermediate or transition state is determined and the ratio of these stabilities is known as Φ values are then compared. Determining many Φ values for residues throughout the native protein allows interfering with the structure of possible intermediates and the rate-limiting transition state (Fersht et al., 1992).

Fluorescence spectroscopy at single molecule resolution can be another powerful method to study different aspects of protein folding. The extremely sensitive nature of fluorescence spectroscopy allows extracting crucial information contained in the ensemble of molecules being studied as well as the time trajectories of individual molecules (Basak et al., 2014). For instance, the method of fluorescence correlation spectroscopy (FCS) provides sensitive information on the diffusion coefficient of a protein labelled with fluorescent dye (Moens et al., 1972; Krichevsky et

al.,2002) and single molecule Förster resonance energy transfer (FRET) has been successfully employed in several studies to characterize the conformational dynamics of experimental systems (Sharma et al., 2008; Chakraborty et al., 2010). In our study, among other methods, we present a novel way of applying FCS technique to measure protein folding rate at single molecule level.

The transition between the intermediate and native states for the proteins that populate intermediates is usually a rate limiting step. It involves the stabilisation and formation of final secondary structure, as well as conversion of disordered side chain conformations to the specific native state rotamers and packing of the side chains within the native structure. Recent advances in mass spectrometry made possible describing separate steps of folding for a range of proteins (Morozova-Roche et al., 1999; Englander et al., 2000). One of the major methods applied in the current study is the method of hydrogen/deuterium exchange monitored by mass spectrometry, which is described in more detail in the following chapter.

2.1.5 Hydrogen/deuterium exchange monitored by mass spectrometry

The method of hydrogen exchange mass spectrometry is based on a fundamental chemical reaction unique to amide hydrogens in proteins – a continuous exchange of amide hydrogens with the hydrogens in the surrounding solution (Figure 2.4). In usual aqueous buffer this reaction is undetectable; however, if all-H_2O solvent is replaced with an isotope of hydrogen such as deuterium then the exchange process can be followed. For most hydrogen exchange experiments, deuterated water (D_2O) is used (Englander 2000; Marcsisin et al., 2010).

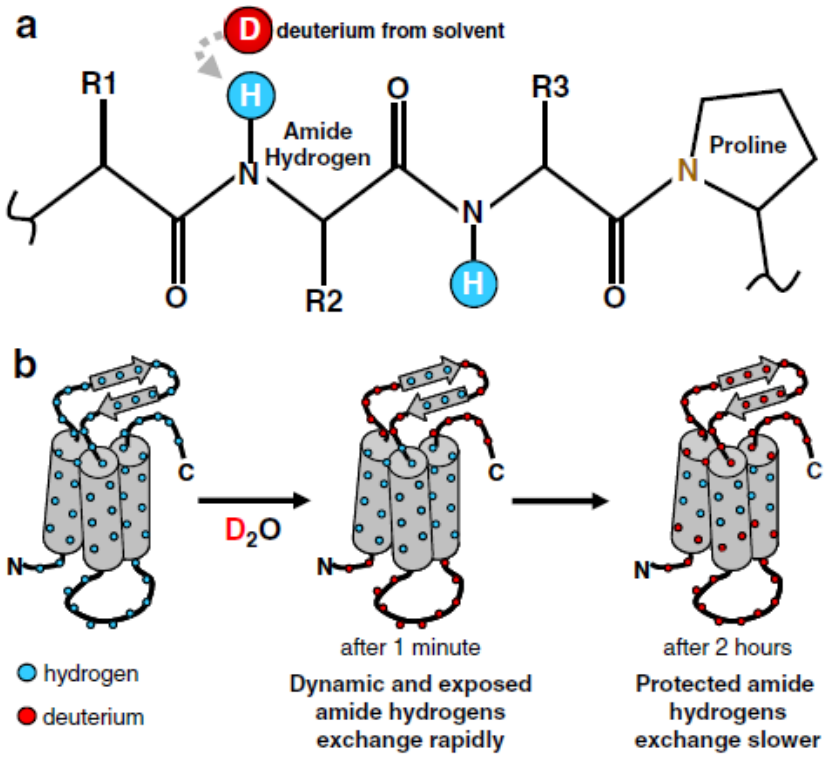


Figure 2.4 Amide hydrogen exchange in proteins.

(a) The backbone amide hydrogens (blue) are in continuous exchange with hydrogens in the solvent. Hydrogen bonded to carbon does not exchange. (b) The exposed regions (such as loops) of proteins exchange rapidly while protected compact regions exchange slower. (adapted from Marcsisin et al., 2010).

In folded proteins, amide hydrogens display a variety of exchange rates depending on their position within the protein and whether they are involved in intramolecular hydrogen bonding. The relationship between hydrogen exchange and solvent accessibility/hydrogen bonding in proteins is shown in Figure 2.4b. Highly dynamic and solvent-exposed regions (such as the loops connecting the alpha helices) will exchange rapidly whereas less dynamic regions and regions involved in hydrogen bonding networks (like as β-sheets or α-helices) will exchange slower (Smith et al., 1997).

As the mass of hydrogen is 1.0078 Da and the mass of deuterium is 2.0141 Da, deuterated proteins will have a larger mass than non-deuterated proteins (Katta et al., 1991). A typical workflow for a hydrogen exchange experiment monitored with mass spectrometry is presented in Figure 2.5. Protein solutions equilibrated at room temperature, physiological pH, and in all-H_2O buffer, are diluted 10- to 20-fold into the identical buffer except with all D_2O. The labelling is then quenched by adjusting the pH of the sample to 2.5 and lowering the temperature to 0 °C. These conditions decrease the rate of amide exchange up to ~5 orders of magnitude and thus ensure retention of the deuterium label for MS analysis (Smith et al., 1997). The quenched sample is sprayed directly into a mass spectrometer (using liquid chromatography) to determine the mass of the protein. Following quenching proteins can also be digested using various acid proteases, e.g. pepsin. With a digestion experiment, deuterium can be localized within the short peptides produced by the digestion.

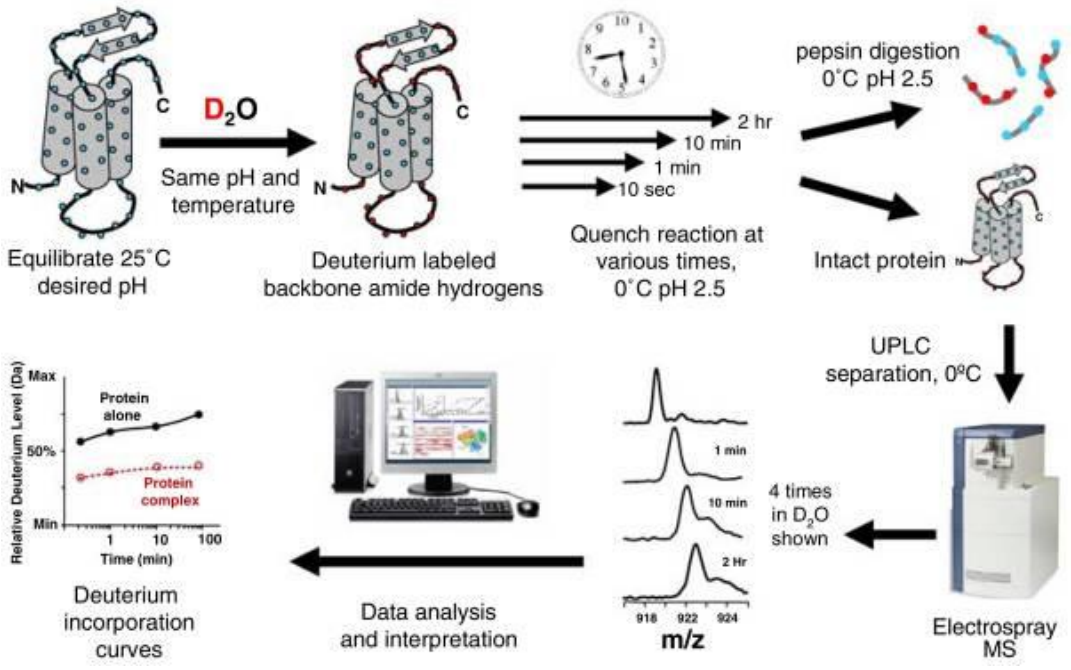


Figure 2.5 Workflow of a hydrogen exchange mass spectrometry experiment

Protein samples are equilibrated at the desired temperature and pH, in aqueous buffer. Protein solutions are then diluted with the identical buffer containing 99.9% D_2O instead of H_2O. The exchange reaction is quenched by lowering the pH to 2.5 and the temperature to 0 °C. Deuterated, quenched protein can then be directly injected into a mass spectrometer or digested with a protease prior to liquid chromatography and mass analysis. The mass spectra and the uptake of deuterium over time are analysed for the full-length protein, or for each of the peptic peptides (adapted from Marcsisin et al., 2010).

Hydrogen exchange mass spectrometry can provide information about conformation changes upon ligand/inhibitor binding or protein-protein interaction, protein folding and unfolding pathways, as well as some insights into the structure of proteins difficult to crystallize. For instance, HX MS was used to study the structural changes induced by pH changes in the capsid protein of the brome mosaic virus (Wang et al., 2001), to investigate conformational changes in the HIV-1 capsid protein during the HIV assembly and maturation (Lanman et al., 2004). Most recently, it has been applied to determine dynamic properties of processivity clamp proteins from different species, revealing a wide range of their dynamic behaviour despite a high level of tertiary structure conservation in these proteins (Fang et al., 2014).

HX MS also allows distinguishing populations of molecules in solution. Structurally different co-existing populations will incorporate different amounts of deuterium and this can be observed in the raw mass spectra. If the refolding rate of unfolded and exposed regions of a protein is slower than the deuterium labelling rate (Figure 2.5), the unfolded species will be highly deuterated and therefore have a higher mass than the folded species (Marcsisin et al., 2010). The rate of conversion of the folded to the unfolded species will indicate the rate of protein unfolding

in solution. The process of protein folding can be followed in a similar setting, providing us with an advantageous tool in the current study.

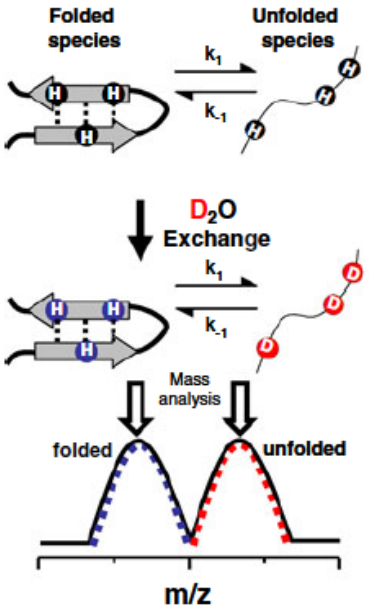


Figure 2.6 - Using hydrogen exchange to monitor protein unfolding dynamics.

Two populations in the mass spectra represent the folded state (blue distribution) and the unfolded state (red distribution). The appearance of two distributions occurs when the rate of interconversion of the two populations (i.e., folded and unfolded) is slower than the amide exchange rate (Weis et al., 2006). Unfolding of the protein molecule leads to its full deuteration, resulting in the higher mass (adapted from Marcsisin et al., 2010).

2.2 Protein folding in the cell

The intracellular environment is highly crowded, with protein concentration reaching up to 400 mg/ml, corresponding to a volume fraction of macromolecules of 20–40% of the total cellular volume (Zimmerman and Minton, 1993; Cheung et al., 2013). This leads to a volume exclusion or ‘macromolecular crowding effect’ (the term introduced by Minton in Minton, 1981), which impacts the behavior of biopolymers inside a cell. One of the major consequences

of the crowding effect is the increase in the intermolecular binding constants between partially folded states, which leads to an increased probability of aggregation during folding (van den Berg et al., 1999).

Various methods have been developed to mimic the macromolecular crowding effects, such as adding inert synthetic polymers (crowders: PEG, Ficoll, and dextran) into a test tube (Zimmermann et al., 1993). The assumption of crowders as hard core spheres allowed their modelling *in silico*. For example, the group of Elcock (McGuffee et al., 2010) created a detailed model of the *Escherichia coli* cytoplasm, including 50 of the most abundant types of macromolecules at experimentally measured concentrations. Their simulations were in accordance with the experimentally measured folding stability of several proteins in the cytoplasm of *E. coli*.

Conditions in the cell are likely to affect protein folding pathways and it has been speculated that protein structure could be dynamically 'tuned' by changing the microenvironment within the cell (Wirth et al., 2013). Different types of proteins might exploit properties of the cellular environment to increase folding efficiency. For example, the folding mechanism for a large protein is likely be altered by folding vectorially (from N to C terminus) during translation or secretion. Several proteins are known to have co-translational folding mechanisms in which the energy landscape for folding is significantly altered versus refolding *in vitro*, leading to significant amounts of native-like structure formation for the N-terminal portion of the chain. (Frydman, J. et al. 1999; Ugrinov et al., Biophys. J 2010; Braselmann et al., 2013).

A nascent chain of average length (~300 amino acid residues in *E. coli*) spends about 15 sec in the unfolded state on the ribosome, exposing hydrophobic residues, which has been earlier suggested to increase the potential risk for nascent chain aggregation and misfolding (Jaenicke et

al.,, 1991). However, more recent studies revealed that the three-dimensional organization of individual ribosomes in polysomes maximizes the distance between nascent chains and reducing the probability of unproductive interactions between nascent chains (Ellis and Hartl, 1999; Brandt et al., 2009).

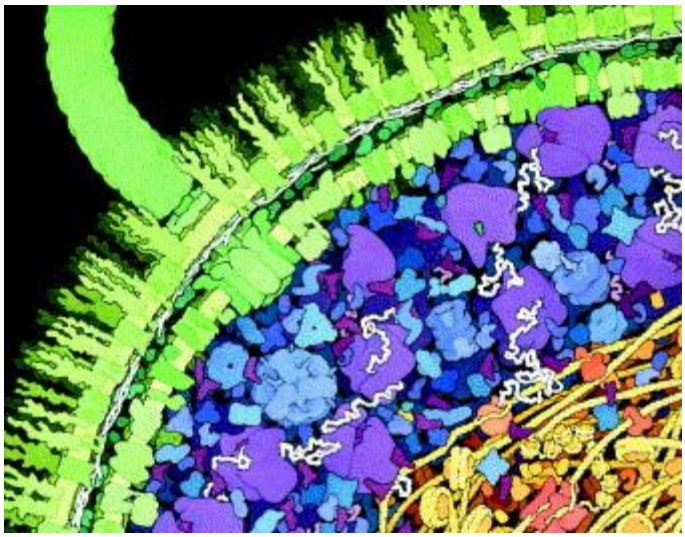


Figure 2.7. The crowded state of the *E.coli* cytoplasm

Representation of the approximate numbers, shapes and density of packing of macromolecules inside a cell of *Escherichia coli*. Small molecules are not shown. The sizes, shapes and numbers of macromolecules are in the order of actual cytosolic concentration. (adapted from Ellis et al., 2001).

In another study, the folding energy landscape in a cell-like environment was also explored for apoflavodoxin (an a/b protein) in the presence of Ficoll 70 by a combined approach of the far UV CD experiment and the coarse-grained molecular simulations (Stagg et al., 2007). The amount of experimentally measure α-helical content at a high concentration of Ficoll 70 was found to be greater than that in aqueous solution, whereas computer simulations showed more formation of native contacts upon the addition of inert Ficoll crowders, suggesting that the folding pathways of a protein under a heterogeneous intracellular condition may be distinct from those in a test tube.

Mutations in protein sequences that destabilize a protein can lead to their increased aggregation propensity in the highly crowded cytoplasm, as well as cause the loss of protein function. Aggregation of misfolded proteins creates toxicity (toxic gain of function) and may lead to severe disorders. For instance, retinitis pigmentosa mutations in the highly abundant photoreceptor protein rhodopsin affect its folding and transport and eventually result in photoreceptor cell death and blindness (Mendes et al., 2005). Serious neurodegenerative conditions, including Alzheimer's disease, Parkinson's disease, Huntington's disease and prion disease, result from the aggregation of a diverse set of peptides and proteins converting into amyloid-like fibrillar assemblies. Another serious disease involving amyloids is type II diabetes. The common structural feature of amyloid is its cross β-fold in which the protein, irregardless of its native structure, is converted into a β-strand fibril (Saibil, 2013).

Although the structural and mechanistic bases of cytotoxicity remain obscure, there is evidence for membrane damage by oligomeric intermediates in amyloidogenesis, as well as the overload of protein quality control systems. In healthy individuals, these processes are prevented or rescued by a subset of protein known as molecular chaperones (Balch et al., 2008; Powers et al., 2009).

2.2.1 Molecular chaperone systems

Molecular chaperone can be defined as a protein participating in the folding or assembly of another protein without being a part of the final structure (Hartl FU. 1996) Molecular chaperones are a set of protein families that act on a variety of non-native substrates and assist in folding, unfolding and homeostasis of cellular proteins.

Chaperones can be classified into different groups on the basis of sequence homology and many of them are stress proteins (heat shock proteins, Hsps, as their synthesis is induced under

conditions of stress. Apart from their role in protein folding, chaperones have a wide range of functions in proteome maintenance, assisting in macromolecular complex assembly, protein transport and degradation, as well as dissociation of aggregates and refolding of stress-denatured proteins (Kim et al., 2013).

Chaperones are known to recognize non-native states of many different proteins by interacting with exposed hydrophobic sequences, which later are buried inside the native protein structure. Most of the main chaperones possess ATPase activity and require cycles of ATP binding and hydrolysis to act on non-native polypeptides, facilitating their folding or unfolding while others only protect nascent protein subunits during their assembly processes (Mayer et al., 2010).

Binding (and rebinding) of non-native proteins to chaperones prevents aggregation and reduces the concentration of free folding intermediates. Achieving efficient folding is possible when the rate of folding is faster than the rates of aggregation or chaperone rebinding. In those cases when protein folding is significantly slower, the protein is transferred to a different chaperone system – such is the interplay between, for instance, Hsp70 chaperones and the chaperonins (Hsp60s). (Kim et al., 2013). If the protein is unable to refold, it may be transferred to the degradation machinery. In cases when the concentration of folding intermediates exceeds the available chaperone capacity in vivo, protein aggregation occurs, which often induces further cellular stress response, increasing the amounts of chaperones.

2.2.2 The chaperone network in the cytosol

The cytosolic chaperone system is organized as a highly cooperative network, highly conserved throughout evolution (Figure 2.8) (Kerner et al., 2005, Albanese et al., 2006). In all 3 domains of life – bacteria, archaea, and eukarya, the newly synthesized polypeptide firstly interacts with ribosome-binding chaperones (trigger factor (TF), nascent-chain-associated complex (NAC), and specialized Hsp70s) (Del Alamo et al., 2011). Later it is transferred to the next tier of chaperones that do not have direct affinity to the ribosome, e.g. the classical Hsp70 system (Calloni et al., 2012). Protein may start folding co-translationally, while still bound to the ribosome, and finish post-translationally, as they are released from the or after being transferred to downstream chaperones (Hsp60s and Hsp90 system) (Bukau et al., 2000).

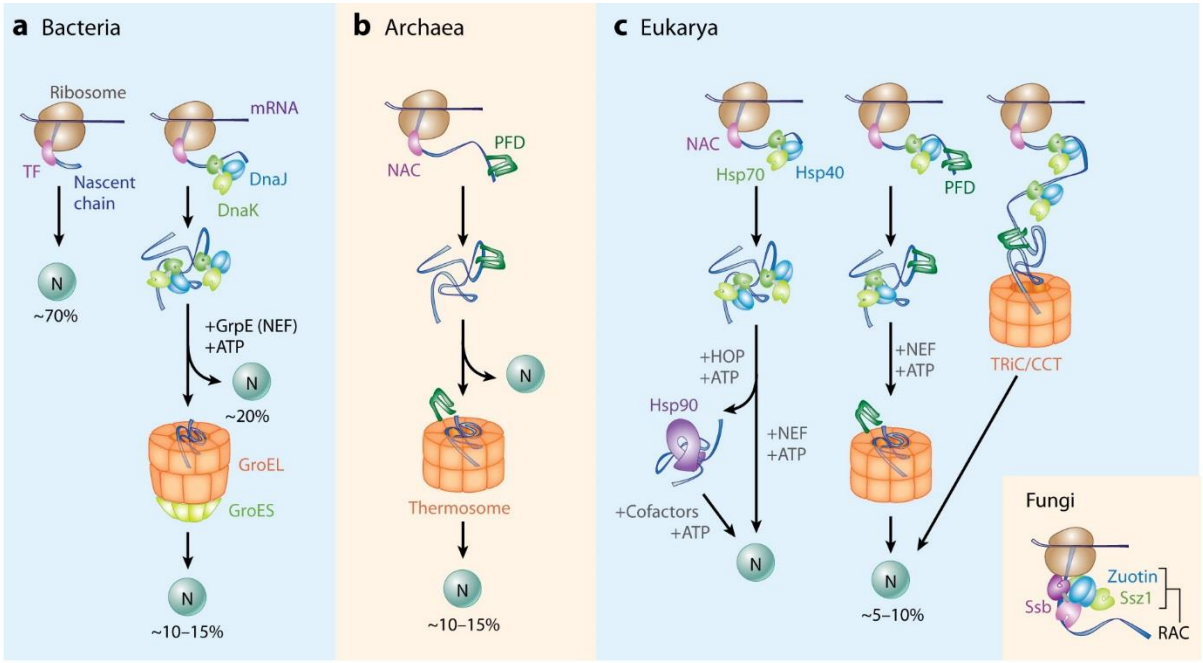


Figure 2.8 Organization of chaperone pathways in the cytosol.

Chaperone network in: Bacteria (a), Archaea (b), and Eukarya (c) Percentages indicate the approximate protein flux through the various chaperones. (adapted from Kim et al., 2013).

2.2.3 Ribosome–associated chaperones

The nascent polypeptide chain is topologically restricted on the ribosome. Since protein chain synthesis proceeds in a vectorial manner, the full network of long-rage interactions cannot be formed until the C-terminal region of the protein emerges from the ribosome tunnel. Therefore, nascent polypeptide chains usually expose extensive hydrophobic patches and are significantly prone to aggregation (Hartl et al., 2011). Ribosome-binding chaperones (trigger factor in prokaryotes and specialised Hsp70 complexes in archaea and eukaryotes) prevent emerging protein chains from aggregation and unfavourable interactions during translation, by shielding hydrophobic segments. (Figure 2.8) (Bukau et al., 2000, Preissler et al., 2012).

Trigger factor (TF) is an abundant bacterial protein of ~50 kDa, which interacts with most newly synthesized cytosolic proteins, binding to the large ribosomal subunit at the exit of the ribosomal tunnel. In vitro, TF was shown to bind to nascent chains as short as ~60 amino acid residues, whereas in vivo it binds ribosomes when nascent chains have reached ~100 amino acids in length (Oh et al., 2011). This allows time for the nascent chain on the ribosome to interact with a variety of targeting factors (e.g., signal recognition particle) and modifying enzymes (Ullers et al., 2003, Bingel-Erlenmeyer et al., 2008). TF is then released from the nascent chain in an ATP independent manner, permitting folding or transfer of the polypeptide to downstream chaperones such as DnaK, the major Hsp70 chaperone in bacteria (Calloni et al., 2012).

TF is absent in eukaryotes, however, other structurally unrelated chaperone systems such as ribosome-associated complex in *Saccharomyces cerevisiae* (RAC) and nascent chain-associated complex (NAC, in archaea and eukaryotes) may fulfil a similar role (Bukau et al., 2000, Gautschi et al., 2002). In fungi, RAC cooperates with ribosome-binding isoforms of Hsp70.

NAC is a heterodimeric complex of α- and β- subunits, with molecular masses of 33 kDa and 22 kDa, respectively, which associates with ribosomes and short nascent chains (Preissler et al., 2012). Although the exact role of NAC in folding is not established, in yeast, NAC function appears to be reminiscent of the interplay between TF and DnaK in bacteria (Koplin et al., 2010)

2.2.4 The Hsp70 system

In bacteria and eukaryotic cells, chaperones of the Hsp70 family form a central "hub" in the cytosolic chaperone network (Frydman et al., 2001, Calloni et al., 2012). They interact with multiple nascent and newly synthesized polypeptides, directing them for refolding or translocation between cellular compartments, as well as participate in protein disaggregation and transferring substrates to the degradation machinery (Saibil 2013). 700 cytoplasmic proteins were identified as Hsp70 interactors *in vivo*, among which 180 proteins are particularly Hsp70-dependent due to their high aggregation propensity (Calloni et al., 2012). Hsp70 chaperones exist in many orthologs in different cellular compartments and are highly interactive, functioning with many partners and cofactors.

The role of Hsp70 in folding was proposed to be stabilizing the unfolded state or unfolding proteins until they can spontaneously fold upon reaching their correct cellular destination (Saibil 2013). Upon release from Hsp70, polypeptides were shown to collapse into their native fold in free solution or to rebind to Hsp70 when they failed to reach the correctly folded state (Sharma et al., 2010). Proteins that are unable to utilize Hsp70 for folding are transferred to the chaperonin or the Hsp90 system (Kim et al., 2013). In addition to its role in folding, Hsp70 participates in the disassembly of the clathrin coat on membrane vesicles disassembly after completion of clathrin-mediated endocytosis. Hsp70 also cooperates with Hsp110 in eukaryotes in disaggregating large aggregates (Rothnie et al., 2011, Shorter 2011).

The Hsp70 structure consists of two domains: an ATPase domain (often also referred to as a nucleotide-binding domain, or NBD) and a substrate-binding domain (SBD). The chaperone activity of Hsp70 depends on internal dynamic interactions between these domains and external interactions with co-chaperones such as the Hsp40 proteins (such as DnaJ in *E.coli*) and nucleotide exchange factors (NEFs, which stimulate ADP release and nucleotide exchange after ATP hydrolysis) (Mayer et al., 2010).

The ATPase domain has the same fold as actin and hexokinase, with two flexible domains surrounding a deep, nucleotide-binding cleft that closes around ATP (Saibil 2013, Figure 2.9a). The substrate-binding domain has a brick-like shape and consists of a β-sandwich subdomain with a cleft capped by a mobile α-helical lid. Binding of Hsp70 substrate happens inside the cleft and is stabilized by closing the lid. The SBD binds to 5–7-residue hydrophobic peptide segments typically flanked by positively charged residues. The interaction between the SBD and the substrate backbone is mediated by hydrogen bonds, whereas binding of hydrophobic side chains of the substrate happens mainly through van der Waals contacts (Bukau et al., 1998, Kim et al., 2013). The substrate-binding domain can exist in two states – open state, which is stimulated by ATP binding to the ATPase domain, and closed state, which is triggered by ATP hydrolysis. The two domains are connected by a flexible hydrophobic linker, a key site in Hsp70 allosteric regulation.

The Hsp70 mechanism of action involves several key steps. In the ADP-bound or nucleotide-free state, the NBD is connected by a flexible linker to the SBD, with the lid domain locking a peptide substrate into the binding pocket (Zhu et al., 1996). ATP binding causes the closure of the nucleotide-binding cleft, creating a binding site on the NBD for the interdomain linker (Figure 2.9b). Linker binding causes the SBD and the lid domain to bind different sites on the

NBD, resulting in a widely opened substrate-binding site that enables rapid exchange of polypeptide substrates. After hydrolysis, the domains separate and the lid closes over the bound substrate. Such binding and release of extended regions of polypeptide chain are thought to unfold and stabilize non-native proteins either for correct folding or degradation (Saibil 2013).

Hsp70 acts in protein folding together with two co-chaperones – Hsp40 and NEF, which regulate its reaction cycle. The Hsp40 family is very diverse, with many specialized members targeting Hsp70 to specific sites or functions (Kampinga et al., 2010). All Hsp40 proteins contain a J-domain and act as the primary substrate recruiters for Hsp70 as well as stimulate the ATPase activity of Hsp70. They interact with both the nucleotide- and substrate-binding domains of Hsp70. The interaction of Hsp40 with Hsp70 stimulates the rate of hydrolysis of Hsp70-bound ATP to ADP over 1000-fold, leading to stable substrate binding by Hsp70 in the closed conformation. Subsequent binding of NEF to the NBD of Hsp70 catalyses the exchange of ADP to ATP, which in turn opens the SBD and initiates substrate release (Kim et al., 2013, Figure 2.9b). Individual levels of Hsp40 and NEF proteins in the cell are lower than those of Hsp70. However, eukaryotic genomes encodes multiple versions of J-domain proteins and NEFs, thus diversifying the Hsp70 function and its substrate specificities (Kampinga et al., 2010).

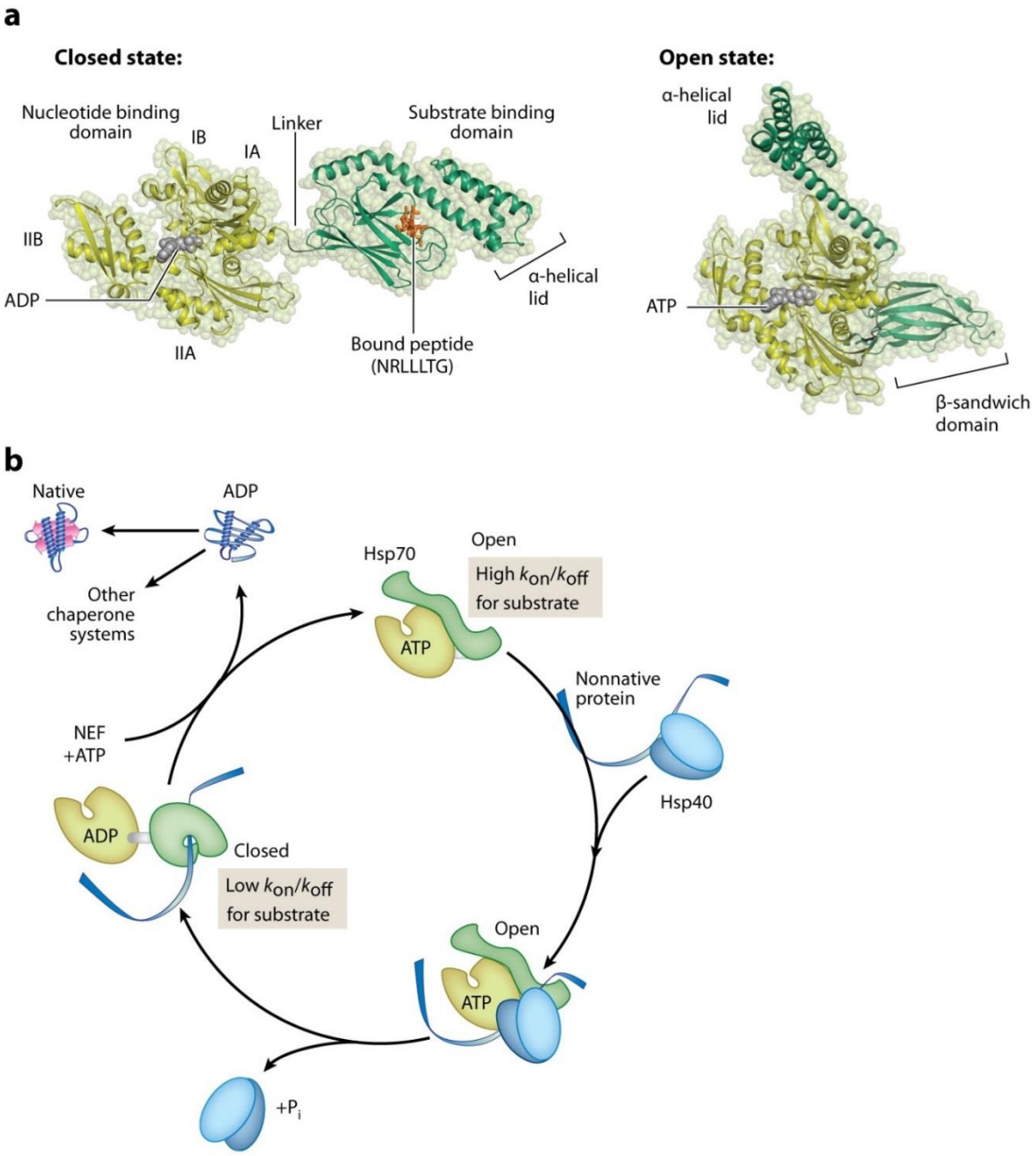


Figure 2.9. The structure and reaction cycle of the DnaK system

(A) Structure of the Hsp70 chaperone. Nucleotide-binding domain and substrate-binding domain are shown in yellow and green ribbons, respectively. (B) Reaction mechanism of Hsp70 chaperone. In an ATP-bound state, Hsp70 binds the non-native polypeptide delivered by Hsp40. ATP hydrolysis causes dissociation of Hsp40 from the complex and trapping the substrate in the substrate-binding cleft by closing the lid. NEF assists in exchanging the ADP to ATP, triggering substrate release (adapted from Kim et al., 2013)

2.2.5 The chaperonins

A special group within the broad class of molecular chaperones are the chaperonins - large, barrel-shaped protein complexes of ~800 kDa that consist of two rings stacked back to back and have the ability to bind and engulf unfolded/misfolded proteins. The chaperonins have an essential role in *de novo* protein folding and the refolding stress-denatured proteins (Hemmingsen et al., 1988, Goloubinoff et al., 1989, Bracher et al., 2011). The central cavity formed by each ring constitutes an active site in which a substrate protein is bound, encapsulated and released via a complex multi-step cycle (Lucent et al., 2009). Chaperonins interact with 10% of all cellular proteins and are therefore required for cellular viability (Hartl & Hayer-Hartl 2002, Fenton & Horwich 2003). Interestingly, unlike GroEL/ES or thermosome, the TRiC/CCT chaperonin of the eukaryotic cytosol is not stress-inducible (Horwich et al., 2007).

Chaperonins are divided into groups I and II. Group I chaperonins are present in bacteria (GroEL/ES) and cellular organelles of endosymbiotic descent, such as mitochondria and chloroplasts. Each ring of the group I chaperonin contains 7 identical subunits and the substrate is encapsulated inside the tetradecameric complex capped by a co-chaperonin lid (GroES in bacteria) (Tang et al., 2007). Group II chaperonins are found in archaea (thermosome) and in the eukaryotic cytosol (TRiC/CCT). The ring of a group II chaperonin usually consists of eight or nine subunits, encoded by several different genes. For instance, thermosome is comprised of 2 to three different types of subunits, whereas TRiC/CCT chaperonin contains eight different subunits in each ring (Horwich et al., 2007).

The ring structures of chaperonins bind non-native proteins via a hydrophobic lining of an open ring and then mediate ATP-triggered release followed by folding to the native state in an

encapsulated cavity. Substrate binding to the hydrophobic surface of the apical domains may in some cases effectively unfold a misfolded substrate which is then allowed another attempt of refolding inside the hydrophilic cavity (Dill & Chan 1994). During ATP hydrolysis, folding inside a sequestered chamber allows the protein to travel down the smooth energy surface as compared with the rough funnel containing kinetic "traps" that would exist in the bulk solution (Brinker et al. 2001, Chakraborty et al., 2010). Folding in the cavity is followed by release into the bulk solution whether or not the polypeptide has reached the native state (Weissman et al. 1996). If the polypeptide is still non-native, another round of binding and encapsulation follows. In the cell, this can result in a kinetic partitioning among the different chaperones and proteases (Hartl and Hayer-Hartl, 2009).

Members of the two chaperonin families function via a similar overall mechanism, however the major difference can be observed in their architectures (Figure 2.10): Type I chaperonins employ a detachable "lid" structure (GroES/Hsp10) that binds to the GroEL tetradecamer following the binding of ATP type II chaperonins, on the other hand, employ a built-in α-helical protrusions of the apical domains to close the cavity (Figure 2.10b, top view) (Meyer et al., 2003).

The ATP-directed reaction cycles of the two chaperonin families are mostly similar, directed by virtually identical equatorial ATP-binding domains. In the GroEL system, a positive cooperativity of ATP binding to subunits within a ring occurs via a concerted mechanism, whereas there is a negative cooperativity of ATP binding between the two rings (Yifrach & Horovitz 1995). In the thermosome and CCT systems the binding of ATP to subunits within one ring was found to be sequential, however for both systems, negative cooperativity between the

rings ensures asymmetric behaviour of the complex as a two-stroke machine (Reissman et al., NSMB 2007, Zhang et al., 2010).

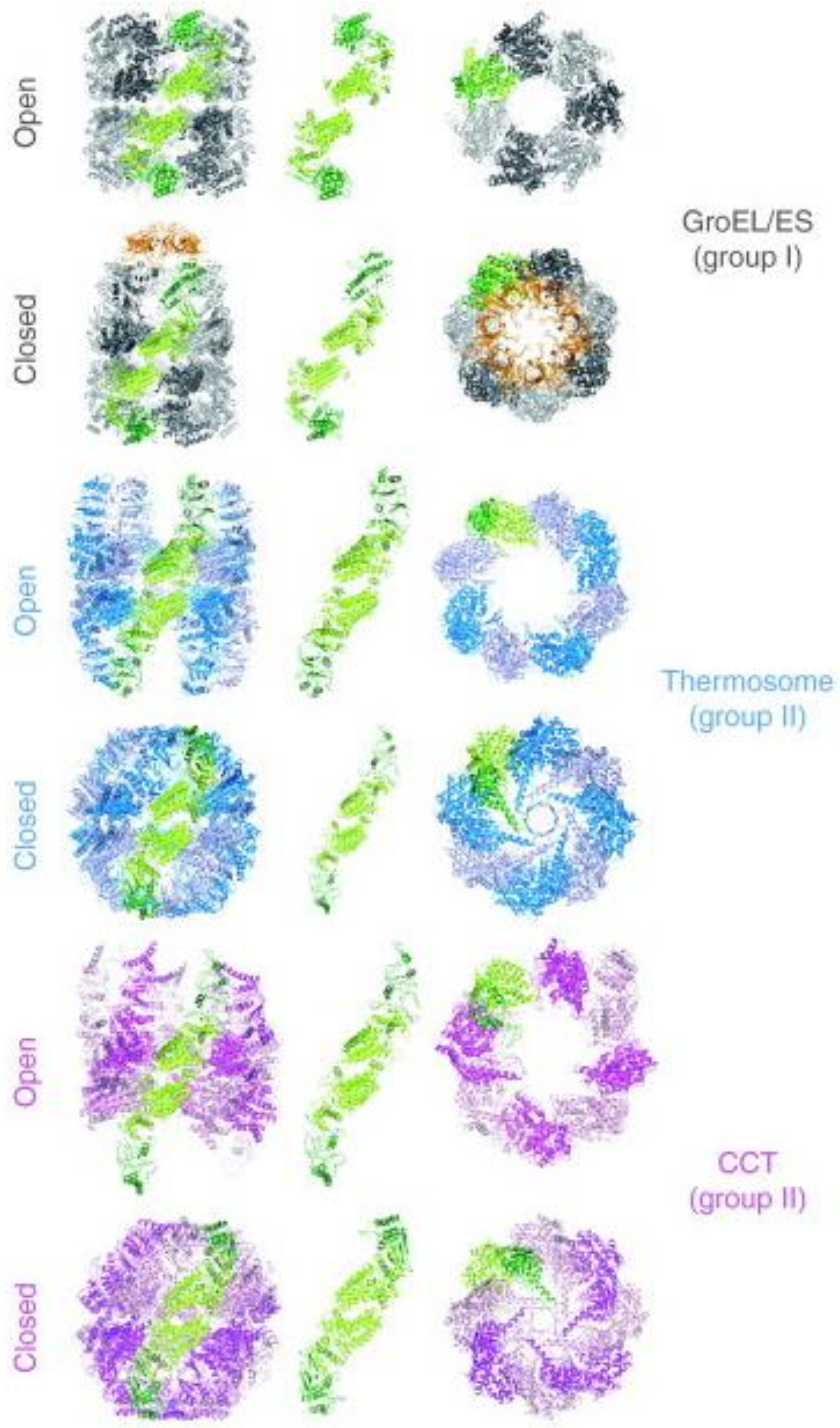


Figure 2.10. Structure representation of the chaperonins of Group I and Group II.

The first column shows a side view of each structure, highlighting the geometry of a pair of subunits, one in each ring. The middle column shows a detailed view of this pair. Lime green - the equatorial and intermediate domains; dark green – the apical domains. The third column

depicts a top view where the chaperonin is in the closed state. The structure coordinates used were: GroEL/ES closed 1AON, GroEL open 3E76, thermosome closed 1A6D, thermosome open (Cpn-Δlid) 3KFK, CCT closed 3IYG and CCT open 2XSM. Adapted from Yebenes et al., 2011.

2.2.6 The chaperonin system GroEL/ES of *E. coli* – structure and mechanism

GroEL/ES is the most extensively studied chaperone system to date. The crystal structure of an open (apo-) GroEL tetradecamer, without its co-chaperonin GroES, was first identified by Braig et al., in 1994. In an open state GroEL is a 15 nm long cylindrical structure of two back-to-back rings, each composed of seven 57 kDa subunits (Figure 2.11, a-c). The size of the GroEL molecular machine in its open state is about 146 Å in height and the central cavity is about 47 Å in diameter, with the walls lined by a band of continuous hydrophobic surfaces (Horwich et al., 2007).

Each GroEL subunit is composed of three domains, an equatorial domain located at the bottom of the ring and forming contacts with the other ring, an apical domain at the terminal end of the subunit and a smaller intermediate domain covalently connecting the other two allowing for rigid-body movements (Figure 2.11c, Saibil et al., 2013). Each equatorial domain contains an ATP pocket where binding and hydrolysis of the nucleotide occurs. The equatorial domains undergo subtle cooperative movements and are responsible for the asymmetric behaviour of the machine, ensuring that only one ring is folding active at a time (Clare et al., 2012). The apical domain, placed at the terminal end of the cylinder ~40Å from the equatorial ATP binding pocket, contains a hydrophobic polypeptide-binding surface facing inside the central cavity (Fenton et al., 1994). The seven apical domains of an open GroEL ring form a smooth hydrophobic surface that selectively captures non-native polypeptides (Farr et al., 2000). Finally, the intermediate domain ensures the flexibility of the subunit structure allowing the apical domains to open via

elevation and twisting movements (Fenton et al., 1994). The movements of the intermediate and apical domains, and of the machine overall, are rigid-body movements as was observed by different methods in free solution, such as cryo-EM and X-ray crystallography (Saibil et al., 2013).

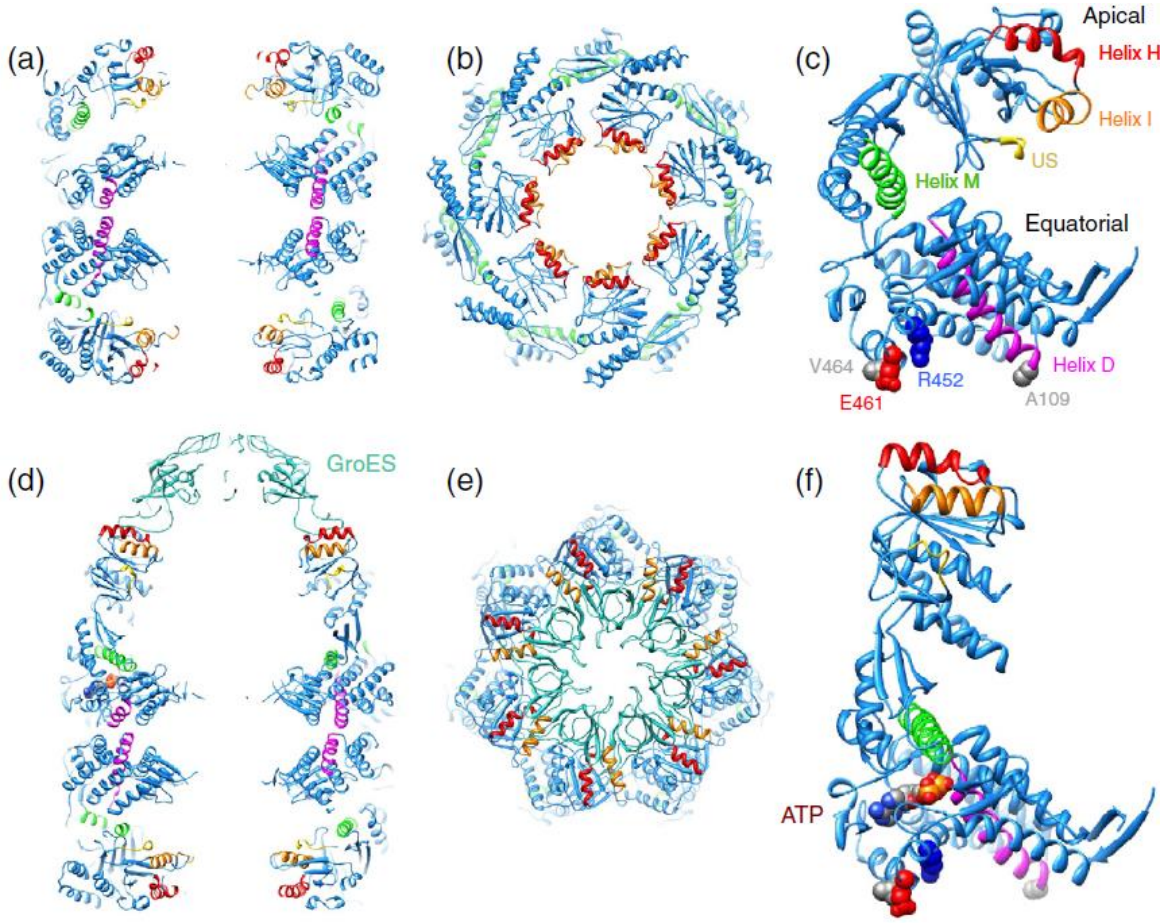


Figure 2.11. Crystal structures of GroEL and GroEL–GroES complexes.

(a) Side view slice through the GroEL structure (1OEL); (b) Top view from the outside of an open GroEL ring; (c) A subunit of GroEL. The apical and equatorial domains are labelled, as well as the main functional helices within each domain and the residues forming the inter-ring contacts (E461, R452, V464, and A109). (d–f) Similar views of the closed complex GroEL–GroES–ATP (1SVT). Adapted from Saibil 2013.

The main steps in the structural transition from an open hydrophobic ring to a closed folding chamber have been determined using computational methods as well experimentally using cryo-EM microscopy (Yang et al., 2009; Clare et al., 2012). The process is initiated by the substrate binding to the hydrophobic sites, followed by concerted ATP binding to a GroEL ring and initial contact with GroES, and concludes with additional rigid-body movements of the GroEL apical domain to form the closed GroEL/ES complex. The encapsulated substrate then remains inside the cage during the time required for ATP hydrolysis (half-time ~5-10 s), which allows substrate folding to take place. However, ATP hydrolysis is not required for protein folding but rather drives the chaperonin machine forward through its reaction cycle. Once ATP has been hydrolysed to ADP in the GroES-bound ring, seven other ATP molecules bind to the opposite ring, which in turn triggers the allosteric discharge of GroES, ADP, and substrate, from the folding chamber (Hartl et al., Nature 2011). If the released substrate is not correctly folded, it can rebind to the chaperonin and undergo another cycle of encapsulation inside the GroEL/ES complex (Rye et al., 1999). Scheme of the mechanism of GroEL/ES-assisted protein folding is depicted in Figure 2.12.

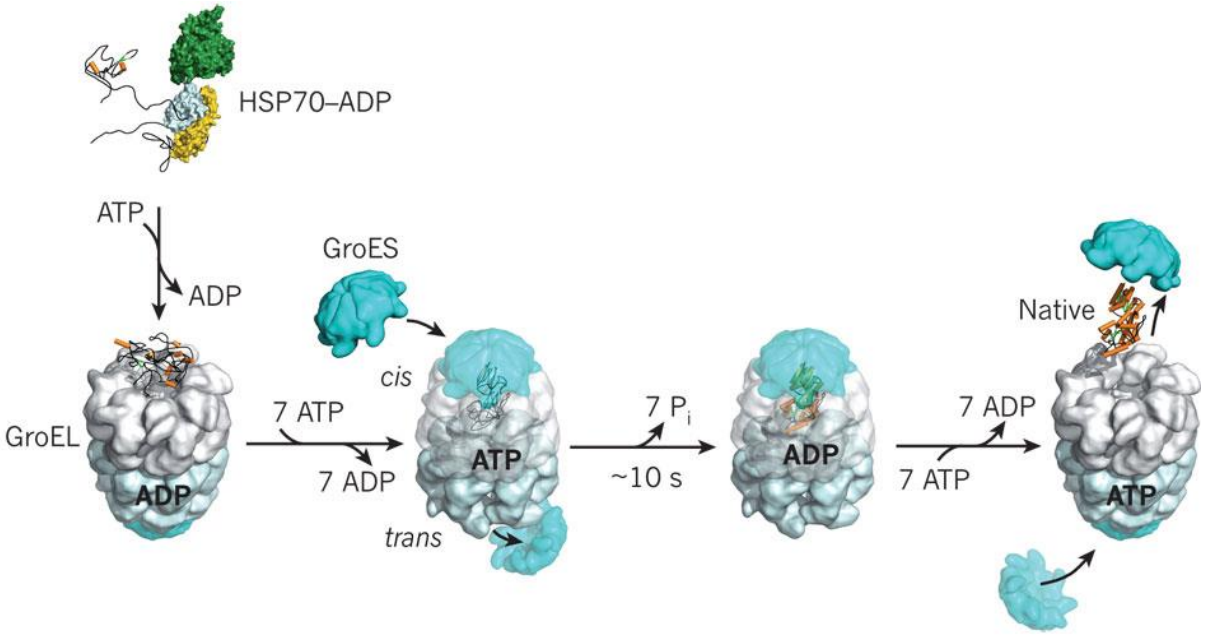


Figure 2.12. Folding in the GroEL–GroES chaperonin cage.

The substrate delivered by Hsp70 binds to the apical domains of GroEL, which leads to its partial unfolding. GroEL then binds seven molecules of ATP and GroES, forming a closed (cis-complex) with the substrate inside the cavity. Once ATP has been hydrolysed in a cis-ring, another unfolded substrate molecule binds the opposite (trans-) ring, together with seven other ATP molecules and GroES. This results in a dissociation of the cis-complex and a release of substrate from the cis-cavity. PDB structure 1AON was used to build the scheme. Adapted from Hartl et al., Nature 2011.

Chaperonins are intricate allosterically driven machines. Transition from the apo-state to the GroES-bound form, as has been shown through comparison of crystal structures, involves substantial rigid-body rotations about interdomain hinge points within each GroEL subunit, as well as local rearrangements of secondary structure elements within each domain (Xu et al., 1997). Cooperative ATP binding in the equatorial domains induces a downward rotation of the intermediate domains (Ranson et al., 2001, Clare et al., 2012). The presence of ATP and GroES causes a major elevation and twist of the apical domains (Roseman et al., 1996, Xu et al., 1997), moving the hydrophobic polypeptide-binding site on each apical domain (Helices H and I in

Figure 2.12) into a different orientation, slightly away from facing the cavity to a position where they bind to mobile loops of the GroES (Clare et al., 2012). A substantial overlap between the substrate binding and GroES binding part of the hydrophobic binding sites has been observed in mutational studies, explaining the mechanism of substrate displacement from the apical domains into the folding chamber without their escape into free solution (Fenton et al., 1994). It has been proposed that the continuous collar of binding sites is essential for productive folding of strictly GroEL-GroES dependent substrates, allowing substrate binding to several adjacent apical domains (Farr et al., 2000). The interaction of the substrate with the Group I chaperonins occurs via multiple hydrophobic interactions (Elad et al., 2007).

Several studies have shown that the substrate on the apical domain of GroEL is stretched and essentially unfolded (Falke et al., 2005, Elad et al., 2007, Sharma et al., 2008, Koculi et al., 2011). The mechanical forces applied on the substrate can potentially result in unfolding of trapped, misfolded proteins (Lin et al., 2008). Displacement of the substrate from the apical domains into the central cavity leads to a compaction of the substrate molecule in a confined environment of the chaperonin, with an upper limit of less than 60 kDa for protein subunits that can be encapsulated. During this transition, the hydrophobic binding surface of the apical domains is replaced with a hydrophilic, net negatively charged lining of the cavity inner wall (Xu et al., 1997; Chaudhry et al., 2003). The lack of exposed hydrophobic sites together with confinement, blocks further misfolding or aggregation pathways and promotes the refolding of the substrate. For substrates that are too large to be encapsulated, GroES may act allosterically facilitating productive folding of the substrate on the open ring (Paul et al., 2007).

A number of theories have been proposed in the literature since the discovery of the GroEL/ES machine describing how chaperonin acts on its substrates. The three major theories that ought to

be addressed in detail are passive cage theory, active cage theory and the theory of iterative annealing.

The Anfinsen cage model (passive cage) suggests that GroEL has only a minimal effect on the behavior of its protein substrates, does not modify the folding pathway or accelerate folding kinetics of its substrate (Chen et al., 2001; Horst et al., 2007; Tyagi et al., 2011). The passive cage theory suggests that the chaperonins encapsulate their substrates and act solely as aggregation-prevention devices by simulating an ‘‘infinitely dilute’’ environment inside the complex. According to a number of calculations, chaperonins significantly reduce the time that proteins spend in the cytosol before folding (Jewett and Shea, 2008).

Expansion of substrates on the apical domains of GroEL has been suggested to be a critical aspect of its function. It is known that many GroEL-dependent proteins typically endure several ATPase cycles before folding (Kerner et al., 2005). The average time for an obligate GroEL substrate to fold is ~30–60 s, equivalent to 3-10 ATPase cycles, assuming the speed of hydrolysis of approximately 7-10 s per cycle. The accumulated body of evidence suggests that the folding of these proteins is rate-limited by kinetic intermediates (van der Vies et al., 1992; Gorovits et al., 1998; Jennings et al., 1993). Some substrates may also encounter topological frustration, with incorrect intra-chain contacts made too early and therefore preventing a protein from reaching the native state (Onuchic et al., 2004). The theory of iterative annealing postulates that substrate denaturation via multiple cycles of binding and release and may free proteins from kinetic traps, leading to faster folding of frustrated proteins. Iterative annealing has been predicted mathematically to lead to a refolding rate acceleration as has been observed in various polymer simulations (Jewett et al., 2004).

A number of experiments, however, have demonstrated that GroEL/ES can assist protein folding in the absence of cycling (Weissman et al., 1996; Hayer-Hartl et al., 1996; Brinker et al., 2001; Chakraborty et al., 2010). Some other proteins can fold in presence of GroEL alone as has been shown for hen lysozyme (Coyle et al., 1999) and barnase (Zahn et al., 1996). Confining a protein in the chaperonin cage would eliminate extended conformations, thus reducing the conformational entropy of the unfolded state and increase the stability of the folded state. Eliminating a number of conformations with high entropy would also increase substrate folding rates as has indeed been shown in a number of studies. For instance, significant rate acceleration has been observed for a double mutant of maltose binding protein (DM-MBP) in the presence of GroEL and SR-EL (Tang et al., 2006; Chakraborty et al., 2010; Gupta et al., 2014).

Despite the vast amount of existing data, it remained unclear to date which of the mechanisms described above would hold true for a natural substrate of GroEL/ES, in particular, one of its obligate substrates. The study presented here, for the first time provides evidence that folding of a strictly dependent GroEL substrate inside the chaperonin cage is facilitated via an active cage, confinement-mediated mechanism.

2.3 Substrates of GroEL and GroES

Estimations of the amount of GroEL/ES in the cell under normal conditions indicate that no more than 5% of all cellular proteins can be actively using GroEL/ES for their folding (Lorimer et al., 1996). This early data prompted a series of studies aimed at identifying *in vivo* substrates of the chaperonin system.

The first study, which has been conducted by Hartl and colleagues, employed methods of protein immunoprecipitation in the presence of EDTA (to prevent ATP-dependent release of substrates) to isolate GroEL-bound substrates for their subsequent separation on 2D gels and identification by mass spectrometry (Houry et al., 1999). Among 300 proteins described in this study were found components of the transcription/translation machinery and metabolic enzymes. It was found that approximately one third the proteins were structurally unstable and required repeated binding to GroEL for their conformational maintenance. Chaperonin substrates were mostly proteins that contained several domains with mixed alpha/beta-folds, which exhibited relatively large hydrophobic surfaces and were highly aggregation-prone.

In a subsequent study, to ensure that substrate binding occurred in vivo and not during cell lysis, substrate-containing GroEL–GroES complexes were isolated from *E.coli* spheroplasts using immobilized affinity chromatography (Kerner et al., 2005). Subsequent mass spectrometry analysis led to identification of about 250 proteins interacting with GroEL under normal growth conditions. All substrates were partitioned into three classes based on the level of on their GroEL-dependency. Class I substrates (38 proteins) were found inside GroEL/ES complexes, however, their efficient folding can also occur in the absence of the chaperonin. Class II substrates (126 proteins) require GroEL/ES for their successful folding, but can also be assisted

by Hsp70 system chaperones. Finally, 84 proteins assigned to class III are stringent substrates of GroEL and at 37°C can only fold in presence of the chaperonin system. Importantly, the amount of proteins identified as Class III is similar to the earlier findings by the Hartl group in 1999 in vivo, which confirms the previously used method for isolating GroEL-substrate complexes from the living cells.

In a more recent study by an independent group, all previously found GroEL substrates have been confirmed and their abundance shown to be significantly reduced upon GroEL depletion in E.coli cells (Fujiwara et al,. 2010). Additionally, all substrates of Class III were re-examined and a new Class IV was formed to include the most stringent 49 substrates of the chaperonin that aggregated or degraded upon depletion of GroEL. Testing each of the Class IV proteins in vitro confirmed that they all indeed require GroEL/ES for their proper folding. Figure 2.13 presents a scheme of how all chaperonin substrates are subdivided into classes.

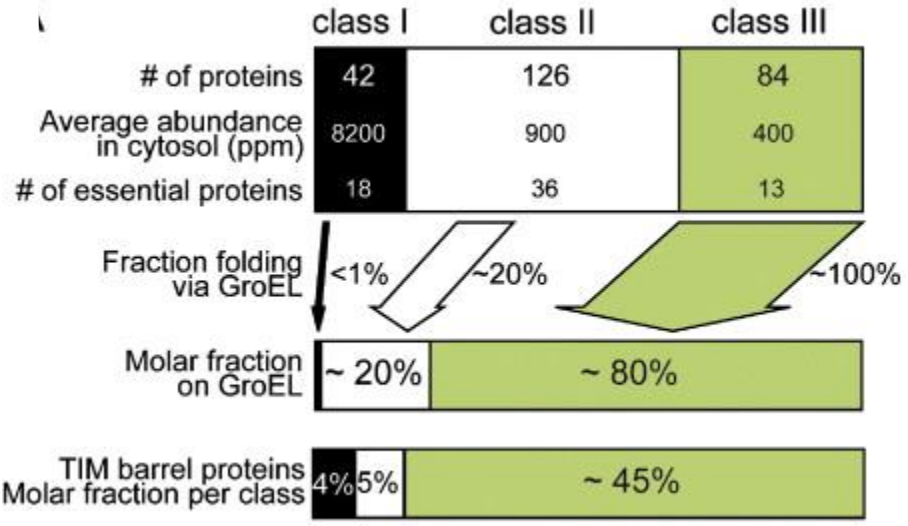


Figure 2.13. Classification of GroEL-interacting proteins.

All GroEL-interacting proteins were subdivided into three Classes, according to the extent of their chaperonin-dependence. Adapted with modifications from Kerner et al., Cell 2005.

All identified GroEL substrates have been extensively analysed in an attempt to unravel specific structural features or motifs that would explain their strong chaperonin-dependence. As mentioned above, these proteins exhibit increasing aggregation propensity, with Class III proteins being highly aggregation-prone. Another feature of the chaperonin substrates is their size range that corresponds to the properties of the GroEL cage, which only allows proteins of up to 60 kDa in size to be fully encapsulated (Horwich et al., 2007). Important and rather unexpected was the discovery by Kerner et al. that the almost half of the Class III substrates of GroEL share $(\alpha\beta)_8$ TIM-barrel, compared to only 6.8% TIM-barrel fold representation among all E. coli lysate proteins. On a large scale, however, it remained unclear to date, what distinguishes GroEL/ES-dependent TIM barrels from all other TIM barrels that fold without the chaperonin assistance. Several studies attempted to address this question employing computational approach, analysing such features as sequence length, overall hydrophobicity and charge, solvent accessibility, amino acid frequencies and codon optimality measures, but none of them showed significance difference between chaperonin-dependent and independent TIM-barrel proteins (Chapman et al., 2006; Azia et al., 2012).

Interestingly, GroEL sunstrates were found to be less hydrophobic than GroEL-independent proteins (Raineri et al., 2010), which is consistent with another finding by Niwa et al. that proteins with poor solubility (including members of classes III and IV) are not enriched in hydrophobic amino acids (Niwa et al., 2009). Importantly, homologs of the substrates of *E.coli* GroEL in *Ureaplasma urealyticum*, an organism that lacks the chaperonin system (Glass et al., 2000), were found to contain a higher amount of hydrophobic amino acids, suggesting a possible compensation for the chaperonin deficiency in this bacteria.

2.3.1 TIM-barrel protein fold

The TIM-barrel fold is the most common enzyme fold in the Protein Data Bank (PDB) and has been identified in about 10% of all known proteins. It is present in various enzyme families, catalysing completely unrelated reactions (Hegyi et al., 1999). The TIM-barrel domain is typically comprised of about 200-250 residues and can be present in single-domain enzymes, such as hevamine (Terwisscha van Scheltinga et al., JMB 1996), or be a part of large, multimeric, multi-domain enzymes such as β-galactosidase (tetrameric, five domains, 1023 residues per polypeptide chain as described in Juers et al., 1999).

Figure 2.14. Structure of a TIM-barrel fold, illustrated using trypanosomal TIM (5TIM in PDB, Borchert et al., Structure 1993).

A. Top view (along the barrel axis) of the TIM-barrel, into the active site. B. Side view of the TIM-barrel. Adapted from Wierenga et al., FEBS Lett 2001, with modifications.

The first protein found to carry a TIM-barrel was triose phosphate isomerase, which subsequently gave a name to this protein fold (Banner et al., 1975). The structure of a typical TIM-barrel fold (TIM protein from trypanosoma) is presented in Figure 2.14. Interestingly, the most hydrophobic region of the TIM-barrel was found to be not the inner part of the β-barrel, but the core region between the β-strands and α-helices (Nagano et al., 1999). The active sites of all TIM-barrel enzymes are located at the C-terminal ends of the β-strands and are shaped by residues of the eight loops following after the β-strands. Whereas the βα loops are important for the function, the αβ loops at the N-side of the molecule are believed to be more important for the stability of the TIM-barrel protein (Urfer et al., 1992).

No significant amino acid sequence similarity has been detected in TIM-barrel proteins from different families. However, complementing sequence alignments with structural analysis revealed the presence of physicochemically similar clusters of residues in the same topological positions of the TIM-barrel domain (Selvaraj et al., 1998). It was suggested that these residue clusters, defined by long-range hydrophobic and electrostatic interactions, may direct and determine the common TIM-barrel folding pattern. This hypothesis is supported by the data obtained from studying circular permutated sequence variants of a TIM-barrel enzyme TrpF, which have been shown to fold similarly to the wild-type enzyme (Luger et al., 1989). Another example comes from the proof-of-principle study by Sterner and co-workers, who have shown that a single point mutation is sufficient to convert a TIM-barrel enzyme HisA catalysing a step in histidine biosynthesis into a protein with TrpF, which in turn is a part of the tryptophan biosynthesis pathway (Jurgens et al., 2000).

An interesting feature of HisA protein and its homolog from the histidine synthesis pathway, HisF, is that these TIM-barrels have an internal twofold repeat pattern in their sequence, indicating that these proteins likely evolved from a common ancestor (Fani et al., 1994). It was suggested that the ancestor of these proteins appeared as a result of a gene duplication and fusion event, from a half-barrel folding unit. Indeed, the N-terminal and C-terminal half-barrel domains of HisF have now been shown to be separate folding units. The half-barrel constructs, HisF-N and HisF-C can be expressed separately and isolated as inactive homo-oligomers (Hoecker et al., 2001). Mixing the purified half-proteins or gene co-expression in vivo allowed to obtain fully functional heterodimers, suggesting that at least in some cases the TIM-barrel may have been formed by modular construction from half-barrel units (Gertl et al., 2001).

2.3.2 GroEL substrate - dihydrodipicolinate synthase (DAPA) from *E.coli*

Dihydrodipicolinate synthase (DAPA, EC 4.2.1.52) is an essential enzyme of the lysine biosynthesis pathway across different species. It was found in both plants and bacteria, but not in animals, rendering it a potential target for herbicides and antibiotics (Hutton et al., 2007). The reaction catalysed by DAPA is the condensation of pyruvate (pyr) and (S)-aspartate semialdehyde (ASA). The reaction is initiated by condensation of pyruvate with an active site lysine residue Lys161 forming a Schiff base, followed by a subsequent tautomerization and an aldol-type reaction with ASA, producing the acyclic enzyme-bound intermediate ASA:Pyr. Transimination of the ASA:Pyr intermediate is thought to yield the cyclic alcohol (S)-4-hydroxytetrahydrodipicolinic acid (HTPA), with simultaneous release of the active site lysine residue. The scheme of the DAPA-catalysed enzymatic reaction is shown in Figure 2.15.

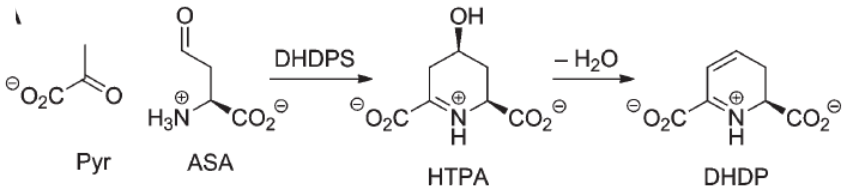


Figure 2.15. Scheme of the enzymatic reaction catalysed by DAPA.

Condensation of pyruvate and ASA to form HTPA, catalysed by DHDPS is then followed by dehydration to give dihydrodipicolinate (DHDP).

DAPA protein from *E. coli* is a homotetramer, existing in as a 'dimer of dimers' with a so-called 'tight' and 'weak' dimer interfaces (Figure 2.16). Residues T44, Y107 and Y133 of the catalytic site are located at the tight-dimer interface. Therefore, each tight dimer contains two complete active sites within the TIM-barrel domain and an allosteric site that binds two (S)-lysine molecules to mediate feedback inhibition and is located in a cleft between the subunits (Dobson et al., 2004). Residue Y107 from one subunit of the tight-dimer protrudes into the active site of the adjacent subunit, forming a part of a catalytic triad that is essential for the enzyme activity (Blickling et al., 1997). Proper assembly of the tetramer is required for catalysis as has been shown in a mutational study where a central residue in the weak dimer-dimer interface was replaced with glutamate (L197D) to produce dimeric variants of DAPA with severely reduced catalytic function (Griffin et al., 2010). If an additional mutation Y107W is introduced into the tight-dimer interface, the resulting enzyme exists in primarily monomeric state in free solution and is highly aggregation-prone, consistently losing its residual enzymatic activity within 24 hours after purification (Muscroft-Taylora et al., 2010).

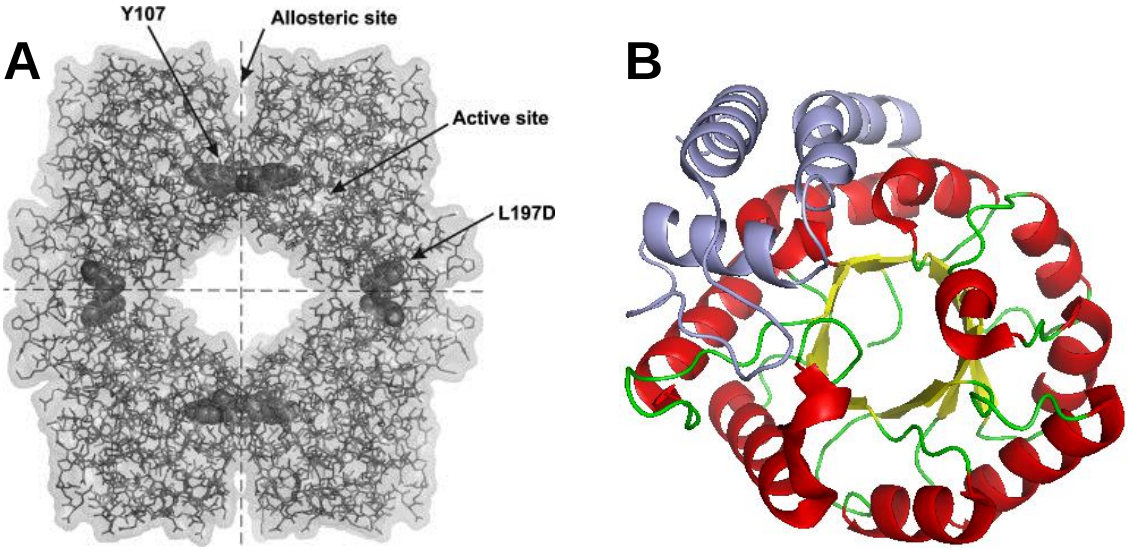


Figure 2.16. Crystal structure of dihydrodipicolinate synthase from *E.coli*.

A. Homotetrameric structure of *E. coli* DAPA indicating symmetry-related residues at the weak-dimer interface (L197D, horizontal) and tight-dimer interface (Y107W, vertical). The (S)-lysine allosteric binding site is located at the intradimeric cleft. Adapted with modifications from Dobson et al., 2004 B. Monomeric unit of *E. coli* DAPA illustrating N-terminal TIM-barrel domain and 3 C-terminal a-helices (light blue) (PDB 1yxc).

2.3.3 GroEL substrate - N-acetylneuraminate lyase (NANA) from *E.coli*

Another Class III substrate of GroEL carrying TIM-barrel domain is enzyme N-acetylneuraminate lyase (NANA, EC 4.1.3.3). NANA belongs to the dihydrodipicolinate synthase family of enzymes and is closely structurally related to DAPA (Izard et al., 1994). The enzyme splits N-acetyl-neuraminic acid (NAN) into its constituent N-acetyl-D-mannosamine and pyruvate and can, under suitable conditions, synthesize NAN from these same molecules (Figure 2.17A). Like DAPA, NANA is also a homotetrameric enzyme, with the monomers arranged as a 'dimer of dimers'. The active site of NANA is also located inside the TIM-barrel and is centred about the catalytic residue Lys165, which forms a Schiff base with the first substrate, pyruvate, to facilitate the required aldol chemistry, however, the substrates and products for the two

enzymes are very different. The sialic acids are a large family of sugars derived from the parent compound N-acetylneuraminic acid (Aisaka et al., 1991), widely utilised on the surface of cells in animals and on the surface of a number of pathogenic bacteria, where they are thought to provide camouflage from the host immune system or to facilitate tissue invasion.

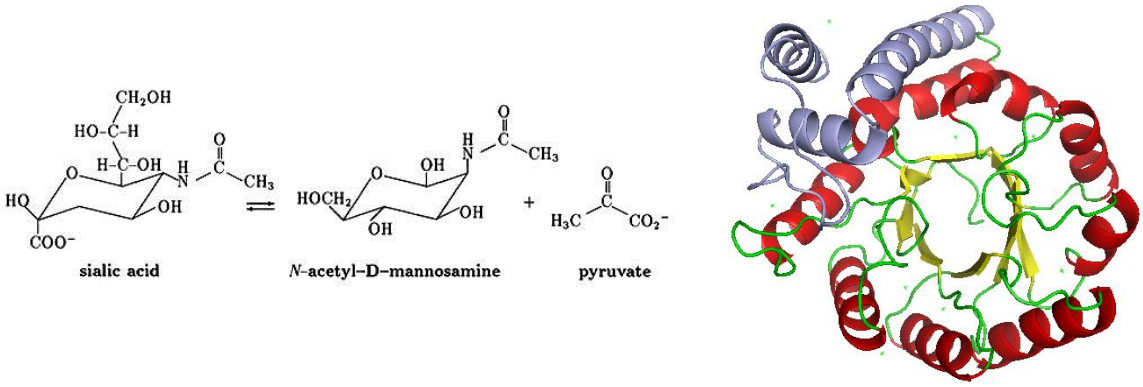


Figure 2.17. N-acetylneuraminate lyase – enzymatic reaction and crystal structure.

A. Cleavage of N-acetylneuraminic acid (sialic acid) by N-acetylneuraminate lyase. Adapted from Izard et al., 1994. B. Crystal structure of NANA monomer illustrating N-terminal TIM-barrel domain and 3 C-terminal a-helices (light blue) (PDB 1NAL).

3 Materials and methods

3.1 Materials

3.1.1 Chemicals

Acetic acid	Merck
Adenosine triphosphate, disodium salt (ATP)	Sigma-Aldrich
Agarose (SeaKem LE)	Cambrex Bio Science
Anti-rabbit IgG secondary antibody	New England Biolabs
Ammonium persulfate (APS)	Sigma-Aldrich
Ampicillin	Merck
Atto-532 C5 maleimide	Atto-Tec
Atto-647N C5 maleimide	Atto-Tec
Bacto agar	Difco
Bacto trypton	Difco
Bacto yeast extract	Difco
Bromophenol blue	Sigma-Aldrich
Calcium chloride	Merck
Complete EDTA-free protease inhibitor	Roche
Coomassie brilliant blue R-250	Roth
Dithiothreitol (DTT)	Roche
ECL™ detection kit	Amersham Pharmacia Biotech
Ethanol	Merck
Ethidium bromide	BioRad
Ethylenediaminetetraaceticacid –sodium salt (EDTA)	Merck
Glycerol	Merck
Glycine	Roth

Guanidium hydrochloride (GuHCl)	Sigma-Aldrich
HEPES	Sigma-Aldrich
Hydrochloric acid (37%)	Merck
Isopropyl-β-D-thiogalactopyranoside (IPTG)	BioMol
Magnesium chloride	Merck
β-mercaptoethanol	Sigma-Aldrich
Methanol	Merck
Phenyl-methyl-sulfonyl-fluoride (PMSF)	Sigma-Aldrich
Polyacrylamide/bisacrylamide solution 30 % (30 : 0.8)	Roth
Potassium hydroxide	Sigma-Aldrich
Rabbit anti-DAPA antibody	Sigma-Aldrich
Sodium chloride	Merck
Sodium dodecylsulfate (SDS)	Sigma-Aldrich
Sodium hydroxide	Sigma-Aldrich
N, N, N', N'-Tetramethylethylendiamine (TEMED)	Sigma-Aldrich
Tris-base	Sigma-Aldrich
Tween-20	Calbiochem

3.1.2 Enzymes

Benzonase	Merck
Lysozyme	Sigma-Aldrich
Pfu DNA polymerase	Stratagene
Restriction enzymes	New England Biolabs
T4 DNA ligase	New England Biolabs
Centricon 30 kDa cut-off	Amicon
Microcon 10 kDa cut-off	Amicon
Microcon 30 kDa cut-off	Amicon

Nitrocellulose transfer membrane	Whatman Schleicher & Schuell
Sterile filter 0.22 μm	Millipore
Sterile filter 0.45 μm	Millipore
Electrophoresis power supply Power PAC 300	Bio-Rad
Fluorescence spectrometer Fluorolog 3	HORIBA Jobin Yvon
FPLC systems	
EmulsiFlex high pressure homogenizer	Avestin
Gilson Pipetman (2, 10, 20, 100, 200, 1000 μl)	Abimed
HDMS ESI-QToF mass spectrometer	Waters Synapt
Incubators Innova 4430	New Brunswick Scientific
Luminescent Image Analyzer LAS-3000	FUJIFILM
Mini Trans-Blot Electrophoretic Transfer Cell	Bio-Rad
Microtime 200 confocal microscope	PicoQuant
PCR-Thermocycler T3	Biometra
pH meter Accumet Basic	Fisher Scientific
SMART system	Amersham Pharmacia Biotech
Sonicator Ultrasonic Processor XL	Misonix Inc.
Spectrophotometer DU 640 UV/VIS	Beckmann
Spectrophotometer LS50	Perkin-Elmer
Spectropolarimeter J-715	Jasco
Stopped-Flow Reaction Analyser Sx.18MV	Photo Physics
SZX10 Stereomicroscope	Olympus
Thermomixer Comfort	Eppendorf
Vortex	Ikamag
Water bath	Bioblock Scientific

3.1.3 Media

LB medium:	10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, (+ 15 g/l agar for solid medium). Adjusted to pH 7.0 with NaOH (Sambrook *et al*., 1989).
SOC medium:	20 g/l tryptone, 5 g/l yeast extract, 0.5 g/l NaCl, 0.186 g/l KCl, 0.95 g/l MgCl2. Twenty ml of filter sterilized 1M glucose were added after autoclaving (Sambrook et al., 1989).
MacConkey medium:	40 g/l of MacConkey base was autoclaved. Maltose was then added to a final concentration of 1% (solid medium).

3.2 Methods

3.2.1 Strains and Plasmids

The E. coli strains DH5α and BL21 (DE3) Gold (Stratagene) were used for cloning and protein expression, respectively. The genes *dapA* and *nanA* from E. coli were cloned into the vector pET22b (pET22b-DAPA; pET22b-EcNANA) and *nanA* from Mycoplasma synoviae into pET11a (pET11a-MsNANA) using Nde1 and BamHI restriction sites. The DAPA mutant, DAPA-293C (in which the three surface cysteines were replaced with seines, C20S, C141S, C218S, and an additional cysteine added to the C-terminus) was generated by QuikChange mutagenesis (Stratagene) of the wild-type *dapA* gene. The authenticity of each construct was confirmed by DNA sequencing.

3.2.2 Protein Expression and Purification

Chaperone proteins GroEL, GroES, DnaK, DnaJ and GrpE were purified as previously described (Brinker et al., 2001; Hayer-Hartl et al., 1996; Kerner et al., 2005). GroEL substrate proteins DAPA (dihydrodipicolinate synthase), DAPA-293C and NANA (N-acetylneuraminic acid aldolase) from E. coli (EcNANA) and from M. synoviae (MsNANA) were purified from the soluble fraction of BL21 (DE3) Gold cells.

Generally cells were grown at 20°C to an OD of 0.45 and protein expression induced by addition of 1 mM IPTG (Roth). After 4 h at 20°C, the cells were harvested by centrifugation (3200 x g, 45 min, 4°C) and subsequently frozen in liquid nitrogen as a suspension in 200 mM Tris-HCl pH 7.5, 10 mM DTT. Thawed cells were incubated for 1 h at 4°C in the presence of complete protease inhibitor (Roche), 1 mg lysozyme L-1 (42.35 U mg-1; Sigma) and 10 U benzonase L-1 (Novagen) and subsequently lysed on ice by sonication with a tip sonicator (Misonix Sonicator 3000, power output 7 in pulse mode, 10 x 30 s pulses interrupted by 90 s pause). All purification steps were performed at 4oC, the theoretical mass of the purified proteins confirmed by MS and the protein concentration determined spectrophotometrically at 280 nm. Chaperonin concentrations refer to the oligomeric state, while concentrations of the substrate proteins refer to the monomeric state.

3.2.3 DAPA Purification

DAPA protein purification was essentially as previously described (Laber et al., 1992) with minor modifications. After removal of cell debris and membranes by ultracentrifugation (200 000 x g, 45 min, 4°C), the supernatant was fractionated by chromatography on Source 30Q (Pharmacia Biotech) (20 mM Tris-HCl pH 7.5, 1 mM DTT, gradient from 0.1 M to 1 M NaCl). Fractions containing DAPA were pooled and adjusted to 20 mM Tris-HCl pH 7.5, 2 M (NH4)2SO4, 1 mM DTT and any precipitated protein was removed by centrifugation (3 200 x g, 30 min, 4°C). The supernatant was next applied to a phenyl-Sepharose CL-4B column(GE Healthcare) equilibrated in buffer (20 mM Tris-HCl pH 7.5, 1 mM DTT) and proteins eluted with a gradient from 2 M to 0 M (NH4)2SO4. Fractions containing DAPA were pooled, dialyzed against *buffer A* (20 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM DTT) and applied to a MonoQ

HR 16/10 column (Pharmacia Biotech) (20 mM Tris-HCl pH 7.5, 1 mM DTT, gradient from 0.1 M to 1 M NaCl).

DAPA containing fractions were pooled and subjected to Sephacryl S300 HiPrep 26/60 (Amersham Biosciences) gel filtration chromatography in buffer A. Fractions containing DAPA were concentrated at 4^{o}C using Vivaspin (MWCO 10 kDa; GE Healthcare) and supplemented with 5 % glycerol, flash-frozen in liquid nitrogen and stored at -80^{o}C.

3.2.4 EcNANA and MsNANA Purification

The purification of EcNANA and MsNANA was performed as previously published (Aisaka et al., 1991) with some modifications. After removal of cell debris and membranes ultracentrifugation as for DAPA above, the supernatant was subjected to heat precipitation at 70°C for 10 min and then immediately cooled on ice. Precipitated material was removed by centrifugation (20 000 x g, 4°C, 15 min) and the supernatant fractionated by chromatography on Source 30Q (Pharmacia Biotech) (20 mM Tris-HCl pH 7.5, 1 mM DTT, gradient from 0.01 M to 1 M NaCl). Fractions containing EcNANA (or MsNANA) were pooled and dialyzed against 20 mM Tris-HCl pH 7.5, 10 mM NaCl, 1 mM DTT. The protein was next applied to a MonoQ HR 16/10 column (Pharmacia Biotech) (20 mM Tris-HCl pH 7.5, 1 mM DTT, gradient from 0.01 M to 1 M NaCl). EcNANA (or MsNANA) containing fractions were pooled, dialyzed against 20 mM Tris-HCl pH 7.5, 0.01 M KCl, 1 mM DTT, concentrated using Ultricon concentrators (MWCO 10kD) (3 200 x g**,** at 4°C). In the case of EcNANA, the concentrated protein was applied onto a hydroxyapatite column (BioRad) (20 mM Tris-HCl pH 7.5, 1 mM DTT, gradient from 0.01 M to 0.1 M KCl) and the pooled fractions containing EcNANA were subjected to gel filtration on a Sephacryl S300 HiPrep 26/60 column (Amersham Biosciences) in buffer A. In the case of MsNANA, the concentrated protein was only subjected to gel filtration on Sephacryl

S300 HiPrep 26/60 column (Amersham Biosciences) in buffer A. After gel filtration, EcNANA (or MsNANA) containing fractions were concentrated at 4^{o}C using Vivaspin (MWCO 10 kDa; GE Healthcare) and supplemented with 5 % glycerol, flash-frozen in liquid nitrogen and stored at -80^{o}C.

3.2.5 Refolding, Assembly and Enzymatic Assays

3.2.5.1 Spontaneous Refolding of DAPA

DAPA was unfolded in 7.2 M GuHCl/10mM DTT or 7.2 M urea/10 mM DTT by incubation for at least 1 h at 25^{o}C. Spontaneous refolding was initiated by diluting the protein 100- to 200-fold into *buffer B* (20 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10 mM pyruvate) at the temperatures and final DAPA monomer concentrations indicated in the figures. Refolding was stopped after different times by addition of a 4-fold molar excess of GroEL D87K (GroEL Trap) (Farr et al., 1997) over DAPA monomer and DAPA enzyme activities were measured as previously described (Kerner et al., 2005) after incubation for at least 1 h at 25^{o}C to ensure complete assembly. Briefly, the product of DAPA enzymatic assay (2S,4S)-4-hydroxy-2,3,4,5-tetrahydrodipicolinate, reacts with o-aminobenzaldehyde under acidic conditions which results in the formation of diazaanthracene, purple chromophore absorbing at 550 nm (Mitsakos et al., 2011). Activities are expressed as percentage of the activity of the native protein.

3.2.5.2 GroEL/ES-assisted Refolding of DAPA

DAPA was unfolded as above and diluted at least 100-fold into refolding buffer B containing 2 µM GroEL/4 µM GroES. In order to ensure full capture of DAPA by GroEL the concentration of GroEL was kept constant at 2 µM as the concentration of DAPA was varied from 0.2 µM to 1.0 µM. Refolding was initiated by addition of 5 mM ATP and stopped at different times by adding an equal volume of 100 mM CDTA. DAPA enzymatic activity was measured as above.

3.2.5.3 SREL/ES-assisted Refolding of DAPA

Urea-denatured DAPA was diluted 100-fold into low salt *buffer C* (20 mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$, 10 mM pyruvate) containing 2 µM SREL and 4 µM GroES. Refolding was initiated by adding 2 mM ATP at 25^{o}C. At different time points aliquots were withdrawn from the refolding reaction and mixed with an equal volume of 100 mM CDTA and 60 mM GuHCl to stop the reaction and facilitate disruption of the SREL/ES complex. DAPA enzymatic activity was measured as above.

3.2.5.4 DAPA Assembly

Spontaneous and GroEL/ES-assisted assembly of DAPA was monitored in refolding reactions essentially as described above except that DAPA enzymatic activities were measured directly at the different time points without stopping refolding.

3.2.5.5 Spontaneous Refolding of E. coli NANA (EcNANA)

EcNANA was unfolded in GuHCl and refolding in *buffer D* (20 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM MgCl2) as described above for DAPA. EcNANA enzymatic activity was determined in a coupled assay containing 10 mM N-acetylneuramic acid (Sigma), 4.0 mM NADH (nicotinamide adenine dinucleotide), 3.0 U LDH (lactate dehydrogenase) (Roche) by measuring the first derivative of the decrease in absorbance of NADH at 340 nm (JASCO V-560 UV/VIS spectrometer). NANA catalyzes the cleavage of N-acetylneuraminic acid to form pyruvate and N-acetylmannosamine.

Activities are expressed as percentage of the activity of the native protein.

3.2.5.6 GroEL/ES-assisted Refolding of EcNANA

GuHCl-denatured EcNANA was diluted 100-fold into refolding buffer D containing 0.8 µM GroEL and 1.6 µM GroES. The final concentration of EcNANA was 0.4 µM. Refolding was initiated and stopped as described for DAPA at 25^{o}C.

3.2.5.7 Spontaneous Renaturation of Mycoplasma synoviae NANA (MsNANA)

MsNANA was unfolded in GuHCl as above and renaturation in buffer D was performed essentially as described for DAPA assembly with direct enzymatic assay as the read-out. MsNANA activity was determined by coupled enzyme assay as for EcNANA and the activities expressed as percentage of the native protein.

3.2.6 Circular Dichroism (CD) Spectroscopy

DAPA or MsNANA (1 μM) was incubated at 10°C for 16 h at increasing GuHCl concentrations in refolding buffer B. The steady state ellipticity was measured at 220 nm. Ellipticity was also measured immediately upon 67-fold dilution of denatured protein (7.2 M GuHCl/10mM DTT) into buffer B containing GuHCl to give final concentrations of 100 mM to 0.5 M GuHCl. Far UV CD spectroscopy was performed at 10^{o}C on a J-715 spectropolarimeter (JASCO) equipped with Peltier-thermostat using 0.1 cm cuvettes.

3.2.7 Fluorescence Correlation Spectroscopy

3.2.7.1 Protein Labeling

For single molecule experiments, the mutant DAPA-293C was labeled with either Alexa647 (Invitrogen) or Dy530 (Dyomics) using maleimide chemistry. The purified protein in buffer A/10 mM DTT was first buffer exchanged on a NAP5 column (Amersham Biosciences) equilibrated in buffer B and immediately mixed with a 1.2 molar excess of dye molecules and incubated for 30 min at 20°C. Free dye was removed using a NAP5 column equilibrated in buffer B/10 mM DTT and concentrated using Vivaspin (MWCO 10 kDa, GE Healthcare). The degree of labeling (DOL) was controlled with absorption spectroscopy (DAPA: ε280 = 12950 M^{-1} cm^{-1}; Alexa647: ε_{max} = 265000 M^{-1} cm^{-1} cf_{280} = 0.023; Dy530: ε_{max} = 100000 M^{-1} cm^{-1} cf_{280} = 0.15) using the following equation,

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

and found to be >90 %. The absence of free dye in the sample was confirmed by fluorescence correlation spectroscopy (FCS). Specific labeling at C293 was confirmed by mass spectrometry.

3.2.7.2 Fluorescence Correlation Spectroscopy

FCS measurements using pulsed interleaved excitation (PIE) (Muller et al., 2005) were performed on a Microtime 200 inverse time-resolved fluorescence microscope (PicoQuant), which was maintained at a constant temperature of 20°C. For excitation of Alexa647 and Dy530 picosecond pulsed diode lasers at 640 nm (LDH-PC-640B) and at 530 nm (LDH-P-FA-530) were used, respectively. Each laser had a laser power of 60 µW measured before the major dichroic. The lasers were pulsed with a rate of 26.6 mHz. The excitation light was guided through a water immersion objective (60 x 1.2 NA, Olympus) into the sample cuvette (Ibidi). The emitted fluorescence was separated from excitation light by a dichroic mirror (Z532/635RPC), guided through a pinhole (75 µm) and in case of cross correlation split according to wavelength by a beam splitter (600 DCXR) onto photon avalanche diodes (SPADs) (PDM series, MPD). The emission light was cleaned up by emission band pass filters (HQ 90/70 and HQ 580/70, Chromas) in front of the respective detector. Detection was performed using time correlated single photon counting, making it possible to correlate any given photon with the excitation source. In case of auto correlation measurements, after-pulsing artifacts were removed using fluorescence lifetime filters (Symphotime, PicoQuant) (Gregor and Enderlein, 2007).

3.2.7.3 Refolding Measured by Fluorescence Correlation Spectroscopy (FCS)

Refolding kinetics of spontaneous and assisted refolding were measured for 100 pM DAPA-293C-Alexa. DAPA-293C-Alexa was denatured as above. Spontaneous refolding was initiated

by dilution of the unfolded protein into buffer B to a final concentration of 100 pM at 20°C and stopped after different times by addition of 2 µM GroEL. For assisted refolding, the unfolded protein was diluted to 100 pM final concentration into buffer B containing 2 µM GroEL. Refolding was initiated by addition of 4 µM GroES and 5 mM ATP and stopped after different times by addition of Apyrase (Sigma). By stopping the folding reaction not-yet folded DAPA monomers will be bound by GroEL whereas folded monomers remain free in solution. The significant size difference of folded DAPA monomer (~31 kDa) and non-native DAPA in complex with GroEL (~830 kDa) results in different diffusion rates (102 ± 2 $\mu m^2\ s^{-1}$ and 49 ± 1 $\mu m^2\ s^{-1}$, respectively) which can be monitored using FCS. FCS measurements were performed within 10 min after stopping the reaction.

The auto-correlation data was fitted with the following one triplet one diffusion equation using the Symphotime software (PicoQuant):

$$G(\tau) = \left[1 - T + T \times e^{\left(-\frac{\tau}{\tau_T}\right)}\right] \times \left[\rho \times \left(1 + \frac{\tau}{\tau_D}\right)^{-1} \times \left(1 + \frac{\tau}{\tau_D \times \kappa^2}\right)^{-1/2}\right]$$

The mean diffusion time τD of particles through the focal spot is described by the structural parameter κ = z0/ω0 where z0 and ω0 denote the axial and radial dimensions of the confocal volume, respectively. The amplitude of the correlation function is denoted by ρ. The first term is used to compensate for fast dynamics arising from dye photophysics such as triplet blinking with the amplitude T on the timescale τT (Widengren et al., 1995). The diffusion coefficients were calculated using the following equation

$$D = \frac{\left(V_{eff} \times \pi^{-3/2} \times \kappa^{-1}\right)^{2/3}}{4 \times \tau_D}$$

by calibrating the confocal volume Veff with Atto655 dye, for which accurate diffusion parameters have been published (Müller et al., 2008). To analyze refolding kinetics the mean

diffusion time, reflecting a shift of DAPA molecules from GroEL-bound (D = 49 ± 1 $\mu m^2/s$, τD = 1.02 ± 0.02 ms) to free (D = 102 ± 2 $\mu m^2/s$, τ_D = 0.52 ± 0.01 ms) was plotted against the refolding time and fitted with a single exponential rate.

3.2.7.4 Dual Color Fluorescence Cross Correlation Spectroscopy

Dual color FCCS (dcFCCS) was employed to demonstrate the absence of inter-molecular association during spontaneous refolding of DAPA at 100 pM. DAPA-293C was labeled with either Alexa647 or Dy530 as above. The labeled proteins were denatured (at 5 nM in 7.2 M GuHCl/10 mM DTT) and diluted into buffer B to a final concentration of 50 pM each. FCCS was recorded with PIE during refolding at 20°C. As a positive control, the two labeled and unfolded molecule populations were mixed 1:1 at a concentration of 100 nM each and allowed to refold and assemble (note that folded DAPA will not assemble at 100 pM). The assembled tetramer was then diluted to a final particle concentration of 100 pM for dcFCCS analysis. A 1:1 mixture of Dy530 and Alexa647, again at 50 pM concentration each, was used as a negative control.

3.2.8 X-ray Crystallography

For crystallization of MsNANA, the screens Magic1 and Magic2 from the MPIB crystallization facility were used. An orthorhombic crystal form was obtained with a precipitant containing 50 mM MES pH 6.0, 4 % MPD, 0.2 M NaCl and 35 % PEG 400 (Magic1, condition C6) at 4ºC. A monoclinic crystal form was obtained with 50 mM MES pH 6.0 and 50 % MPD (Magic2, condition C12) at 4ºC. The crystals were directly picked from the drop and immediately cryocooled with liquid nitrogen.

3.2.8.1 Structure Solution

Diffraction data were collected at beamline X10SA of the Swiss Synchrotron Light Source (SLS; Villigen, Switzerland). The data were integrated with XDS (Kabsch et al., 2010). Subsequently programs as implemented in the CCP4i graphical user interface were used (Collaborative Computational Project, 1994). Pointless, Scala and Truncate were employed to scale the data and generate structure factor amplitudes (Evans et al., 2006; Evans et al., 1997). The orthorhombic crystal form was solved by molecular replacement using the program Molrep (Vagin and Isupov, 2001) with the coordinates of NANA from *H. influenzae* (PDB code 1F5Z; Barbosa et al., 2000) as a search model. After several rounds of refinement using Refmac (Murshudov et al., 1997), a nearly complete model was auto-build with ArpWarp (Langer et al., 2008). The model was completed by cycles of manual building with Coot (Emsley and Cowtan, 2004) and refinement with Refmac. This model served as a search model for structure solution of the monoclinic crystal form with Molrep.

3.2.8.2 Structure Analysis

For structure analysis the programs Lsqman (Kleywegt and Jones, 1994), Pymol (http://www.pymol.org/) and SC (Lawrence and Colman, 1993) were used.

3.2.9 Hydrogen/Deuterium Exchange (H/DX)

3.2.9.1 Spontaneous Refolding of DAPA and MsNANA

The reaction scheme shown in Figure 4.4A was followed. DAPA or MsNANA were denatured in 7.2 M GuHCl/10 mM DTT and diluted 100-fold into refolding buffer B at 10^{o}C to a final concentration of 2.4 µM. At different times of refolding, an aliquot of 40 µL was withdrawn and added to 360 µL of pulse-labeling *buffer E* (20 mM Tris-HCl, 20 mM KCl, pD 7.5, 99.9% D2O) to a final concentration of 90 % D2O. H/DX was allowed to proceed for 12 s before quenching

by addition of 8 µL of 10 % formic acid (final pH 2.5), followed by immediate analysis by LCMS.

As controls, denatured and native DAPA or MsNANA were subjected to D2O pulse-labeling and LC-MS analysis. To account for back-exchange, a reaction containing 99.9 % D2O was prepared by diluting denatured DAPA (6 M GuDCl/10 mM DTT) into refolding buffer B made with 99.9% D2O and after 12 s the reaction was quenched and immediately subjected to LC-MS at 0^{o}C.

3.2.9.2 Chaperonin-assisted Refolding of DAPA

GuHCl-denatured DAPA was diluted 100-fold into refolding buffer B containing 4.8 µM GroEL or 4.8 µM SREL. The chaperonin-DAPA complex was purified on a Superdex 200 gel filtration column and concentrated to 2.4 µM DAPA at 4oC using 0.5 mL Amicon Ultra filter (MWCO 30 kDa; Millipore). An aliquot of 50 µL of GroEL-DAPA (or SREL-DAPA) complex was mixed with 0.5 µL of 4.8 µM GroES in refolding buffer B (or buffer C in case of SREL-DAPA) and 1 µL of 250 mM ATP was added to initiate refolding at either 10^{o}C or 25^{o}C. At different times of refolding, aliquots of 20 µL were withdrawn and added to 180 µL of pulse-labeling buffer E. H/DX labeling was allowed to proceed for 12 s at 10°C before quenching as above, followed by processing of samples for peptide analysis (see below). For controls, GroEL-DAPA and SRELDAPA complexes were analyzed in the absence of ATP and GroES as above.

3.2.9.3 Intact Protein Mass Analysis and Data Processing

Protein samples (400 µL) were injected immediately after pulse-labeling and acid quenching into an Alltech analytical in-line guard column, packed with POROS 20-R1 reversed-phase media (PerSeptive Biosystems) that was placed in an ice-bath. 400 µL ice-cold 0.1 % formic acid was immediately injected (time required to complete this step was ~10 s) to wash the protein. The

guard column was then switched to position it in line with the HPLC inject port with a flow rate of 200 μL min^{-1} of 15% acetonitrile/0.1% formic acid, pH 2.5 for 2 min. The protein was eluted at 40 μL min^{-1} using a 1.5 min gradient of 15-75 % acetonitrile directly into a Waters Synapt G1 mass spectrometer with a standard electrospray interface. The injector, column and all associated tubings were kept at 0°C using an ice bath to minimize back exchange (Zhang and Smith, 1993). All experiments were conducted under identical conditions to allow comparison of relative deuterium levels (Wales and Engen, 2006); deuterium levels were not corrected for back exchange and are therefore reported as relative deuterium levels (Wales and Engen, 2006). Mass accuracy of less than 20 ppm was maintained by infusing horse heart myoglobin into the mass spectrometer at the end of each chromatographic gradient. Intact mass spectra were deconvoluted using MassLynx (Waters). All intact-protein H/DX-MS experiments were conducted at least in triplicate. Differences in the intensity of the species were observed but no new peaks were detected.

3.2.9.4 Peptide Mass Analysis and Data Processing

Each quenched sample was injected into a Waters nanoACQUITY UPLC with H/DX, as previously described (200 μL sample into a 100 μL loop) (Wales et al., 2008). The protein passed through a Poroszyme immobilized pepsin cartridge (Applied Biosystems) accommodated within the H/DX manager at a flow rate of 100 μL min-1 and temperature of 20°C. The peptic peptides eluting from the pepsin column were trapped and desalted for 3 min at 100 μL min^{-1} and then separated in 6 min with 8-40 % acetonitrile gradient in 0.1 % formic acid pH 2.5, at 40 μL min^{-1}. The cooling chamber of the H/DX manager, housing all the chromatographic elements, was held at 2.5°C. The separation column was a 1.0×100.0 mm ACQUITY UPLC C18 BEH (Waters) containing 1.7 μm particles (back pressure averaged 7800 psi). The average amount of

back-exchange using this experimental setup was 20-25 %, based on analysis of highly deuterated peptide standards. Deuterium levels were not corrected for back-exchange and are therefore reported as relative (Wales and Engen, 2006); all experiments were done under identical experimental conditions thus negating the need for back exchange correction. The UPLC was performed with protonated solvents, allowing deuterium to be replaced with hydrogen from side-chains and amino/carboxyl termini that exchange much faster than backbone amide linkages (Englander and Kallenbach, 1983). All experiments were performed between 2-5 times. The error of determining the deuterium levels was ±0.20 Da in this experimental setup (Burkitt and O'Connor, 2008; Houde et al., 2011). Mass spectra were obtained with a Waters Synapt G1 with a standard ESI source (Waters Corp., Milford, MA, USA) over an m/z range of 50-1700. Mass accuracy was ensured by calibration with Glu-fibrinogen peptide, and was less than 10 ppm throughout all experiments. Identification of the peptic fragments was accomplished with at least 4 replicate MSE (Plumb et al., 2006) analyses using Identity Software (Waters Corp., Milford, MA, USA). MSE was performed by a series of low-high collision energies ramping from 5–30 V, ensuring proper fragmentation of the peptic peptides. The mass spectra were processed with DynamX software (Waters Corp., Milford, MA, USA) by identifying the isotopic distribution (from +1 to +6 charge state, depending on the peptide). Isotope distribution and peak selection was verified manually for all peptides. In many cases, the isotope distributions were bimodal and all isotopes of the entire bimodal pattern were 15 selected for processing. The first-moment centroid m/z values of the distributions were calculated by DynamX and plotted with respect to refolding time. The unfolded reference or the GroEL-DAPA or SREL-DAPA references were arbitrarily given a refolding time of 0.05 min of the log time scale of these graphs and the native reference was given a value of 100 min. The relative

deuterium incorporation was calculated by subtracting the centroid of the isotopic distribution for peptide ions of the native reference from the centroid of the isotopic distribution for peptide ions from each refolding sample.

4 Results

This work was performed in close collaboration with Dr. Florian Georgescauld in the laboratory of Prof. Dr. F. Ulrich Hartl. Dr. Florian Georgescauld was involved in planning the experimental design as well as performing pulse-labelling H/DX coupled to liquid chromatography (LC) and MS experiments. Dr. Andreas Bracher solved the structure of MsNANA protein.

4.1 GroEL/ES accelerates the refolding of its natural substrate, dihydrodipicolinate synthase (DAPA)

4.1.1 DAPA is a Class III GroEL substrate that can refold spontaneously at low temperatures

In order to investigate the mechanism of GroEL/ES-assisted protein folding, a suitable substrate is required for which the spontaneous and chaperonin-assisted refolding could be compared.

A number of studies of GroEL/ES-assisted protein folding were performed using different model proteins such as lysozyme, rhodanese or DM-MBP, none of which is a substrate of GroEL *in vivo* (Weber F et al., 2000; Rozema et al.,1996; Chakraborty et al., 2010). It is of particular interest, therefore, to investigate folding of a natural substrate of the chaperonin, especially that of a strictly chaperonin-dependent (Class III) substrate identified in a recent study by Kerner *et al.,* 2005, and compare spontaneous and GroEL/ES-assisted refolding of such a protein. However, such studies are difficult, due to a high aggregation propensity of all obligate substrates of GroEL.

In our study, we found that one of the obligate substrates of GroEL (dihydrodipicolinate synthase, DAPA from *E.coli*) was able to refold spontaneously *in vitro* with a high yield at 25°C and at lower temperatures (Figure 4.1). The amount of enzymatically active protein observed

upon refolding of DAPA at different temperatures correlated well with the amount of soluble protein in the refolding sample analysed by SDS-PAGE (Figure 4.1, B and C). At a physiological temperature for *E.coli* growth, 37°C, no activity was observed upon refolding and all protein was insoluble. At 25°C, the recovery of DAPA enzymatic activity was ~75% and at 15°C and lower temperatures, the amount of active soluble protein reached almost 100%.

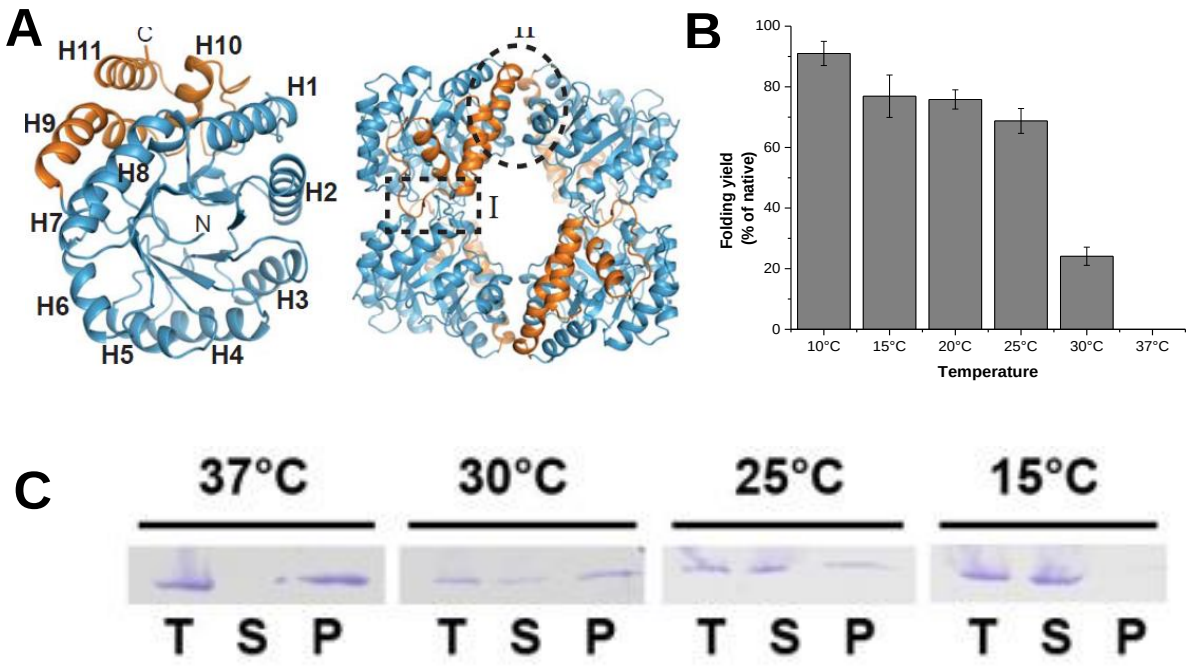


Figure 4.1: Dihydrodipicolinate synthase from *E.coli* (DAPA) refolds spontaneously at low temperatures.

(A) Structure of E. coli DAPA monomer (left) and tetramer (right) in ribbon representation (PDB 1YXC). The $(\beta\alpha)_8$ TIM-barrel domain is shown in blue and the C-terminal domain in gold. Helices H1 to H11 and locations of the strong and weak interfaces of the tetramer (interface I and II, respectively) are indicated.

(B) and (C) Yield of spontaneous DAPA refolding at different temperatures. (B) Refolding was initiated by diluting GuHCl-denatured DAPA into refolding buffer B to a final concentration of 200 nM monomer and yields analysed by enzyme assay after 1.5 h (15°C-37°C) and after 16 h (10°C). Folding yields are plotted as DAPA activities in % of native enzyme control incubated at the respective temperature. (C) Refolding was initiated as in B to a final concentration of 500 nM monomer and yields analysed by SDS-PAGE after ~3 hours. Total (T), Supernatant (S) and Pellet (P) protein fractions are shown.

4.1.2 DAPA is efficiently folded upon translation by GroEL/ES but not Hsp70 system

The above mentioned experiments represent the reactivation of a denatured protein. In order to investigate *de novo* DAPA folding, similar to what occurs *in vivo* upon translation, we performed cell-free synthesis of DAPA from a DNA template using the *in vitro* PURExpress system containing all necessary components for transcription and translation of a protein chain.

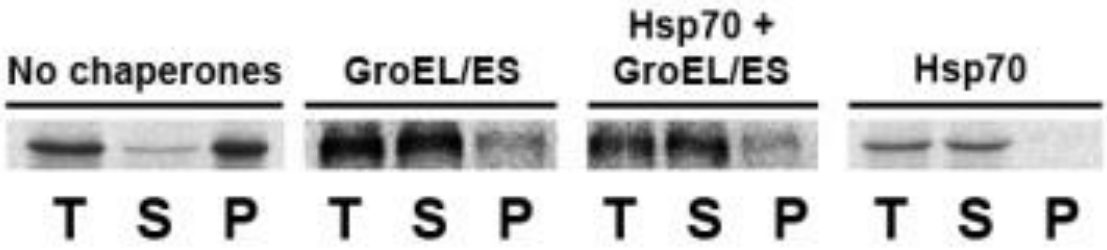


Figure 4.2: Aggregation of DAPA in cell-free transcription/translation system *In vitro* is prevented by DnaK/DnaJ/GrpE and GroEL/ES chaperones.

The coupled *in vitro* transcription and translation of DAPA was performed using the PURExpress system (New England Biolabs). DNA template containing the DAPA gene was added to the *in vitro* translation reactions containing no chaperones, Hsp70 system (DnaK/DnaJ/GrpE) alone, GroEL/ES alone or Hsp70 + GroEL/ES (see Materials and Methods for more details). Total (T), supernatant (S) and pellet (P) protein fractions of each sample were analysed by SDS-PAGE followed by autoradiography imaging (FUJIFILM Image Reader).

Like Class III GroEL substrates, DAPA does not fold spontaneously and upon synthesis in the absence of the chaperonin at 37°C the protein is resolved in the pellet fraction of the translation lysate (Figure 4.2). However, DAPA aggregation is efficiently prevented when either Hsp70 system (DnaK/DnaJ/GrpE) or GroEL/ES or both these systems are present in the solution.

In order to elucidate whether Hsp70 system alone could efficiently fold DAPA to its native state or if, as has been shown in Kerner *et al.*, 2005 it can only prevent aggregation of the Class III GroEL subastrates, we performed *in vitro* experiments of DAPA refolding at 37°C in presence or

absence of the chaperones. As shown in Figure 4.3, in the absence of chaperones DAPA is unable to regain its enzymatic activity whereas in the presence of Hsp70 system only about 40% of the substrate protein is refolded slowly within 1 hour.

Importantly, the apparent refolding rate of DAPA in presence of the Hsp70 system is significantly slower than even the rate of spontaneous DAPA refolding at 25°C or 10°C (Figure 4.4), (half-time of spontaneous DAPA refolding at 25°C is ~3.6 min, half-time of DAPA refolding with Hsp70 system is >1 hour). This phenomenon can be explained if the substrate is not stably bound to DnaK during the whole time of the experiment and individual molecules of DAPA are occasionally released into free solution where they can refold, aggregate or to be captured back by the Hsp70 system. Since the concentration of the free substrate molecules in the solution in this case would be extremely low, some of the molecules will have a chance to refold spontaneously which is reflected by the appearance of enzymatic activity. Simultaneously, a competition between spontaneous refolding and rebinding of unfolded molecules to Hsp70 would likely slow down the refolding, which is reflected by a half-time of the DnaK/DnaJ/GrpE-assisted reaction longer than 25 minutes. Aggregation, as shown later in this study (Figure 4.5), is irreversible and therefore has no contribution to the apparent refolding rate.

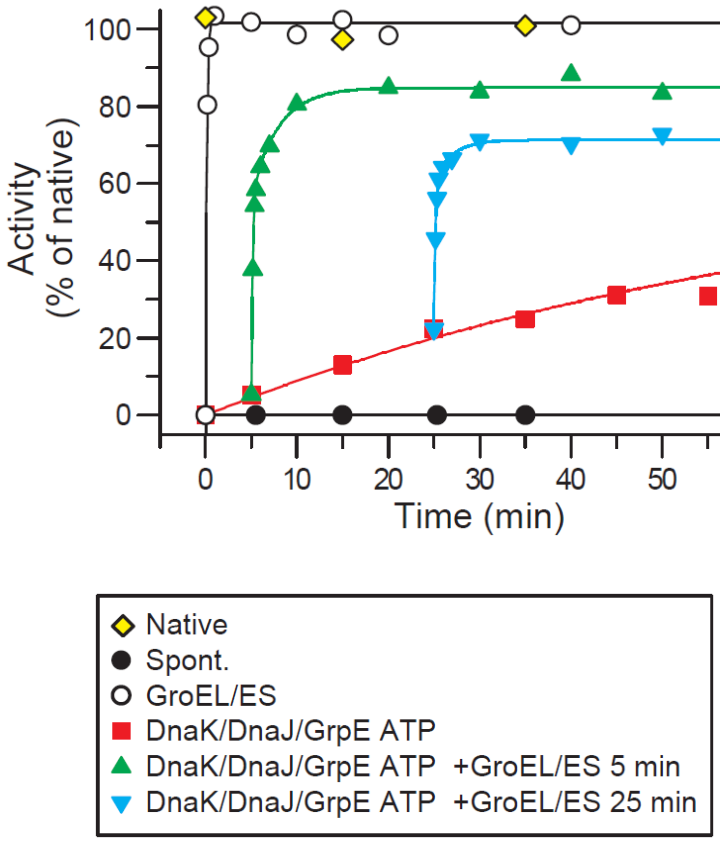


Figure 4.3: Hsp70 chaperone system keeps DAPA in a refolding-competent state for subsequent transfer to GroEL/ES at 37°C.

Refolding was measured at 200 nM final DAPA concentration by diluting denatured DAPA into refolding buffer B or containing 1.25 µM DnaK/0.63 µM DnaJ/1.25 µM GrpE or 2 µM GroEL/4 µM GroES or containing 1.25 µM DnaK/0.63 µM DnaJ/1.25 µM GrpE but with addition of 2 µM GroEL/4 µM GroES after 5 and 25 min after initiating folding. Assisted refolding was initiated by addition of ATP.

Of note is that upon addition of GroEL/ES into the refolding reaction that already contains the Hsp70 system, DAPA reaches its enzymatically active state with a high yield within only 1-2 minutes, which indicates efficient transfer of the substrate from Hsp70 system to GroEL/ES. When added into the reaction at later time points (25 minutes), GroEL/ES is only able to restore

~60-70% of DAPA activity compared to more than 80% when added early on, consistent with the slow release of substrate from the Hsp70 system and its subsequent aggregation in the solution. Taken together, our results show that the Hsp70 system efficiently maintained DAPA in a non-aggregated folding-competent state for its subsequent transfer to GroEL/ES and rapid completion of folding. Thus, the ability to promote DAPA folding is unique to GroEL/ES.

4.1.3 GroEL/ES catalyses the refolding of DAPA subunits

Formation of enzymatically active DAPA involves folding of individual subunits, followed by their assembly to dimers and finally tetramers (Reboul et al., 2012). To distinguish these processes and measure subunit folding independent of assembly, we stopped spontaneous refolding reactions by addition of GroEL D87K mutant to trap not-yet folded subunits (Figure 4.4A). This mutant of GroEL is unable to bind and hydrolyse ATP, which prevents the release of the captured unfolded molecules from its apical domains (Weissman et al., 1994). The GroEL/ES-assisted reaction was stopped by addition of CDTA to chelate Mg^{2+} ions and therefore inhibit the ATPase activity of the chaperonin, stopping the folding. To ensure that all individual DAPA subunits assemble into tetramers, we incubated all samples at 25°C for at least 1 hour prior to enzyme assay. The concentration of GroEL and GroES in the refolding reaction was kept close to their physiological levels (2µM GroEL and 4µM GroES in the reaction, as compared to ~3µM GroEL 14-mer in the cell (Ewalt et al., 1997, Mogk et al., 1999).

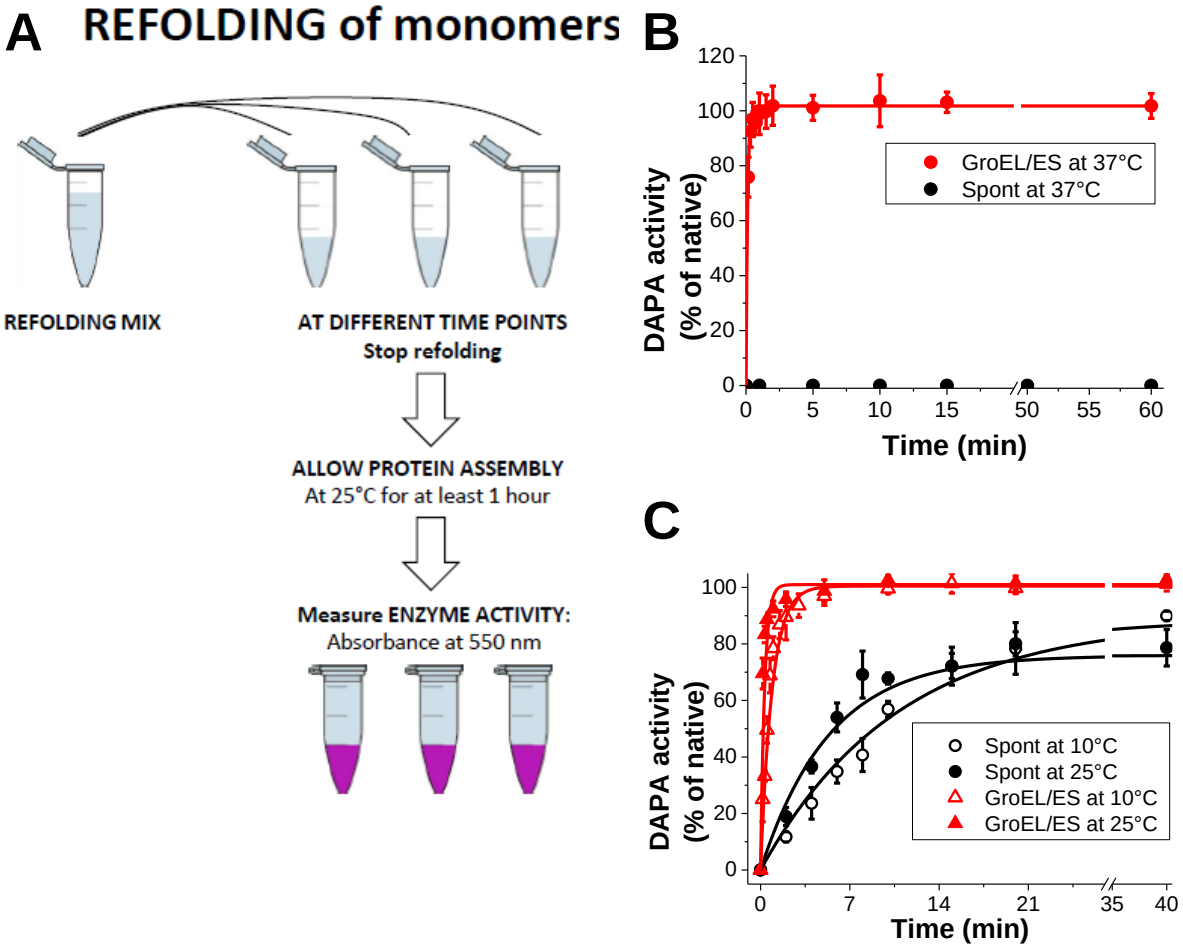


Figure 4.4: GroEL/ES accelerates DAPA subunit folding at different temperatures

(A) Scheme of the experiment measuring refolding of DAPA subunits.

(B) and (C). Refolding was measured at 200 nM final monomer concentration by diluting GuHCl-denatured DAPA 100-fold into refolding buffer B (20 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10 mM Pyruvate) or into buffer B containing 2 µM GroEL and 4 µM GroES. Assisted refolding was initiated by addition of ATP. Spontaneous refolding was stopped by addition of 0.8 µM GroEL Trap (GroEL D87K) and GroEL/ES-assisted refolding was stopped with 50 mM CDTA. Reactions were incubated for 1 h at 25°C to allow for complete assembly prior to enzyme assay. Single exponential rates are indicated. Standard deviations are from at least 3 independent experiments.

Upon dilution from GuHCl, DAPA refolded efficiently at 25°C and 10°C but aggregated at 37°C (Figure 4.4, B and C). The rate of spontaneous folding at 25°C was ~0.19 min^{-1} ($t_{1/2}$ ~ 3.6 min) whereas at 10°C the reaction was about 2-fold slower with a rate of ~0.09 min^{-1} ($t_{1/2}$ ~ 9 min). Strikingly, GroEL/ES accelerated the folding reaction about 30-fold at 25°C to a rate of ~6.0 min-1 ($t_{1/2}$ ~7 sec). At 37°C, where spontaneous refolding of DAPA could not be followed, the chaperonin-assisted reaction was even faster than at 25°C and a yield of 100% was reached within 20 sec after addition of ATP. It is important to note that the time needed to fold all substrate molecules at 37°C is comparable to the rate of protein synthesis in *E.coli* cell (~14 sec, assuming translation at a rate of ~20 amino acids per second; Hartl and Hayer-Hartl, 2009). It would therefore prevent the accumulation and possible aggregation of newly synthesized protein in the cell upon its translation. At 10°C, the acceleration of folding reaction by the chaperonin was ~16-fold with a rate of ~1.5 min^{-1} ($t_{1/2}$ ~ 28 sec). The yield of DAPA refolding in presence of GroEL/ES at all temperatures was 100%, in line with the previously obtained data showing that chaperonins efficiently prevent aggregation of substrate proteins.

4.1.4 Transient aggregation is not the reason for slow spontaneous refolding of DAPA

All above mentioned data strongly implies that GroEL/ES does not merely prevent aggregation of DAPA but is rather a protein folding catalyst. However, a recent report suggested that a reversible aggregation of the unfolded molecules could be the cause of the slow spontaneous folding of GroEL/ES substrates. In previous studies by Tang et al., 2006 and Chakraborty et al., 2010, aggregation of a heterologous chaperonin substrate, DM-MBP, was shown to be irreversible, which suggested that GroEL/ES can accelerate the intrinsic folding rate of a protein. Whether the same would hold true for the natural substrates of GroEL/ES remained to be tested.

To gain insight into this question we performed spontaneous and chaperonin-assisted refolding of DAPA over a wide range of substrate concentrations. If aggregation of substrate molecules were reversible and therefore affected refolding rates, we would expect to observe slower rates of spontaneous refolding at higher concentrations of the substrate. However, the rate of spontaneous folding was independent of DAPA concentration from 200 to 3500 nM at 10°C and from 200 to 1000 nM at 25°C (Figure 4.5). As can be further seen in Figure 4.6, partial aggregation of substrate protein at high concentrations resulted in a reduction in final yield but did not slow the apparent folding rate, indicating that aggregation was irreversible.

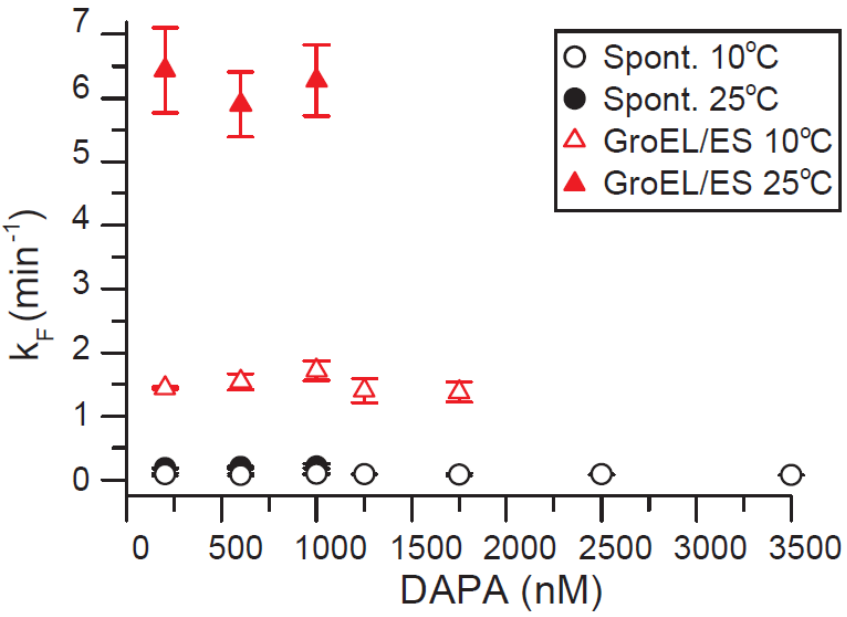


Figure 4.5: Rates of spontaneous and GroEL/ES-assisted refolding are concentration-independent. Spontaneous and GroEL/ES-assisted subunit refolding was performed at 10°C and 25°C as in Figure 4.4 over a range of DAPA concentrations. Standard deviations are from at least 3 independent experiments.

Significant protein aggregation precluded accurate rate measurements at >1000 nM at 25°C (Figure 4.6). However, in line with the notion that protein aggregation is a temperature-dependent phenomenon, at 10°C aggregation of DAPA was noticeably reduced. This later

allowed us to perform the refolding and assembly experiments at a high concentration of substrate and investigate DAPA folding in detail by methods of HD/X coupled to LC and MS (where the minimal concentration required to obtain good signal to noise ratio was ~2μM).

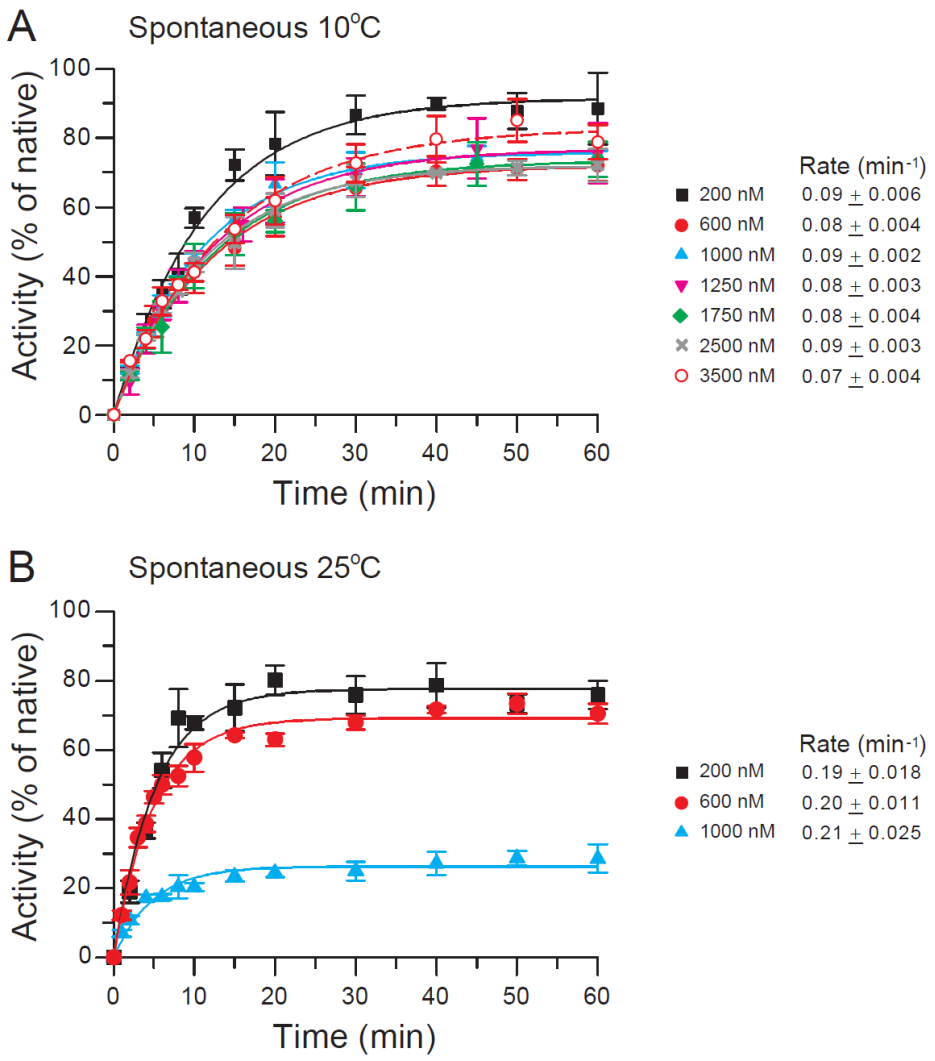


Figure 4.6: The rate of spontaneous refolding of DAPA at 10°C and 25°C is concentration-independent

Refolding was performed at 10°C (A) and 25°C (B) as in Figure 4.5 over a range of DAPA concentrations. Data were fitted with single exponential rates. Standard deviations are from at least 3 independent experiments. Note that the decrease in folding yield at 1000 nM DAPA in (B) is due to aggregation which does not affect the apparent rate of folding.

To further rule out aggregation as the cause of slow spontaneous folding, we used fluorescence cross-correlation spectroscopy (FCCS) to analyse refolding reactions at a very low DAPA

concentration (100 pM), where intermolecular association is excluded (Mukhopadhyay et al., 2007). A mutant of DAPA was designed in which the surface exposed cysteines (C20, C141, C218) were changed to serines and another cysteine was added to the C-terminus (DAPA-293C). DAPA-293C was labelled with Alexa647 or Dy530 dyes and the proteins mixed at equimolar amounts, denatured and allowed to refold at a concentration of 100 pM.

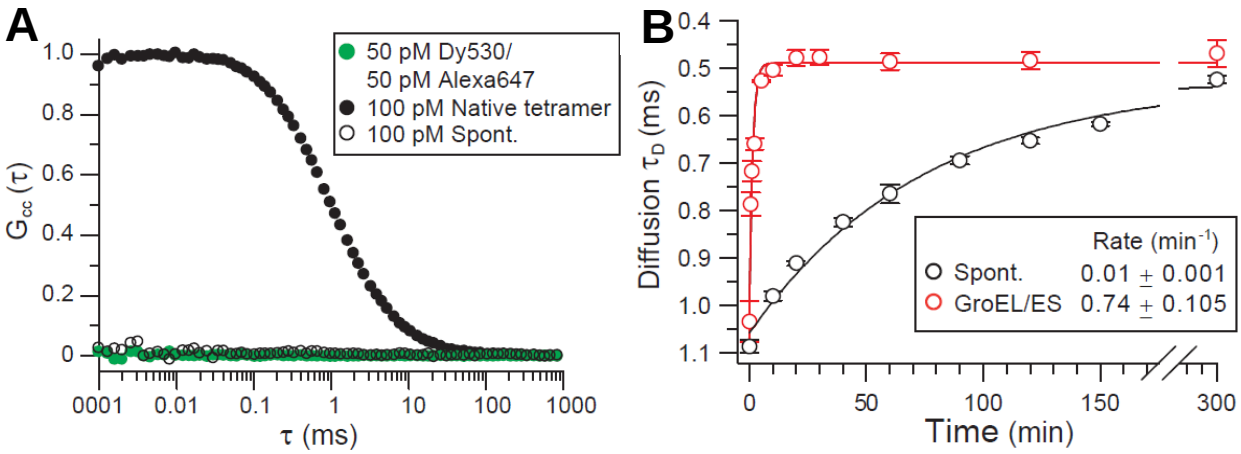


Figure 4.7. GroEL/ES catalyses DAPA refolding at single molecule level

(A) Absence of inter-molecular association during refolding by FCCS. DAPA-293C was labelled with either Dy530 or Alexa647 fluorophores. The labelled proteins were mixed, denatured and diluted into refolding buffer to a final concentration of 50 pM each. FCCS was recorded within the first 30 min of refolding. As a positive control, DAPA-293C tetramer that has been refolded and assembled from Dy530 and Alexa647 labelled subunits at 200 nM concentration was diluted to 100 pM. A 1:1 mixture of the free dyes at 50 pM each was used as a negative control. Experiments in this Figure were performed by Amit Jean Gupta.

(B) Spontaneous and assisted refolding of DAPA-293C labelled with Alexa647 was measured at a final concentration of 100 pM at 20°C. GroEL and GroES, when present, were 2 µM and 4 µM, respectively. Refolding was stopped either by addition of 2 µM GroEL (spontaneous refolding) or by addition of Apyrase (assisted refolding). The difference in diffusion rate between not-yet folded DAPA bound to GroEL and folded DAPA monomer free in solution was monitored by FCS, resulting in rates of subunit refolding. All standard deviations are from at least 3 independent experiments. Experiments in this Figure were performed by Amit Jean Gupta.

No FCCS signal was observed during DAPA refolding, indicating the absence of intermolecular association (Figure 4.7A). In contrast, cross-correlation was observed for 100 pM native tetramer when separately labelled proteins were first allowed to refold and assemble at 200 nM (Figure 4.7A). Taking advantage of the slower diffusion speed of the GroEL-DAPA complex and the ability of GroEL to trap not-yet folded DAPA molecules, we monitored the time dependent decrease in the average diffusion time of DAPA-293C-Alexa and extracted a refolding rate of 0.01 ± 0.001 min^{-1} (Figure 4.7B). Note that DAPA-293C-Alexa is enzymatically active but its spontaneous refolding rate is ~13-times slower than that of wild-type DAPA. Importantly, GroEL/ES accelerated the folding of DAPA-293C-Alexa (100 pM) more than ~50-fold at $20^{o}C$ (Figure 4.7B). These data demonstrate that GroEL/ES functions as a highly efficient catalyst of DAPA folding under conditions where subunit association (due to aggregation or productive assembly) is excluded.

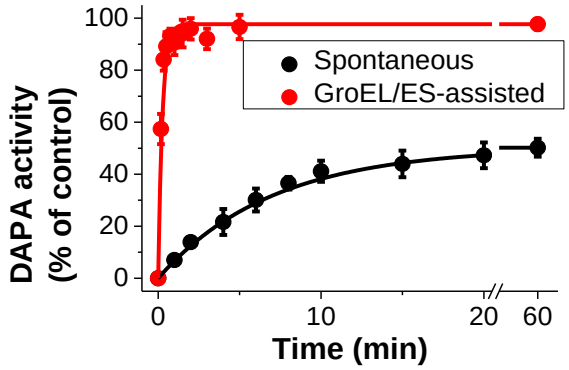
Taken together, results in this chapter argue strongly against the notion of transient aggregation as a rate-limiting step in the DAPA refolding reaction and support of the hypothesis that GroEL/ES is a protein folding "catalyst".

4.1.5 Acceleration of DAPA refolding by GroEL/ES is independent of buffer conditions

However, another recently published study (Tyagi et al., 2011) suggested that the acceleration of protein folding by GroEL/ES can only be observed in certain buffer conditions and strongly depends on the presence of Cl^- ions in the solution. The high salt refolding buffer B used in our experiments contained 100mM KCl and 10mM $MgCl_2$ which corresponds to physiological concentration of Cl^- in *E.coli* cells (Schultz et al., 1962).To elucidate whether the absence of Cl^- in the buffer would change the refolding rate of DAPA, we performed spontaneous and

chaperonin-assisted refolding reactions in buffer where KCl and $MgCl_2$ were replaced with potassium acetate and magnesium acetate, respectively. The pH value of Tris buffer in this case was adjusted using acetic acid instead of HCl. DAPA was unfolded in 8M Urea instead of the previously used 8M GuHCl.

As shown in Figure 4.8A, the rates of spontaneous and chaperonin-assisted refolding in Cl^--depleted buffer are similar to the values obtained in the high salt refolding buffer B. Therefore, the important acceleration of DAPA refolding observed in presence of GroEL/ES system in the above mentioned conditions is due to an intrinsic property of the chaperonin, independent of concentration of the substrate or composition of the refolding buffer.



	k. min^{-1}	
	Spontaneous	GroEL/ES-assisted
High Salt buffer with chloride	0.19 ± 0.04	6.44 ± 0.67
Chloride-depleted buffer	0.16 ± 0.01	5.97 ± 0.48

Figure 4.8: Acceleration of DAPA subunit folding by GroEL/ES at 25°C is independent of buffer conditions

Refolding was measured at 200 nM final monomer concentration by diluting Urea-denatured DAPA 100-fold into refolding buffer without Cl^- (20 mM Tris-Acetate pH 7.5, 100 mM potassium acetate, 10 mM magnesium acetate, 10 mM Pyruvate) or into buffer without Cl^-, containing 2 µM GroEL and 4 µM GroES. Assisted refolding was initiated by addition of ATP. Spontaneous refolding was stopped by addition of 0.8 µM GroEL Trap (GroEL D87K) and GroEL/ES-assisted refolding with 50 mM CDTA. Reactions were incubated for 1 h at 25°C to allow complete assembly prior to enzyme assay. Refolding rates in High salt buffer are taken from Figure 4.4 Standard deviations are from at least 3 independent experiments.

4.1.6 GroEL/ES promotes DAPA subunit assembly

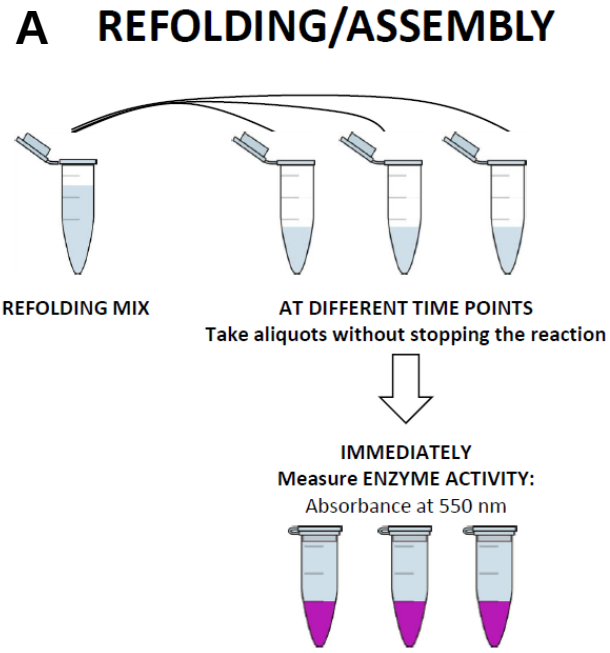
Formation of enzymatically active DAPA requires assembly of individual folded subunits into homotetramers. To find out if subunit assembly is the limiting step on the way to produce the active enzyme, we performed enzyme assays at different time points without stopping refolding reactions (Figure 4.9A). In this case, if subunit folding were the slowest step of DAPA reactivation, the rate obtained in this experiment would be identical to the subunit refolding rate measured earlier. Otherwise, if the limiting step of this process is assembly, we would directly measure the rate of assembly process.

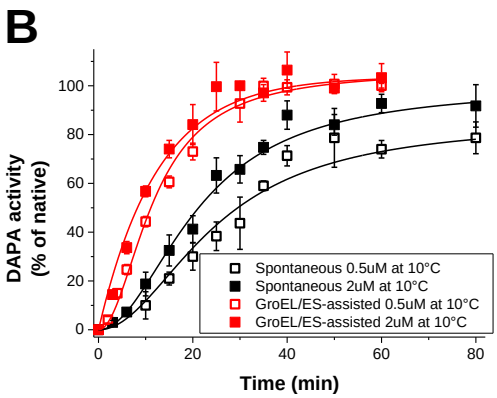
Assuming that the rate of protein assembly is temperature- and concentration-dependent we chose 10°C as the conditions for these experiments to allow for a better resolution. At this temperature, we were able to measure each kinetic point more precisely and could use a higher protein concentration (2µM) to ensure complete assembly of folded monomers.

As shown in Figure 4.9B, the half-time of spontaneous DAPA assembly at 10°C is ~25 min at 0.5µM and ~20 min at 2µM, which is approximately 2-3-fold slower than folding of individual subunits measured in these conditions. Therefore, the assembly of monomers is indeed the limiting step during formation of enzymatically active DAPA. In the presence of GroEL/ES, the half-time of DAPA assembly is ~12 min at 0.5µM and ~8 min at 2µM, which corresponds to a ~2.5-fold acceleration of the assembly process by the chaperonin. This effect is likely to be explained by a higher concentration of assembly-competent, folded subunits in the solution in presence of GroEL/ES at the early time points, as all DAPA monomers are folded by the chaperonin within 2-3 min (Figure 4.4, C).

Therefore, GroEL/ES seems to act as a highly efficient catalyst of DAPA subunit folding which results in a high concentration of folded DAPA subunits, enhancing assembly of DAPA

tetramers. Considering that most of the obligate substrates of GroEL are homo-oligomers, this effect could have biological relevance as subunit assembly is possibly a rate-limiting step of protein biogenesis in the *E.coli* cell.





Assembly rates		
	$t_{1/2}$, min	
	0.5 µM	**2 µM**
Spontaneous	~25	~20
GroEL/ES-assisted	~12	~8

Figure 4.9: DAPA subunit assembly is promoted by GroEL/ES at 10°C.

(A) Scheme of the experiment measuring DAPA subunit assembly.

(B) Refolding/assembly reactions were performed at 10°C with 0.5 µM (empty symbols) or 2 µM DAPA (filled symbols) and 4 µM GroEL/8 µM GroES when indicated. The reactions were not stopped and enzyme activities were measured at the time points indicated. Standard deviations are from at least 3 independent experiments.

4.1.7 A single round of encapsulation inside the chaperonin is sufficient for accelerated substrate folding

The model of iterative annealing suggests that repeated events of substrate unfolding in successive binding/release cycles may reverse kinetically trapped states populated by substrate protein (Lin et al., 2008). However, a single round of substrate binding and encapsulation, using a single-ring mutant of GroEL that binds stably GroES, was shown to be equally efficient in substrate folding (Tang et al., 2006, Chakraborty et al., 2010).

To test whether accelerated DAPA folding may occur in a single round of encapsulation in the chaperonin cage, we used the single-ring mutant of GroEL (SREL, Figure 4.10 A) which carries mutations Arg452, Glu461, Ser463 and Val464 responsible for the major contacts between the two rings of GroEL (Weissman et al., 1996). Residue 452 is replaced by Glu and the other three residues by Ala. These mutations cause the loss of electrostatic interaction between the two rings and result in the formation of a single heptametrical ring. The system allows only one round of ATP hydrolysis upon GroES binding, resulting in a stable SREL/ES complex.

As described in Tang et al., 2006, stable encapsulation of DM-MBP inside the SREL cage can be achieved in low salt buffer. In order to confirm that similar conditions are applicable to study DAPA, we performed the following experiments in the low salt buffer (20 mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$). To avoid increasing salt concentration in the refolding reaction, we used 8M Urea protein denaturant (final concentration upon dilution 80 mM) instead of GuHCl. In order to prove the absence of GroES cycling in these conditions, we measured ATPase activity of GroEL and SREL in low salt buffer and inhibition of the chaperonin ATPase rate by GroES. The ATPase activity of GroEL was inhibited to about 40% in presence of GroES, compared to the value obtained with the chaperonin alone (Figure 4.10B). The ATPase activity

of SREL in the absence of GroES was comparable to the activity of a double-ring chaperonin whereas the addition of GroES into refolding buffer fully inhibits the ATPase activity of SREL. These results indicate the absence of cycling of GroES on SREL in the chosen conditions.

To confirm that DAPA is stably encapsulated inside the SREL cage upon addition of ATP and GroES, we performed size exclusion chromatography to separate and analyse different protein complexes present in the refolding sample. As shown in Figure 4.10C, after the chaperonin cage was sealed with GroES upon addition of ATP, all substrate co-eluted with the SREL/GroES complex indicating that all DAPA molecules had been stably encapsulated inside the chaperonin cage. Thus, the SREL/ES system is suitable for studying a single round of DAPA encapsulation in the cage.

We performed DAPA refolding inside SREL/ES in low salt buffer and compared it to spontaneous and GroEL/ES-assisted refolding in these conditions. Inside SREL/ES, DAPA refolded at a rate of ~3.3 min^{-1} at 25°C (t1/2 ~ 13 sec), i.e. ~17-fold faster than spontaneously (Figure 4.11A). Importantly, a similar rate of refolding was achieved with GroEL WT, indicating that a slower refolding rate inside the chaperonin cage in this buffer as compared to high salt buffer is likely due to a reduced concentration of potassium ions, which are known to enhance the chaperonin affinity for ATP (Viitanen et al., 1990).Note that the yield of spontaneous DAPA refolding in low salt buffer is significantly decreased due to a more pronounced aggregation. However, the spontaneous refolding rate in these conditions remains unchanged, pointing out again that aggregation does not slow down DAPA refolding. Thus, a single round of encapsulation in the SREL/ES cage is sufficient to achieve a significant acceleration of DAPA refolding, arguing against the significance of iterative annealing in substrate folding.

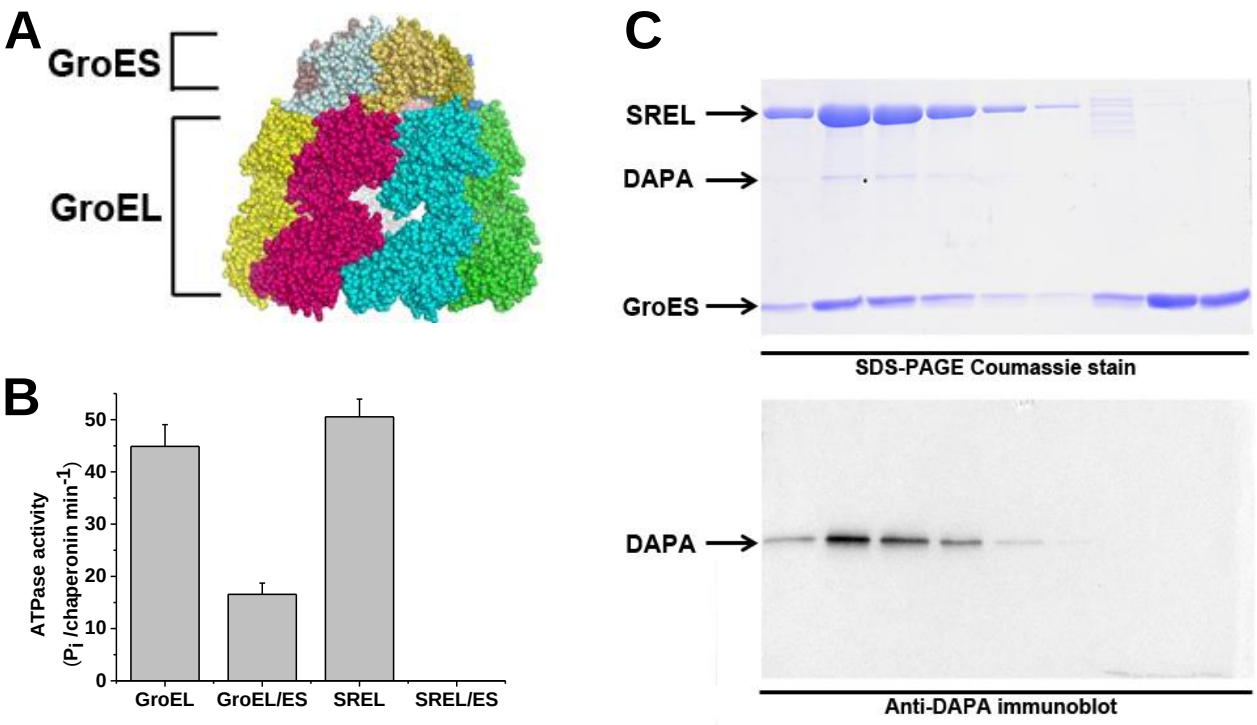


Figure 4.10: Stable encapsulation of DAPA inside SREL/ES complex is achieved at 25°C

(A) Structure of a SREL, a single ring version of GroEL (PDB 1AON, Pymol).

(B) ATPase rate for GroEL, SREL in the absence and presence of GroES was measured as previously described (Tang, 2008) Standard deviations from three different measurements are shown. Note that SREL in the presence of GroES performs only a single round ATP hydrolysis.

(C) The refolding of DAPA was performed in low salt buffer containing 20mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$. 5 min after initiating folding, an aliquot of the reaction was applied on a Superdex200 PC 3.2/10 column (GE Healthcare Life Sciences) using Ettan LC system (Amersham Biosciences) at 25°C. Fractions of 50 µL were collected at a flow rate 60 µL min^{-1} and analysed by SDS-PAGE followed by Coomassie staining and immunoblotting with polyclonal rabbit antibodies against DAPA.

To find out whether efficient folding of the substrate can be achieved within one cycle inside the double ring GroEL, we measured the yield of GroEL/ES-assisted DAPA refolding in the presence of excess of GroEL D87K Trap. When added simultaneously with ATP, GroEL D87K mutant would capture not yet folded DAPA molecules emerging from GroEL WT cage, preventing them from proceeding for another cycle inside GroEL WT. The yield of refolding, therefore, will reflect the amount of substrate that folds inside GroEL/ES cage within a single round.

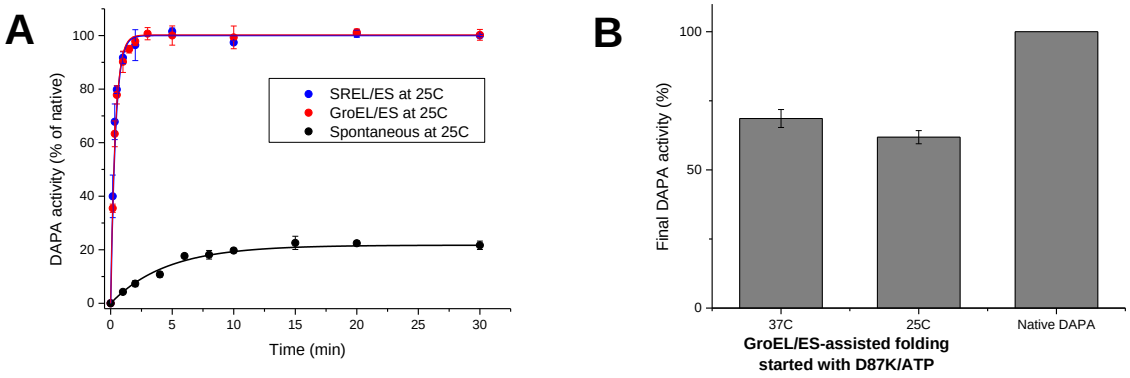


Figure 4.11. A single round of encapsulation in the chaperonin cage is sufficient for DAPA refolding

(A) Spontaneous, SREL/ES-assisted and GroEL/ES-assisted DAPA subunit folding was measured at 25^{o}C at 200 nM final monomer concentration by diluting Urea-denatured DAPA 100-fold into low salt buffer (20 mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$, 10 mM Pyruvate), without chaperones or into the low salt buffer containing 2 µM chaperonin and 4 µM GroES. Assisted refolding was initiated by addition of ATP. Spontaneous refolding was stopped by addition of 0.8 µM GroEL Trap (GroEL D87K) and chaperonin-assisted refolding with 50 mM CDTA and 60 mM GuHCl. Reactions were incubated for 1 h at 25^{o}C to allow for complete assembly prior to enzyme assay. Standard deviations are from at least 3 independent experiments.(B) GroEL/ES-assisted refolding was performed at 25^{o}C as in Figure 4.4 and initiated by addition of 5 mM ATP with 20µM D87K Trap (10-fold access over GroEL) to capture not yet folded molecules of DAPA emerging from GroEL/ES.

Approximately 60% of all DAPA was folded in these conditions at 25°C and ~75% at 37°C (Figure 4.11B), demonstrating that the majority of substrate molecules require only one cycle of encapsulation inside GroEL/ES to reach the native state. In line with the results achieved using SREL/ES, this shows that iterative annealing does not play a significant role in DAPA refolding..

In summary, the results in this chapter argue strongly against the notion of GroEL/ES being a solely passive aggregation prevention device as well as against the theory of iterative annealing. The chaperonin cage is likely to play an active role in protein folding and the mechanism of its action is the focus of the following chapters of this work.

4.2 A highly dynamic intermediate is observed in spontaneous DAPA refolding

In order to elucidate the mechanism of GroEL/ES action, we need to find out what makes a certain subset of proteins chaperonin-dependent. Previous studies attempted to identify the unique features distinguishing GroEL/ES substrates from other proteins but no specific sequence patterns or structural properties could be discovered. Computational studies taking into account the values of protein chain hydrophobicity and net charge, showed that obligate chaperonin substrates are characterized by lower folding propensities compared to GroEL-independent proteins (Tartaglia et al., 2010, Tartaglia et al., 2007). The fact that the obligate GroEL substrates are characterized by an increased aggregation propensity and lower overall hydrophobicity leads to a hypothesis that initial hydrophobic collapse of these proteins may be inefficient (Azia et al., 2011).

In order to gain more insight into the features that make DAPA chaperonin-dependent, we investigated the properties of the intermediates populated along DAPA folding pathway.

4.2.1 Intermediate populated during spontaneous DAPA refolding lacks stable secondary structure

To characterize the intermediate species along the DAPA refolding pathway, we performed isothermal equilibrium GuHCl denaturation experiments after 16 h of incubation in varying concentrations of the denaturant. The amount of secondary structure formed in each sample was estimated by circular dichroism (CD).

DAPA was thermodynamically stable in up to ~1.0-1.5M GuHCl and was fully unfolded only in 4.5-5.0M GuHCl (Figure 4.12). To analyse the secondary structure content in the folding intermediate, we also performed CD measurements immediately upon dilution of the unfolded DAPA into refolding buffer. Final concentration of GuHCl in the samples after dilution was between 100 mM to 0.5M (filled circles in Figure 4.12). The data provides evidence that an intermediate is being formed at an early time point of the refolding reaction. This intermediate presents only ~20% secondary structure compared to the native state and is likely to be the molten globule-like species recognized by the chaperonin.

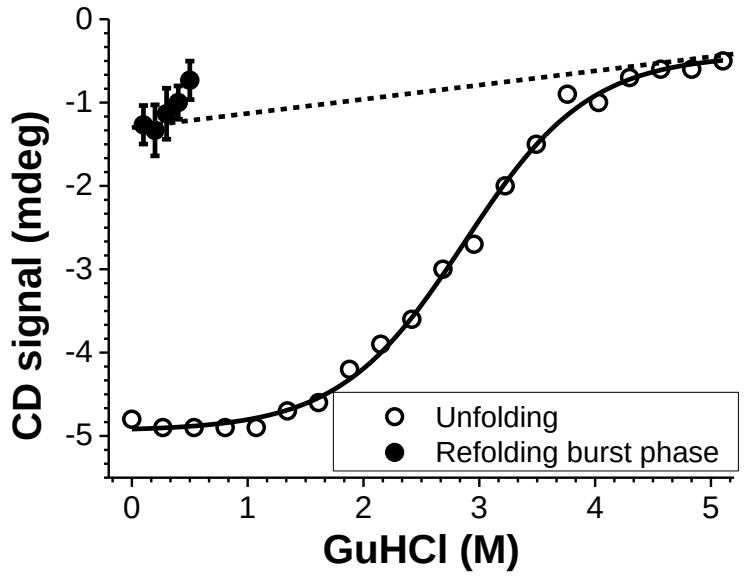


Figure 4.12: Secondary structure in DAPA at different concentration of GuHCl as measured by CD.

Native DAPA (1 µM) was incubated for 16 h in refolding buffer B containing ~70 mM to ~5.7 M GuHCl at 10°C. The steady state CD signal was measured at 220 nm (open circles). CD signals were also measured immediately upon 67-fold dilution of denatured protein into buffer B containing GuHCl to give final concentrations of 100 mM to 0.5 M GuHCl (filled circles).

To further characterize conformational states populated during spontaneous folding of DAPA, we employed pulse-labelling H/DX coupled to liquid chromatography (LC) and MS (Zhang and Smith, 1993). After various times of refolding, the protein was pulse-labeled for 12 s by 10-fold dilution into D_2O buffer (Figure 4.13A). To allow for time resolution of folding and exclude protein aggregation, these experiments were performed at 10°C. Unfolded and native DAPA were used as references. The mass-spectrum of the unfolded DAPA is shown in red in Figure 4.13 B. Peaks corresponding to the unfolding protein species in the spectra are indicated by a red dashed line. Unfolded DAPA had a mass of 31475 Da and incorporated 205 ± 2 deuterons,

which represented ~74 % of total possible deuterium incorporation. Mass-spectrum of the native protein is shown in black and the corresponding peak of the native tetramer is marked with a black dashed line. Native tetramer only incorporated 37 ± 2 deuterons, which corresponded to a mass of 31307 Da.

During the first 70 s of refolding, DAPA populated a distribution of molecules with an average mass essentially identical to that of the unfolded protein. This is consistent with the above mentioned results obtained by CD spectroscopy (Figure 4.12). The wide high-mass peak thus reflects a population of conformationally dynamic molecules lacking stable secondary and tertiary interactions. With time, this highly dynamic species slowly converts into another species with a mass of ~31350 Da (~80 exchangeable hydrogens being deuterated) that most likely represents the folded monomer and disappears with slower kinetics, giving rise to a peak of the native tetramer at ~31307 Da.

For each spectra obtained in the H/DX experiment, we calculated the area under the peak of unfolded DAPA and were therefore able to estimate the kinetics of its disappearance (Figure 4.14A). This is plotted in Figure 4.14B against the kinetics of spontaneous DAPA subunit folding. The peak of unfolded protein converts into the peak of the native monomer with a rate of 0.08 min^{-1}, similar to the rate of subunit folding (0.09 min^{-1}). Since folding of DAPA subunits at each time point was stopped with GroEL, the results indicate that the highly dynamic intermediate is the species recognized by the chaperonin. Of note is also that the peak of the folded DAPA monomer in Figure 4.13B, converts into the native tetramer peak with a slower kinetics reflecting DAPA assembly measured earlier by enzyme activity.

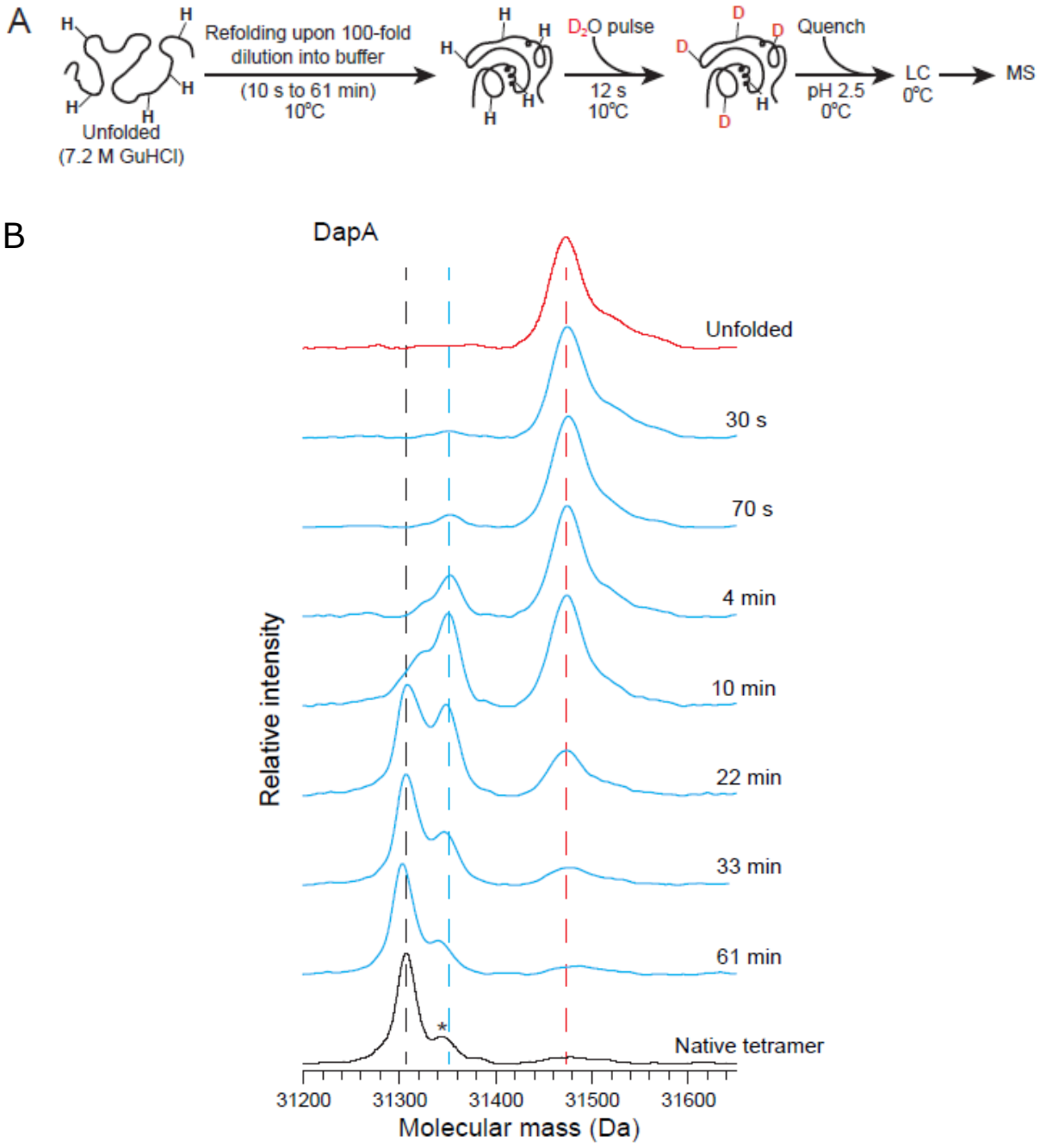


Figure 4.13: Spontaneous refolding of DAPA followed by pulse-labelling H/DX coupled to LC and MS.

(A) Scheme of a typical H/DX pulse experiment.

(B) Positions of unfolded proteins and folded tetramers in the mass spectra are indicated by red and black dotted lines, respectively. The blue dotted line marks the position of folded monomer.

The peak with an asterisk at 31346 Da (10 % of main peak) seen in the native tetramer is due to a potassium adduct. Pulse-labelling H/DX coupled to LC and MS experiments have been performed by Florian Georgescauld.

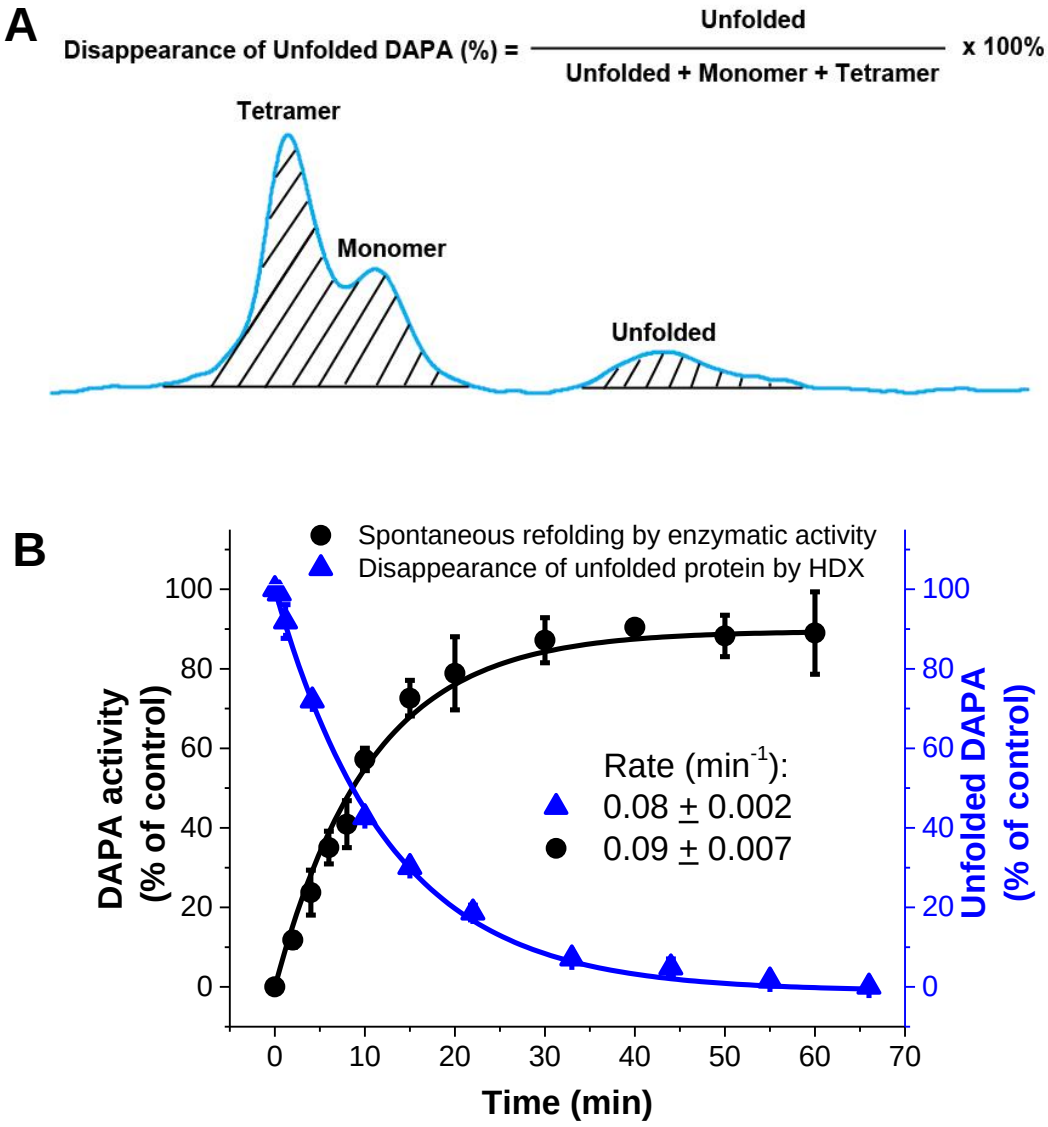


Figure 4.14: Spontaneous DAPA refolding rate followed by enzymatic activity (Figure 4.4) corresponds to the rate of disappearance of the unfolded DAPA peak detected by H/DX.

(A) Scheme illustrating the calculations for Figure B.

(B) Refolding reactions were performed under similar experimental conditions at 10°C. The amount of unfolded protein was calculated as a fraction of the total area of the three peaks of DAPA observed in mass spectra from H/DX experiments. Standard deviations from three different measurements are shown.

Taken together, these results indicate that folding of DAPA subunits follows a two-state behaviour with only unstructured intermediate and folded subunits being populated along the folding pathway, suggesting that despite the complexity of a TIM-barrel fold, the folding pathway of DAPA protein accords with a simple two-state kinetic model..

4.2.2 Chaperonin reduces the entropic component of the energy barrier to folding

Based on the apparent two-state behaviour of the folding reaction, in which the kinetically trapped intermediate and the native state are the main populated species (Figs. 4.13B and 4.14), we used the temperature dependence of the folding rate to extract an approximate estimate of the enthalpic and entropic contributions to the folding energy barrier. We performed spontaneous and GroEL/ES-assisted refolding of DAPA over a range of temperatures from 7°C to 25°C and presented the results in a form of Arrhenius plot (Figure 4.15).

The population of highly dynamic folding intermediates by DAPA suggested the presence of a large entropic component of the energy barrier to folding. Consistent with this notion, the spontaneous folding rate of DAPA proved temperature independent between 15°C and 25°C (Bicout and Szabo, 2000). At lower temperatures, the Arrhenius plot presented a constant slope, indicating the presence of a transition state with both enthalpic and entropic components (Table 4.1). In contrast, the Arrhenius plot of GroEL/ES-assisted folding displayed a linear slope over the whole temperature range, indicating that the activation barrier has gained a significant enthalpic component and the entropic contribution was reduced.

The values of energy barrier to folding as well as its enthalpic and entropic components can be calculated from the Arrhenius plot and are represented in Table 4.1.

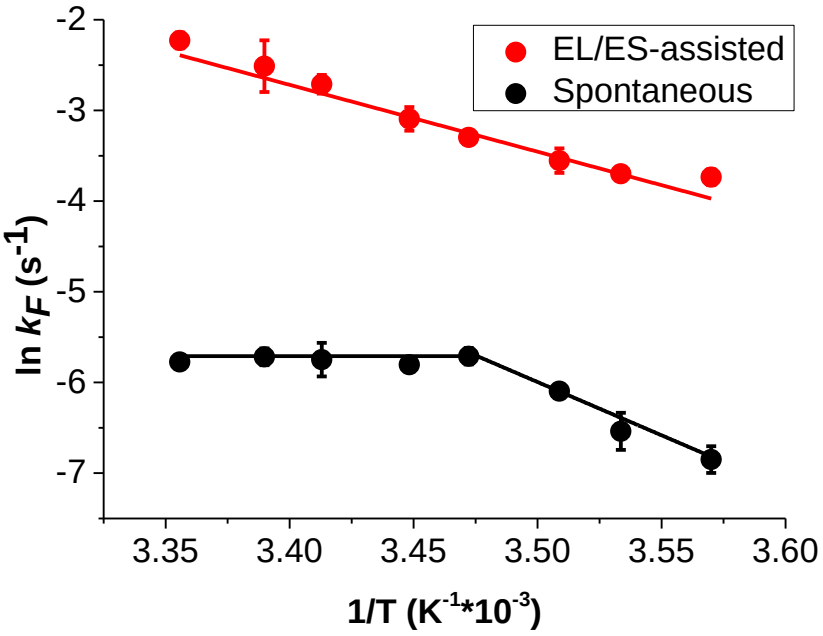


Figure 4.15: Arrhenius plots of spontaneous and GroEL/ES-assisted refolding of 200 nM DAPA in buffer B.

The rates (k_F) were measured as described in the legend of Figure 4.4 at temperatures of 7.0°C to 25°C. Standard deviations from three independent measurements are shown.

According to transition state theory, the rate constant of folding, k_F, of a two-state reaction is defined by the Eyring (or Arrhenius) equation (1):

$$k_F = \frac{k_B T}{h} e^{-\frac{\Delta G_F^{Ea}}{RT}}, \qquad (1)$$

where ΔG_F^{Ea} is the activation energy of folding, k_B is Boltzmann's constant (1.381 x10^{-23} JK^{-1}), and h is Planck's constant (6.626×10^{-34} Js), R is the gas constant (8.31 $JK^{-1}mol^{-1}$) and T is temperature in K.

Using the definition of the free energy:

$$\Delta G_F^{Ea} = \Delta H_F - T\Delta S_F, \qquad (2)$$

the Eyring equation can be rewritten as

$$\ln(k_F) = \left(ln\left(\frac{k_B T}{h}\right) + \frac{\Delta S_F}{R}\right) - \frac{\Delta H_F}{R}\left(\frac{1}{T}\right). \qquad (3)$$

As ln(T) depends weakly on 1/T in the range measured, equation (3) is approximately linear with the slope equal to $-\frac{\Delta H_F}{R}$ and an intercept that depends on ΔS_F. Hence, the enthalpy and entropy contribution to the barrier between the intermediate state (I) and the transition state (TS) can be extracted from the graph. The results are provided in Table 4.1.

Table 4.1: Comparison of folding rate constants (k_F), activation energy of folding (ΔG_F^{Ea}) and the entropy barrier ($T\Delta S_F$) for spontaneous and GroEL/ES-assisted refolding of DAPA WT at 25°C.

	k_F*10^{-3} (s^{-1})*	ΔG_F^{Ea} (kJ mol^{-1})	ΔH_F (kJ mol^{-1})	$T\Delta S_F$ (kJ mol^{-1})
DAPA, spontaneous refolding at 25°C	3.0 ± 0.2	87.2	~0	-87.2
DAPA, GroEL/ES-assisted refolding at 25°C	106 ± 10	78.5	66.6	-11.9
DAPA, spontaneous refolding at 10°C	1.5 ± 0.3	89.3	96.8	7.5
DAPA, GroEL/ES-assisted refolding at 10°C	25 ± 3	77.8	66.6	-11.2

* The apparent refolding rate in high salt refolding buffer B (3 independent experiments).

According to the calculations, in chaperonin-assisted reaction, the entropic component of the energy barrier is significantly reduced and the energy barrier is slightly lower with GroEL/ES compared to spontaneous refolding. This is consistent with the confinement effect of the GroEL cage that has been previously described in Chakraborty et al., 2010.

Thus, the most likely reason for slow spontaneous refolding of DAPA is the presence of a highly dynamic, unstructured intermediate which results in a high entropic factor in the energy barrier

to the native state. Overcoming this barrier presents a major challenge on the substrate folding pathway and one of the ways it can be accomplished is through confinement inside the chaperonin. A different way of avoiding such a kinetic trap has been developed by a GroEL-independent homolog of DAPA which, along with the mechanism of GroEL/ES-assisted refolding, is the focus of the following chapter.

4.2.3 Analysis of DAPA spontaneous and chaperonin-assisted folding at peptide resolution

To obtain futher details and look closely at the common as well as the specific features of DAPA spontaneous and chaperonin-assisted folding, we needed a method that would allow us to dissect the folding pathway of these proteins into smaller steps. H/DX analysis of full-length DAPA upon refolding in the presence of GroEL/ES was not feasible because the signals for GroEL/ES and DAPA extensively overlapped. However, detailed structural information on the spontaneous and assisted folding of DAPA could be obtained by monitoring H/DX protection at peptide resolution.

This study presents the first example of how a method of HD/X couple to mass spectrometry can be applied to study the full refolding pathway of a protein as large as almost 300 amino acids. Attempts have been made previously to employ similar methodology to various proteins, including lysozyme and ribonuclease H (Miranker et al., 1991, Chen et al., 2001, Hu et al., 2013).

We compared spontaneous and GroEL/ES-mediated refolding of DAPA at 10^{o}C, as the low temperature allows us to obtain improved time resolution and to exclude aggregation. In addition, we analysed the refolding of DAPA upon stable encapsulation in SREL/ES, and

compared it to the cycling GroEL/ES reaction. The experiments with SREL/ES were performed at 25^{o}C because the SREL/ES complex is unstable at low temperature.

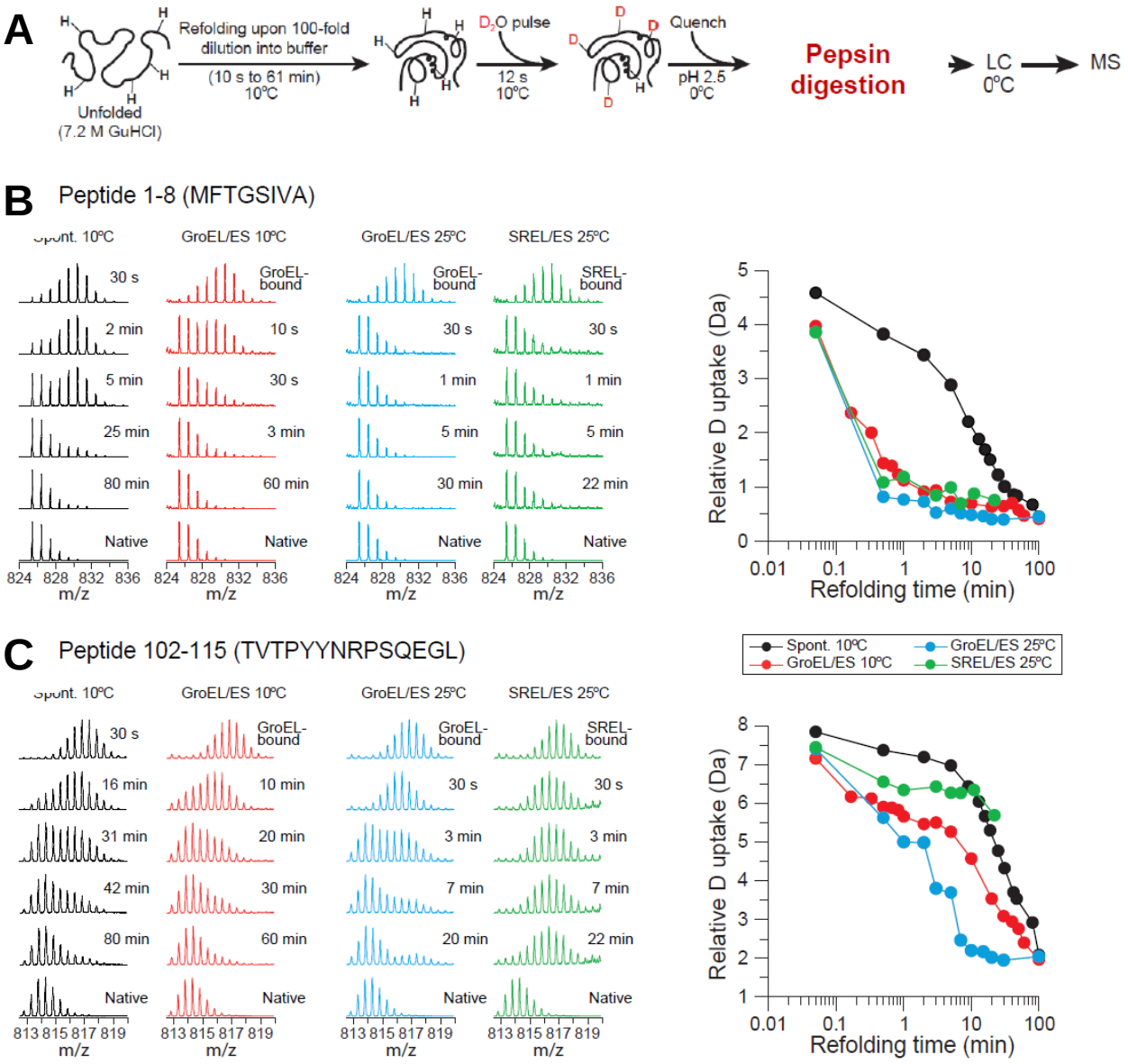


Figure 4.16 Deuterium incorporation into peptides of DAPA during spontaneous and assisted refolding

(A) Schematic representation of the H/DX pulse experiment at peptide resolution. Conditions of this experiment are similar to those presented in Figure 4.13, with the addition of a pepsin digestion step prior to liquid chromatography.

Deuterium incorporation into peptide P1-8 of the TIM-barrel domain of DAPA (B) and into peptide P102-115 of interface I of DAPA (C).

Left panels: Examples of mass spectra for DAPA peptides P1-8 and P102-115 at different times during spontaneous, GroEL/ES-assisted and SREL/ES-assisted refolding at 10°C and 25°C, as indicated. Amino acid sequences of the peptides are indicated in single letter code. Right panels: Deuterium uptake in Da is plotted versus refolding time.

The isotope distributions for nearly all peptides during refolding (spontaneous or assisted) were bimodal, indicating that peptides were either unfolded or folded, but not partially folded. This is illustrated for peptides P1-8 (MFTGSIVA) (Figure 4.16B) and P102-115 (TVTPYYNRPSQEGL) (Figure 4.16C): only two states are apparent, displaying either the same amount of deuterium incorporation as in the unfolded state (represented by the 30 s time point of spontaneous refolding) or limited deuterium incorporation as in the native protein. The mass spectra showed time-dependent transitions from an all-exchangeable to an all-protected population. Strikingly, GroEL/ES accelerated the rate at which P1-8 acquired protection by at least 50-fold ($t_{1/2}$ ~12 s, the duration of the D_2O pulse) compared to spontaneous folding ($t_{1/2}$ ~9 min) (Figures 4.16B). P102-115 acquired protection much more slowly ($t_{1/2}$ ~9 min) with GroEL/ES, but still ~2-fold faster than in spontaneous folding ($t_{1/2}$ ~20 min) (Figures 4.16C), consistent with assembly being enhanced due to accelerated subunit folding (Figure 4.9). Acquisition of protection in distinct peptides generally correlated either with the rate of subunit folding or assembly measured by enzymatic assay, arguing against non-specific effects due to interaction of DAPA with the wall of the GroEL/ES cage.

To distinguish between structure formation upon subunit folding versus assembly, we performed H/DX measurements during stable encapsulation of DAPA in SREL/ES at 25°C (Figure 4.16, B and C, in green). For comparison, folding with GroEL/ES was also analysed at 25°C (Figure 4.16, B and C, in blue). In both systems, P1-8 acquired protection at essentially the same rate,

demonstrating that a single round of protein encapsulation is sufficient to catalyse folding (Figure 4.16B). In contrast, P102-115 of interface I acquired only ~20 % protection in SREL/ES (Figures 4.16C), demonstrating that full protection of this peptide results from subunit assembly after protein release from the chaperonin cage. Similar behaviour was observed for other peptides involved in assembly (see Figure 4.17), further excluding non-specific effects of the cage wall on H/DX protection. Additionally, it proves that once encapsulated, the substrate remains inside the SREL/ES cage during the time of the experiment. Therefore, the conditions we used in this study (low salt buffer and Urea as protein denaturant) are suitable for further studies of substrate folding in non-cycling conditions with SREL/ES system.

In a similar manner, we analysed all the other peptides of DAPA and revealed different categories of protection (Figure 4.17). During spontaneous refolding at 10°C, the peptides in the fastest category were all located in the TIM-barrel domain and acquired protection with a t½ of ~9.5 min, consistent with the rate of subunit folding (Figures 4.17A, Figure 4.4). Three TIM-barrel peptides (P9-24; P63-85; P160-167) acquired protection more slowly. P160-167, containing the conserved K161 involved in binding the substrate pyruvate, has the slowest protection (t½ ~22 min). This is similar to the rate at which peptides of the C-terminal domain and regions at the tight dimer interface (interface I) acquire protection (Figures 4.17A), and corresponds to the rate of tetramer assembly and acquisition of enzymatic activity (Figure 4.9). P63-85 has three of its residues located in interface I, possibly contributing to its slower protection. During GroEL/ES-assisted folding, all TIM-barrel peptides, except P160-167, acquired protection with a t½ of 30 s or less, while several of the C-terminal domain and subunit interface peptides reached protection with a t½ of 6-11 min (Figure 4.17B), corresponding to the rate of assembly in the presence of GroEL/ES (Figure 4.9).

Upon encapsulation in the non-cycling SREL/ES complex, all TIM-barrel peptides (except P160-167) reached protection at similar rates as with the cycling GroEL/ES, while regions involved in assembly showed markedly reduced protection (Figure 4.17B and the panel below). These results demonstrate that a single round of encapsulation in the chaperonin cage is sufficient to achieve full refolding of the substrate, strongly arguing, therefore, against the iterative annealing model.

Our analysis also suggests that during spontaneous folding, the structure of the TIM-barrel evolves in a highly concerted process, with almost all its segments not involved in assembly acquiring H/DX protection simultaneously (t½ ~9.5 min) (Figure 4.17A,). During folding with GroEL/ES, these peptides acquire protection 20- to 50-fold faster (Figure 4.17B insert). Moreover, protection no longer develops simultaneously for all TIM-barrel peptides. Specifically, P62-66, P116-124 and P144-150 mapping to α-helices H2, H4 and H6, respectively, acquire protection 2- to 3- fold faster than peptides with mixed secondary elements: P9-24 (coil and α-helix), P63-85 (αβ), P86-101 (αβ) and P125-132 (coil and β-strand) (Figure 4.17B, insert). These differences in protection suggest that confinement by the chaperonin catalyses folding by promoting local structure formation in amphiphilic α-helices onto which β-strands can dock (Figure 4.17B, insert). This is consistent with GroEL/ES reducing the entropic component of the folding energy barrier, shown previously in Table 4.1.

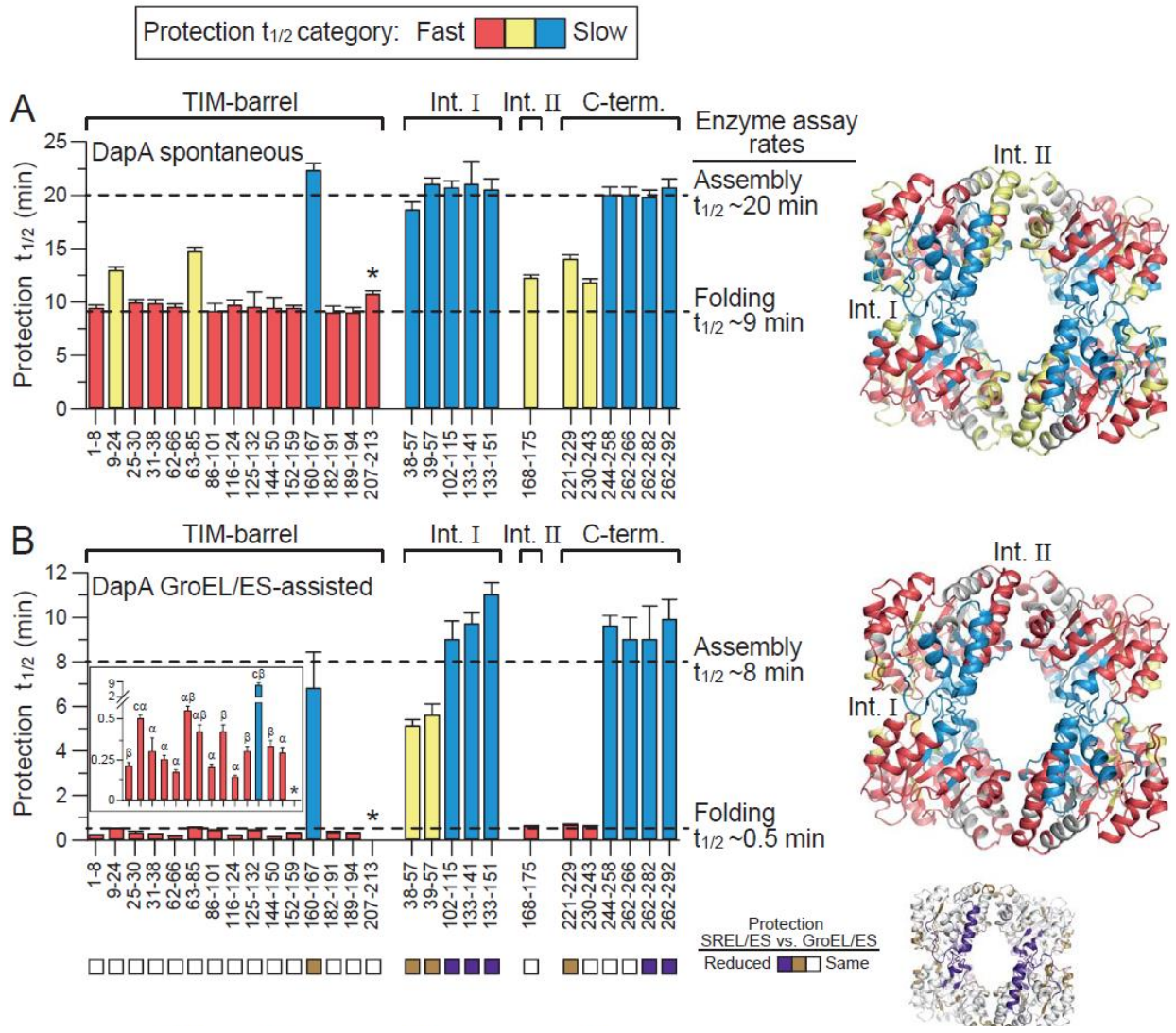


Figure 4.17: Comparison of DAPA spontaneous and GroEL/ES-assisted refolding at peptide resolution

Apparent half-times of H/DX protection for peptides along the amino acid sequence during spontaneous DAPA refolding/assembly (A) and GroEL/ES-assisted DAPA refolding/assembly (B) at 10^{o}C are presented in the bar graphs and are mapped on the tetramer structures of DAPA (right panels). Peptides are assigned either to the TIM-barrel domain, C-terminal domain or interfaces I and II (when at least 25 % of the sequence is in the interface). Half-times of protection are colour coded: Red bars indicate peptides with half-times as fast or faster as subunit refolding in enzymatic refolding assays (Figure 4.4); blue bars denote peptide with half-times of protection as slow as assembly (Figure 4.9), and yellow bar peptides with intermediate half-times of protection. Note that peptide P207-213, indicated by an asterisk, in GroEL/ES-

assisted folding is already fully protected in the GroEL-bound state. The insert highlights the differences in protection of the peptides of the TIM-barrel domain and the secondary structure of the peptides is indicated: α, α-helix; β, β-strand; cα, coil and α-helix; cβ, coil and β-strand. The squares below the bar graph in (B) indicate protection properties of peptides upon refolding with SREL/ES in comparison to GroEL/ES at 25^{o}C. Peptides with reduced protection with SREL/ES are highlighted in the tetramer structure. Experiments in this section were performed by Florian Georgescauld.

4.3 Homologous TIM-barrel proteins from *E.coli* and *M.synoviae* refold differently

About 45% of all obligate substrates of GroEL (Class III substrates) carry the TIM-barrel fold (Kerner et al., 2005). However, there exist many TIM-barrel proteins that do not require chaperone assistance for their folding. Moreover, some bacterial species such as *Mycoplasma synoviae* do not have a GroEL/ES operon in their genome (Vasconcelos et al., 2005). Therefore, homologs of the obligate chaperonin substrates in this organism must have evolved to fold efficiently without the assistance of GroEL/ES. Uncovering the folding mechanism of such a protein would help to understand what distinguishes Class III substrates from their chaperonin-independent homologues and provide insights into the GroEL/ES mechanism.

4.3.1 Unlike the TIM-barrel protein EcNANA from *E.coli*, its homolog from *M.synoviae* refolds spontaneously

Another Class III GroEL substrate that we chose to use in our study is N-acetylneuraminic acid aldolase from *E.coli*, EcNANA (32.5 kDa) (Fujiwara et al., 2010; Kerner et al., 2005). This protein is a paralog of DAPA and is also comprised of an N-terminal TIM-barrel domain and 3 C-terminal α-helices (Figure 4.18). Similar to DAPA, the functional form of EcNANA protein is a tetramer, with two different types of interacting interfaces – tight interface (I) and weak interface (II), depicted with a dashed line. DAPA and EcNANA share about 26% sequence identity (Figure 4.19A), especially in the TIM-barrel domain area, and have a similar catalytic site with the active lysine at the centre of the TIM-barrel.

The close structural homolog of DAPA and EcNANA in *Mycoplasma synoviae* is MsNANA. MsNANA shares about 36% sequence identity with EcNANA (Figure 4.19B). To our knowledge, this protein has never been expressed or studied *in vitro* up to date. Therefore, we

synthesized the gene using the sequence identified by Vasconcelos et al., and purified the protein in a similar manner to EcNANA.

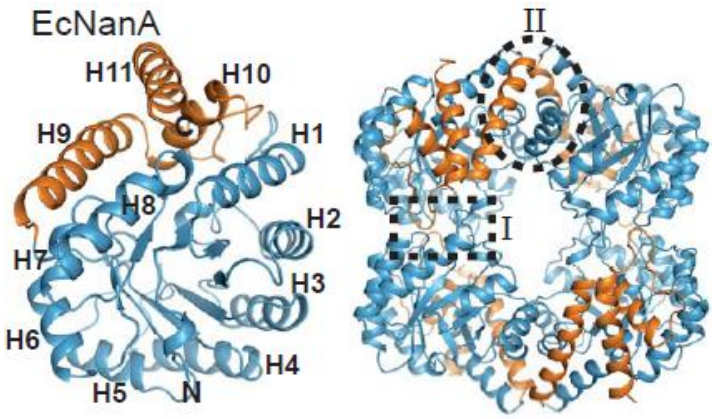


Figure 4.18: Crystal structure of EcNANA (PDB 2WO5).

Left, monomer; right, tetramer. Helices H1 to H11 as well as the locations of the interfaces I and II of the tetramer are indicated.

A

```
                  *         20         *         40         *         60         *         80         *        100        *
DAPA_ECOLI : ---MFTGSIVAIVTPMDEKGNVCRASLKKLIDYHVASGTSAIVSVGTTGESATLNHDEHADVVMMTLDLADGRIPVIAGTGANATAEAISLTQRFNDSGIVGCLTVTPYYN : 108
NANA_ECOLI : MATNLRGVMAALLTPFDQQQALDKASLRRLVQFNIQQGIDGLYVGGSTGEAFVQSLSEREQVIELVAEEAKGKIKLIAHVGCVSTAESQQLAASAKRYGFDAVSAVTPFYY : 111
                    G 6 A66TP D2   6 4ASL44L6 5 6  G   6   G3TGE       E   V6 6    A G4I 6IA  G   TAE   L       G      VTP5Y

                 120        *         140        *         160        *         180        *         200        *         220
DAPA_ECOLI : RPSQEGLYQHFKAIAEHTDLPQILYNVPSRTGCDLLPETVGRLAKVKNIIGIKEATGNLTRVNQIRELVSDDFVLLSGDDASALDFMQLGGHGVISVTANVAARDMAQMC : 218
NANA_ECOLI : PFSREEHCDIYRAIIDSADGLPMVVYNIPALSGVKLTLDQINTLVTLPGVGALKQTSGDLYQMEQIRREHPDLVLYNGYDEIFASGLLAGADGGIGSTYNIMGWRYQGIV : 221
               S E    H54AI    D LP 66YN6P  3G  L    6  L  6  6  6K2 3G1L  6 QI4     D VL  G D      6  G  G I  T N6       6

                    *       240        *         260        *         280        *
DAPA_ECOLI : KLAAEGHFAEARVINQRLMPLHNKLFVEPNPIPVKWACKELGLVATDTLRLPMTPITDSGRETVRAALKHAGLL-- : 292
NANA_ECOLI : RALKEGDIQTAQKLQTECNKVIDLLIKTGVFRGLKTVLHYMDVVSVPLCRKPFGPVDEKYLPELKALAQQLMQERG : 297
              K   EG    A  6     6 1 L        6K     6 6V     R P  P6       64A
```

B

```
                  *         20         *         40         *         60         *         80         *        100        *
NANA_ECOLI : --MATNIRGVMAALLTPFDQQQALDKASLRRLVQFNIQQG-IDGLYVGGSTGEAFVQSLSEREQVIEIVAEEAKGKIKLIAHVGCVSTAESQQLAASAKRYGFDAVSAVTP : 108
NANA_MYSYN : MHNFDKPKGIFPAMVTPFTKDGKLHKAGVKPVVNFLVEKQKVDGIYITGSTGEFLLLSFEDKKEVMKLVAEEANAGRVTLIAGIGSLNIEEMKELAKIAKELKYDAISAITP : 111
                    4G6   A66TPF      L KA 64 6V F 62    6DG6Y6 GSTGE   6 S    4 2V6 6VAE    G46 LIA 6G 6    E3 2LA   AK    5DA6SA6TP

                 120        *         140        *         160        *         180        *         200        *         220
NANA_ECOLI : FYYPFSFEEHCDHYRAIIDSADGLPMVVYNIPALSGVKLTLDQINTLVTLPGVGALKQTSGDLYQMEQIRREHPDIVLYNGYDEIFASGLLAGADGGIGSTYNIMGWRYQG : 219
NANA_MYSYN : YYYNFSENETHHIYEEISKAADIPMLIYMLEQIAGQKVSTDQFGKLIELKNVIGSKYGATDLEAFERLMSKYPDRLFMFAWDEAILAMGLTMGARGFIGSTYNVNAKGANA : 221
              5YY FSF E    Y  I   AD 6PM66Y 6P L G K63 DQ   L6 6  V   K    DL5  E 6    PD 6    5DE  A GL  GA G IGSTYN6

                    *       240        *         260        *         280        *         300
NANA_ECOLI : IVKALKEGDIQTAQKLQTECNKVIDLLIKTGVFRGLKTVLHYMDVVSVPLCRKPFGPVDEKYLPELKALAQQLMQERG : 297
NANA_MYSYN : IIKAWEANDKEAVMKLTHTYNDYVLDLISKGLMQSLKAIMRLHGVLAG-YTRKPFWRYEDEDIKKHAEFITDKYLK-- : 296
              I6KA    D 2   KL    N  6  LI  G6   LK 66    V      RKPF       6
```

Figure 4.19: Pairwise alignments of TIM-barrel proteins DAPA, EcNANA and MsNANA sequences (http://www.genome.jp/tools/clustalw/).

(A) Alignment DAPA and EcNANA amino acid sequences performed using ClustalW algorithm. (B) Alignment of EcNANA and MsNANA amino acid sequences performed using ClustalW algorithm.

Refolding of EcNANA and MsNANA was performed with and without GroEL/ES at different temperatures. In contrast to DAPA, no spontaneous refolding was observed for EcNANA due to pronounced aggregation, even at low temperature on ice and very dilute concentrations like 100 nM. Therefore, all further experiments with EcNANA were performed in presence of the chaperonin in similar conditions to DAPA refolding (in buffer D, at 25°C), and protein concentrations were 400 nM (Figure 4.20A).

In these conditions, EcNANA could refold efficiently with GroEL/ES at a rate of ~1.36 min^{-1} ($t_{½}$ ~30 s). Remarkably, MsNANA renatured spontaneously with a half-time similar to that measured for the assisted refolding of EcNANA which points out that MsNANA is well adapted to fold without the chaperonin.

To find out whether GroEL/ES could further accelerate MsNANA folding, we compared refolding of this protein at 25°C, in presence and in absence of the chaperonin (Figure 4.20B). Refolding rate of MsNANA in both these experiments was essentially identical, which points out that refolding of MsNANA had become fully GroEL-independent.

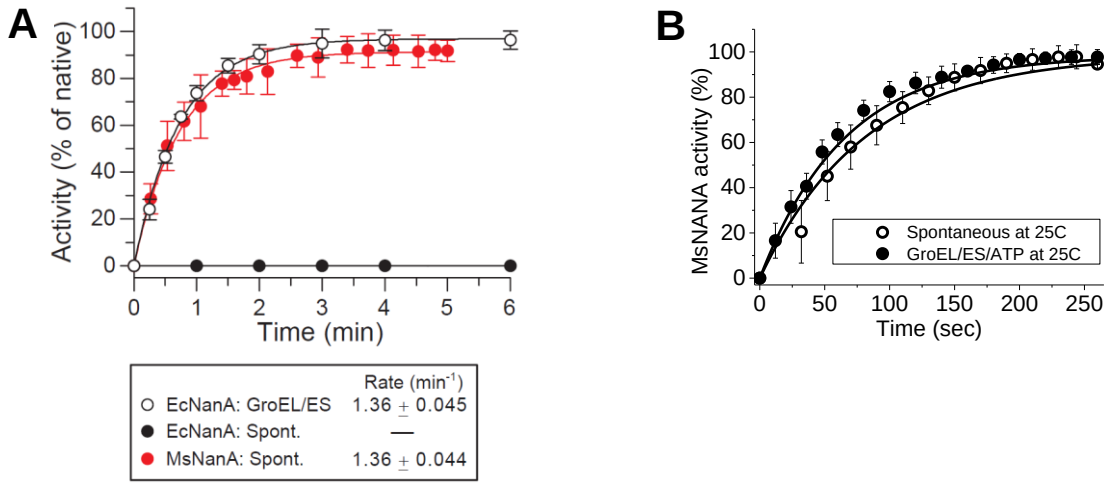


Figure 4.20: Spontaneous and GroEL/ES-assisted refolding of EcNANA and MsNANA occur at a similar rate.

(A) GroEL/ES-assisted refolding of EcNANA and spontaneous refolding of MsNANA occur at similar rates. Spontaneous and GroEL/ES assisted refolding of EcNANA were analysed at 25^{o}C at a final concentration of NANA of 400 nM in buffer D essentially as described for DAPA refolding in Figure 4.4. Spontaneous renaturation of MsNANA (400 nM final concentration) in buffer D was analysed by direct enzyme assay at the time points indicated. The observed kinetics for MsNANA reflect both subunit folding and assembly. Standard deviations are from at least 3 independent experiments.

(B) Refolding of MsNANA with and without GroEL/ES occur at similar rates. Refolding of MsNANA in presence and in absence of the chaperonin were analysed at 25°C at a final concentration of NANA of 400 nM in buffer D as described earlier. The observed kinetics reflect both subunit folding and assembly. Standard deviations are from at least 3 independent experiments.

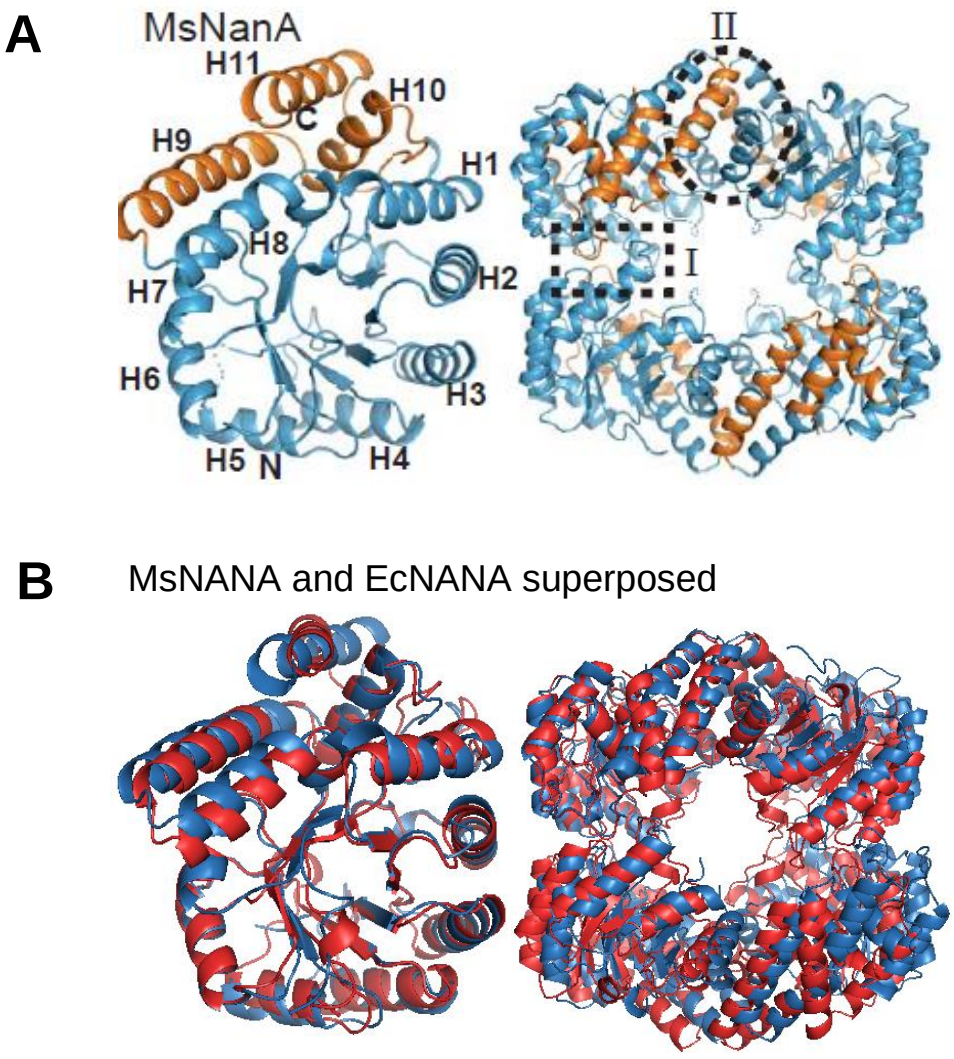


Figure 4.21: Crystal structure of MsNANA (PDB 4N4P, this study) alone (A) or superposed with crystal structure of EcNANA (B). (A) Structure of MsNANA. Left, monomer; right, tetramer. Helices H1 to H11 as well as the locations of the interfaces I and II of the tetramer. (B) Superposition of EcNANA (red) and MsNANA (blue) crystal structures, using PDBeFold service (http://www.ebi.ac.uk/msd-srv/ssm/). Left, monomer; right, tetramer.

Next, we focused on the special structural features that make the MsNANA protein chaperonin-independent. The crystal structure of MsNANA was not known and, in collaboration with Andreas Bracher from the department, we crystallized the protein and solved its structure at 1.8

Å resolution (Figure 4.21A and Table 4.2). Similar to DAPA and EcNANA, this protein is a homo-tetramer, with two different interaction interfaces between subunits shown as dash line in Figure 4.21A. As for its homologs from *E.coli*, each monomer of MsNANA has an N-terminal TIM-barrel domain and 3 C-terminal α-helices. Superposition of structures (Figure 4.21B) shows that EcNANA and MsNANA are highly similar (r.m.s.d. of 1.4 to 1.8Å).

Table 4.2. Data Collection and Refinement Statistics, Related to Figure 4.21

Dataset	**I**	**II**
PBD code	4N4P	4N4Q
Data collection		
Beamline	SLS X10SA	SLS X10SA
Wavelength	0.89997	0.89997
Space group	$P2_12_12_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	59.99, 136.35, 157.25	60.23, 142.44, 80.79
α, β, γ (°)	90, 90, 90	90, 108.27, 90
Resolution (Å) *	48.93 – 1.8 (1.9 – 1.8)	47.48 – 2.0 (2.11 – 2.0)
R_{merge} *	0.176 (0.633)	0.113 (0.499)
I / σ *	11.0 (3.2)	6.7 (2.0)
Completeness (%)*	99.1 (93.9)	99.0 (97.6)
Redundancy *	12.9 (12.2)	3.5 (3.3)
Refinement		
Resolution (Å)	30 – 1.8	30 – 2.0
No. reflections	112371	81494
R_{work} / R_{free}	0.1729 / 0.2068	0.2392 / 0.2830
No. atoms	10378	9824
Protein	9272	9342
Ligand/ion	8	0
Water	1098	482
B-factors (Å^2)	15.57	27.58
Protein	14.43	27.50
Ligand/ion	19.80	-
Water	25.29	28.93
R.m.s. deviations		
Bond lengths (Å)	0.014	0.013
Bond angles (°)	1.301	1.315

Interestingly, MsNANA contains more than twice the number of lysines (34) as the chaperonin-dependent EcNANA and DAPA proteins (16 and 13, respectively), and less than half the number of arginines (5 compared to 12 and 13, respectively). Lysines are generally solvent exposed and are known to confer helical propensity (Groebke et al., 1996), therefore acting as 'gatekeepers' as they exclude wide regions of the conformational space by their desolvation penalties (Oliveberg and Wolynes, 2005). Indeed, the structure of MsNANA shows numerous lysines located on the solvent exposed surfaces of the amphiphilic TIM-barrel α-helices (such as H1, H2, H3) (Figure 4.22). These lysine residues would be predicted to stabilize the α-helices and to facilitate the formation of the hydrophobic core. Moreover, the total number of aromatic residues (Phe and Tyr) in MsNANA is increased to 34 compared to 23 in EcNANA and 12 in DAPA. The aromatic side chains contribute to the hydrophobic core of MsNANA and some engage in edge-face interactions, stabilizing the fold (Burley and Petsko, 1985; Singh and Thornton, 1985). These differences in amino acid composition may explain the efficient folding of MsNANA in the absence of chaperonin.

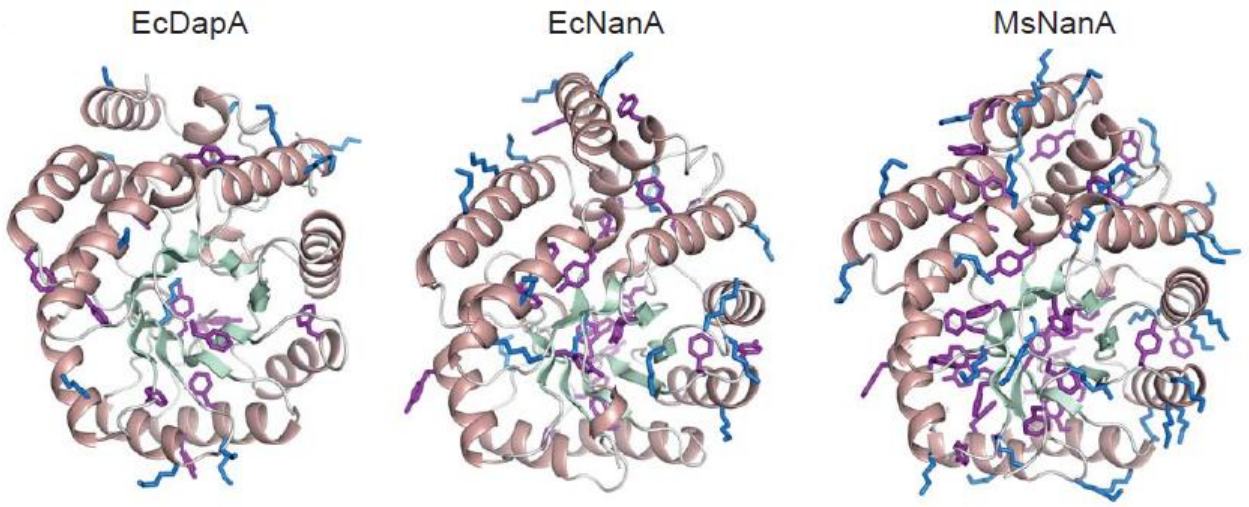


Figure 4.22: Amino acid compositional bias in MsNANA. The monomer structures of EcDAPA, EcNANA and MsNANA are shown in ribbon representations with α-helices and β-strands indicated in salmon and pale green, respectively. The side-chains of lysines and aromatic residues (Phe and Tyr) are highlighted in blue and purple, respectively. Note the relative abundance of lysine residues in α-helices H1-H3 of MsNANA.

To obtain further insights into the nature of the intermediate populated along the MsNANA refolding pathway, we performed spontaneous renaturation experiments over a range of temperatures from 7°C to 25°C and presented the results in a form of Arrhenius plot (Figure 4.21; similar to described previously for DAPA in Figure 4.15). Since enzyme assay of MsNANA does not allow us to distinguish between its refolding and assembly processes, we performed renaturation experiments of this protein over a range of concentrations at 10°C. The renaturation increased with the increase of protein concentration from 100nM to 200nM and remained unchanged between 200 – 1000 nM (Figure 4.23A), indicating that the plateu was reached where assembly is not anymore the limiting step of the renaturation process. Moreover, the MsNANA renaturation rate at 25°C was similar when measured at 400nM and 2uM protein concentration (Figure 4.23B and Figure 4.20A). Finally, to ensure that the rate measured in these

experiments is a true rate of MsNANA subunit folding, we performed the experiments at the high MsNANA concentration of 2 uM at all temperatures.

The Arrhenius plot therefore allows us to extract the thermodynamic parameters of MsNANA monomer folding (Figure 4.23B). The plot displayed a linear slope over the whole temperature range, indicating that the activation barrier contains a significant enthalpic component and the entropic contribution is less important. These results are similar to previously obtained data for GroEL/ES-assisted refolding of DAPA and suggest that GroEL/ES shifts the folding properties of DAPA towards those of the chaperonin-independent protein MsNANA.

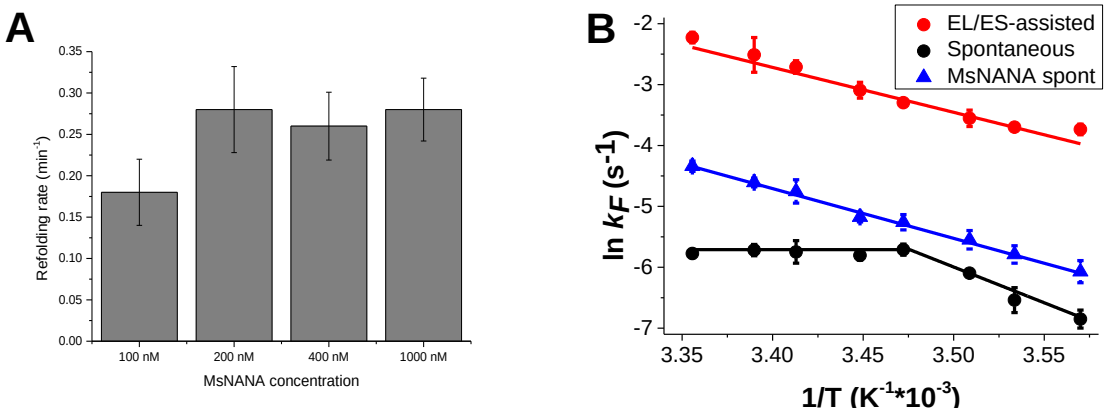


Figure 4.23: Spontaneous subunit folding of MsNANA is characterized by Arrhenius-like behaviour

(A) Rate of MsNANA renaturation over a range of protein concentrations. Spontaneous renaturation of MsNANA was analysed by direct enzyme assay as described in Figure 4.20. Standard deviations are from at least 3 independent experiments.

(B) Arrhenius plots of spontaneous and GroEL/ES-assisted refolding of 200 nM DAPA in buffer B (from Figure 4.15) and spontaneous renaturation of 2 µM MsNANA in buffer D. The rates (k_F) were measured as described earlier, at temperatures of 7.0°C to 25°C. Standard deviations from three independent measurements are shown.

The thermodynamical parameters of MsNANA folding were calculated from the Arrhenius plot as previously and are represented in Table 4.3.

Table 4.3: Comparison of folding rate constants (k_F), activation energy of folding (ΔG_F^{Ea}) and the entropy barrier ($T\Delta S_F$) for spontaneous renaturation of MsNANA at 25°C.

	k_F*10^{-3} (s^{-1})*	ΔG_F^{Ea} (kJ mol^{-1})	ΔH_F (kJ mol^{-1})	$T\Delta S_F$ (kJ mol^{-1})
MsNANA spontaneous 25°C	23 ± 2	82.3	67.8	-14.5
MsNANA spontaneous 10°C	4.9 ± 0.2	81.7	67.8	-13.9

4.3.2 Spontaneous folding of MsNANA followed by HD/X coupled to LC and MS

In order to characterize conformational states populated during MsNANA folding, we employed pulse-labelling H/DX coupled to liquid chromatography (LC) and MS on full-length protein, similar to what has previously been applied for DAPA. The experiments were performed at 10°C, unfolded and native MsNANA were used as references. Mass-spectrum of the unfolded MsNANA is shown in Figure 4.24, in red and the peaks corresponding to the unfolding protein species in other spectra are indicated by a red dashed line.

In contrast to DAPA, MsNANA (208 ± 3 deuterons incorporated in the unfolded state) immediately upon dilution from denaturant populated a broader range of folding intermediates with varying numbers of deuterons incorporated (Figure 4.24). Notably, already after 10 sec of refolding, the intermediate detected was more compact than the unfolded MsNANA species. Later this heterogeneous population eventually converged into a peak around 33640 Da (38 ± 3 exchangeable deuterons), which corresponds to the assembled tetramer. No discernible folded monomers or dimers were detected. Thus, it appears that in case of MsNANA, subunit folding and assembly processes are coupled. This is a striking difference from the refolding pathway of DAPA where a distinct monomeric species is first formed, which then slowly assembles into a tetramer.

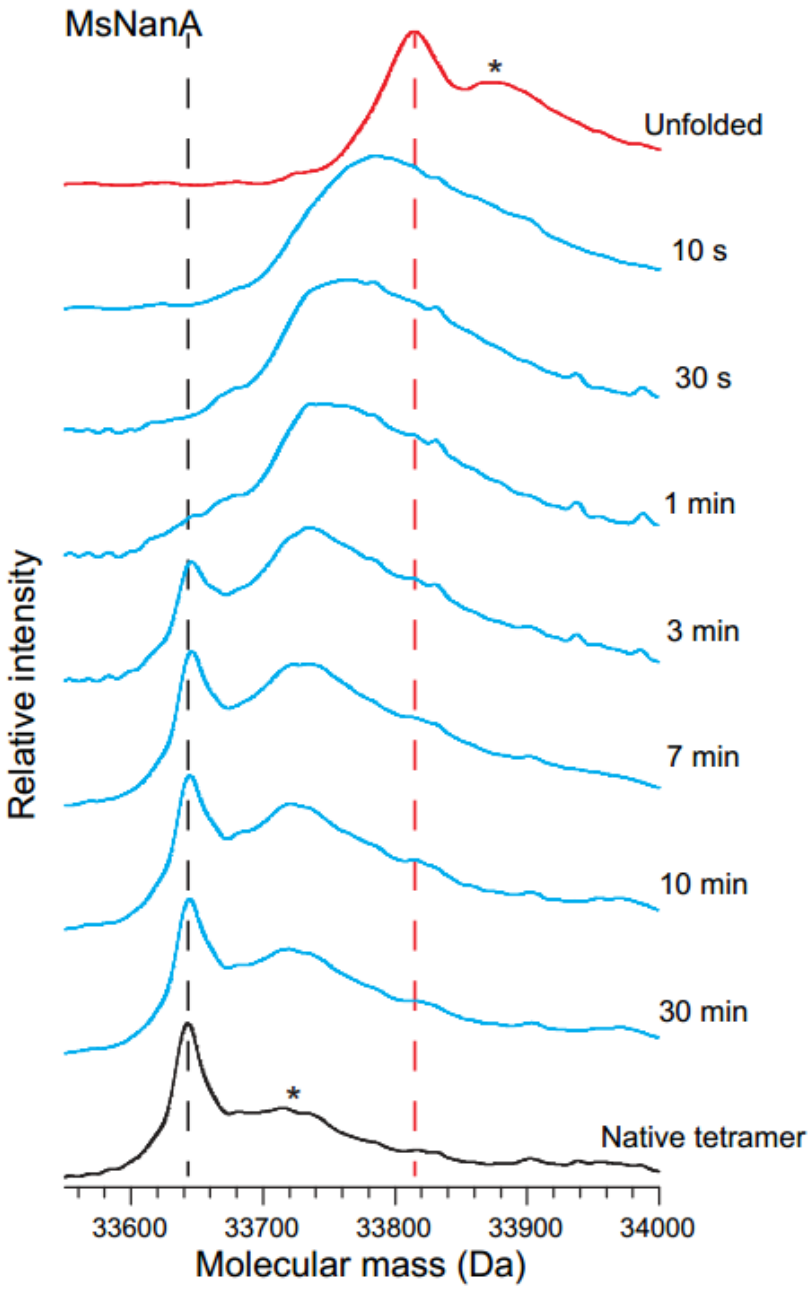


Figure 4.24: Mass spectra during spontaneous renaturation of MsNANA. Positions of unfolded proteins and folded tetramers in the mass spectra are indicated by red and black dotted lines, respectively. Asterisks indicate the presence of potassium and sodium adducts (1 sodium, 1 potassium, 2 sodium and 2 potassium). The experiments were performed by Florian Georgescauld.

4.3.3 Folding mechanism of GroEL/ES-independent MsNANA at peptide resolution

The above mentioned results indicate that refolding of MsNANA and DAPA are significantly different, and, importantly, spontaneous folding of MsNANA seems to resemble the GroEL/ES-assisted folding of DAPA. To obtain insight into the mechanism of MsNANA folding and the features underlying chaperonin-independence of this protein, we also analysed the folding/assembly of MsNANA by H/DX at peptide resolution (Figure 4.25). We detected non-uniform rates of protection for elements within the TIM-barrel domain, with up to 4-fold rate differences ($t_{½}$ ~15 s to ~1.8 min at 10^{o}C), consistent with the population of a broad range of folding intermediates as detected by H/DX of the full-length protein (Figure 4.24).

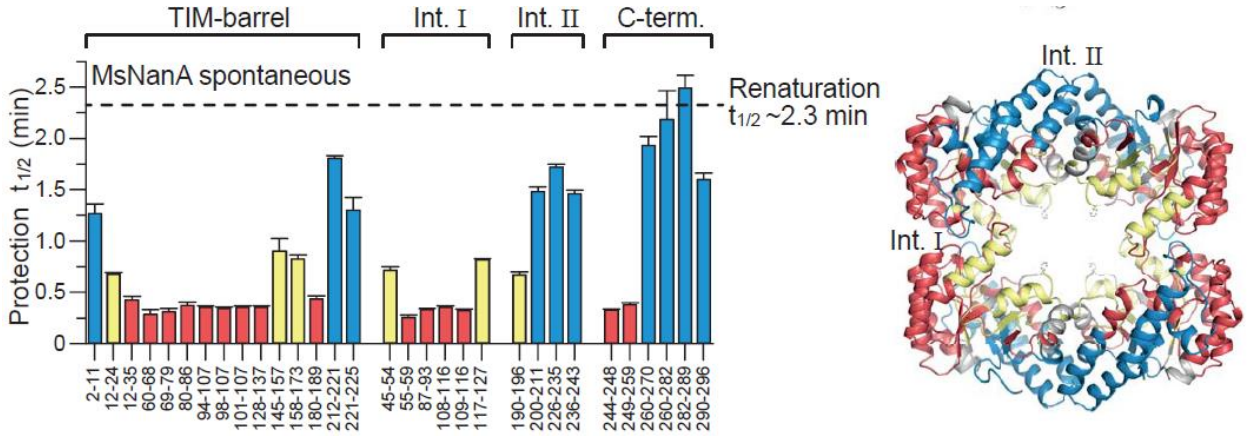


Figure 4.25. Comparison of DAPA and MsNANA refolding at peptide resolution

Apparent half-times of H/DX protection for peptides along the amino acid sequence during spontaneous renaturation of MsNANA at 10°C are presented as for DAPA in Figure 4.17, and are mapped on the tetramer structure of MsNANA (right panel). The half-times of protection are coloured from fast (red <0.5 min) to slow (blue >1.0 min). The experiments were performed by Florian Georgescauld.

Folding of the MsNANA TIM-barrel domain appears to initiate in a nucleus formed by the amphiphilic α-helical peptides H1 to H3 with their corresponding β-strands and H10 in the C-terminal domain, which is adjacent to H1. Structure formation then proceeds in a wave around the TIM-barrel and reaches completion with formation of H8 and the C-terminal domain helices H9 and H11 (Figure 4.25). Helices H1 to H3 are enriched in solvent exposed lysines, and their early formation may be coupled to organization of the hydrophobic TIM-barrel core which is stabilized by numerous aromatic (Tyr and Phe) residues (Figure 4.22). Although both the assisted folding of DAPA and the spontaneous renaturation of MsNANA involve local structure formation in the TIM-barrel, the folding regimes differ significantly. GroEL/ES-assisted folding of DAPA is characterized by multiple foci of initial structure formation, as might be expected for a protein confined in the chaperonin-cage, whereas in the case of MsNANA structure initiates asymmetrically. Another notable feature of MsNANA is that peptides located in interface I acquire protection at the fast rate of TIM-barrel folding (Figure 4.25), which may facilitate subunit assembly to occur coupled with folding. In contrast, interface I of DAPA acquires protection from exchange slower than the TIM-barrel, indicating that subunit folding and assembly are sequential steps (Figures 4.17A and 4.17B).

Thus, the refolding of chaperonin-independent MsNANA indeed, resembles the chaperonin-assisted refolding of DAPA, showing that GroEL/ES shifts the DAPA folding pathway towards that of a GroEL/ES-independent protein. An additional feature of MsNANA renaturation, the coupling of refolding and assembly, is an elegant way developed by this protein to overcome the entropic barrier, which in the case of DAPA is lowered due to confinement inside the GroEL/ES cage. Further studies would be required to elucidate whether such a mechanism is unique for

MsNANA folding or is a common feature shared by homologs of the obligate chaperonin substrates in *Mycoplasma synoviae*.

4.4 Ability to refold substrates is impaired in GroEL mutant with altered cavity charge

Previous studies have shown that the environment inside the chaperonin cage plays an important role in substrate folding, showing that GroEL mutants with altered net charge of the inner wall are unable to accelerate folding of DM-MBP and Rubisco proteins (Tang et al., 2006, Chakraborty et al., 2010).

The wall of the GroEL cis cavity is known to have a net charge of minus 42 (189 negatively and 147 positively charged amino acid residues). A number of negative charges positioned in the apical domain (residues E252, D253, E255, D359, D361, and E363), are arranged in clusters in two circular layers. Most of these residues are highly conserved among GroEL homologs, although they have no apparent role in the basic GroEL functions of substrate and GroES binding (Brocchieri and Karlin, 2000; Stan et al., 2003). In our study we probed a previously described mutant of a single ring version of GroEL, SR-KKK2 (D359, D361, E363) (Figure 4.26A), (Tang et al., 2006) and investigated its ability to refold DAPA.

As previously for SR-EL, a stable encapsulation of urea-denatured DAPA was achieved inside the SR-KKK2/ES complex in the low salt buffer (20 mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$) (Figure 4.26C). Stable binding of GroES to SR-KKK2 in these conditions was also confirmed in the ATPase assay (Figure 4.26D), where GroES fully inhibits the ATPase rate of SR-KKK2. The ATPase activity of SREL in the absence of GroES is comparable to the activity of a double-ring chaperonin in Figure 4.10. However, the dissociation of the SR-KKK2/ES

complex can be achieved by adding high salt concentration and guanidine hydrochloride into the solution (Figure 4.26D) as shown by the recovery of ATPase activity in these conditions.

We then analysed DAPA refolding inside the SR-KKK2 cage by enzymatic activity and found that SR-KKK2 appears to be unable to efficiently refold DAPA (Figure 4.26E). Only residual DAPA activity (under 15%) was detected in the refolding reaction at different time points. As previosuly shown for Rubisco, DAPA refolding appears to be arrested inside SR-KKK2.

Notably, similar to Rubisco, DAPA carries a slightly negative net charge on its surface (~ -4, Innovagen peptide calculator, Figure 4.26B) and therefore it seems plausible that the negative charge of GroEL cavity wall is necessary for creating a non-interacting environment for DAPA refolding. These results are in good agreement with previous studies, which showed that the repulsive environment of the GroEL cage results in acceleration of substrate folding, whereas specific residue-residue interactions between the substrate and the GroEL cavity induce a strong reduction of substrate folding rates (Sirur and Best, 2013). Further studies will be required to describe the mechanism of protein folding arrest in SR-KKK2.

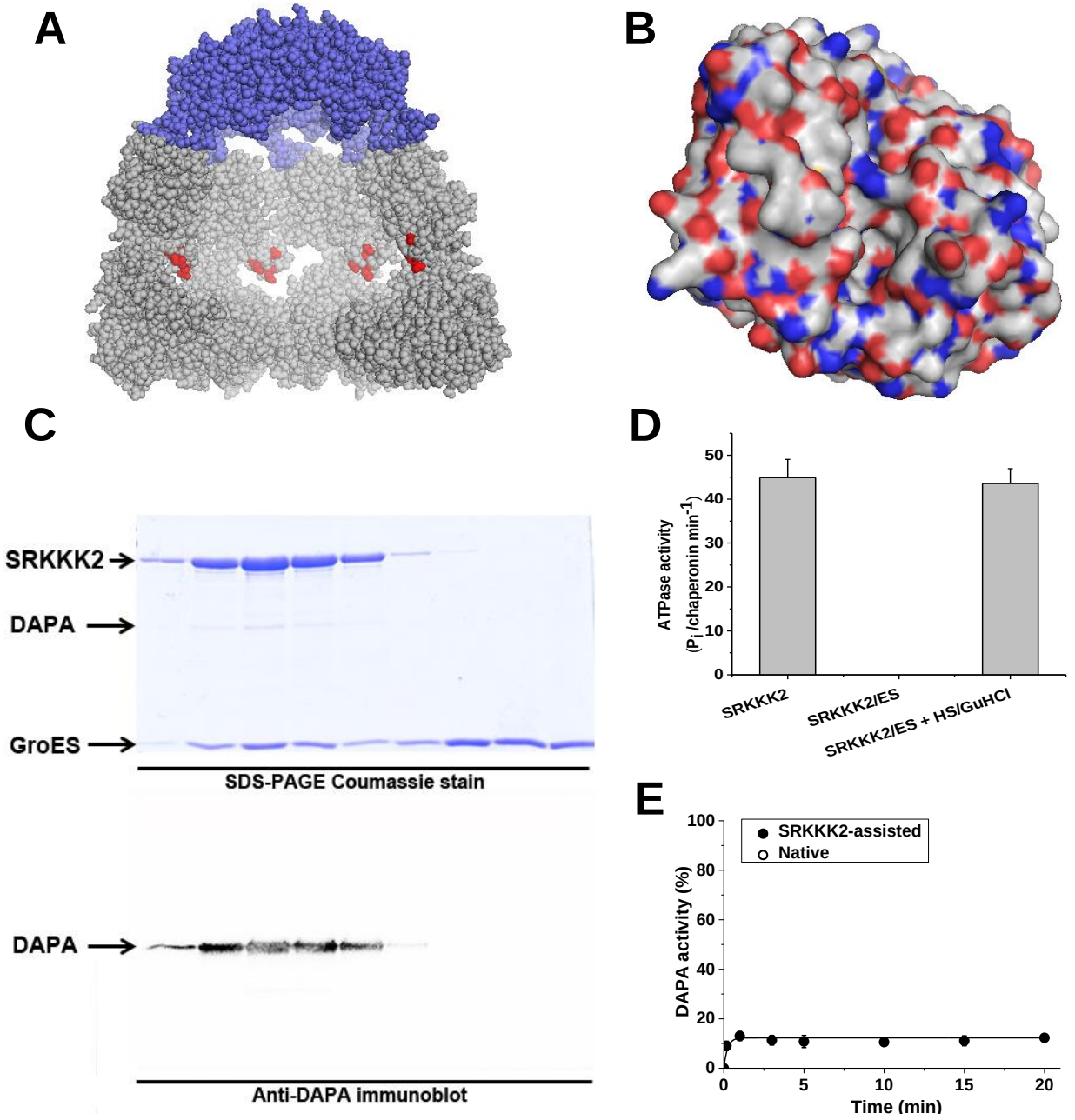


Figure 4.26: Refolding of DAPA inside SRKKK2/ES complex at 25°C is impaired

(A) Structure of a SRKKK2, a single ring mutant version of GroEL (PDB 1AON, Pymol).

(B) Surface of DAPA monomer, positively charged residues are coloured in blue, negatively charged residues in red.

(C) The encapsulation of DAPA was performed in low salt buffer C. 5 min after initiating folding an aliquot of the reaction was applied on a Superdex200 PC 3.2/10 column (GE Healthcare Life Sciences) using Ettan LC system (Amersham Biosciences) at 25°C. Fractions of 50 µL were collected at a flow rate 60 µL min^{-1} and analysed by SDS-PAGE followed by Coomassie staining and immunoblotting with polyclonal rabbit antibodies against DAPA.

(D) ATPase rate for SRKKK2 in the absence and presence of GroES, in low salt buffer C or in high salt buffer B with addition of 150mM HuHCl was measured as previously described (Tang, 2008) Standard deviations from three different measurements are shown. Note that SREL in the presence of GroES performs only a single round ATP hydrolysis.

(E) The encapsulation of urea-unfolded DAPA was performed in low salt buffer C at 25°C. At different time points, the refolding was stopped with 50mM CDTA and the chaperonin cage was opened by adding 100mM KCl with 80mM GuHCl final concentration. DAPA activity was measured as described. Mean and standard deviations from at least three independent experiments are shown.

5 Discussion

In this work, we present an answer to a long-standing question in the field of the chaperonin study, whether GroEL/ES actively promotes protein folding beyond preventing aggregation. We analyzed an obligate substrate of GroEL that refolds spontaneously at low temperatures and were therefore able to compare the refolding pathway of a protein inside and outside the chaperonin cage at peptide resolution.

We found that the chaperonin accelerates DAPA folding up to 30-fold over its spontaneous folding rate. Assembly of DAPA is not significantly affected by the chaperonin (accelerated only ~2.5-fold) which indicates that the main role of GroEL/ES is to accelerate the folding of individual protein subunits. We have also shown by a combination of methods, that transient aggregation is not the reason for slow spontaneous refolding of DAPA but rather, the refolding is slowed down due to the presence of a highly dynamic, kinetically trapped intermediate.

Applying the method of hydrogen-deuterium exchange coupled to liquid chromatography and mass-spectrometry at peptide resolution, we found that the chaperonin catalyses folding of the DAPA TIM-barrel domain. For the first time, we were able to compare the refolding pathway of a GroEL/ES substrate protein inside and outside the cage, and found that the chaperonin modifies the folding pathway of its substrate. The slow spontaneous folding of the TIM-barrel in the absence of GroEL/ES involves a concerted transition from an ensemble of dynamic folding intermediates to the native state that is associated with a high entropic energy barrier. In contrast, folding in the confining environment of the chaperonin cage is characterized by rapid stepwise and therefore less cooperative structure formation, which effectively results in a lower entropic energy barrier to folding.

We were also able to compare the refolding pathways of DAPA and its homolog MsNANA from *Mycoplasma synoviae* organism, whose genome lacks GroEL/ES. We purified and solved the structure of a DAPA homolog MsNANA, and showed that although structurally similar to DAPA, MsNANA is a completely GroEL/ES-independent protein. Applying the method of H/DX coupled to LC and MS, we found that the refolding pathway of MsNANA is profoundly different than that of DAPA. Similar to the case of GroEL/ES-assisted refolding of DAPA, spontaneous folding of MsNANA also employs a segmental folding regime. Moreover, folding and assembly of MsNANA appear to be coupled, which possibly represents an example of how a protein could adapt its refolding pathway to achieve chaperonin-independency. Segmental protein folding is likely to be a mechanism to efficiently avoid kinetically trapped states and ensure proper formation of the native state.

One of the future directions in studying chaperonin-assisted protein folding would be an investigation of the specific properties of the GroEL cage, which contribute to protein folding acceleration and define substrate folding inside the chaperonin. We briefly addressed this question using a mutant of GroEL with an altered charge of the inner cavity wall, SR-KKK2, and found that folding of DAPA inside SR-KKK2 is arrested. We propose that a possible reason for this might be a formation of a misfolded intermediate due to improper interactions between the substrate and the chaperonin wall. This finding demonstrates an active role of the chaperonin in folding its substrates and points out the importance of further studies to corroborate how the charge pattern inside GroEL/ES affects the refolding of its substrates.

We conclude that the chaperonin cage acts as a powerful folding catalyst for a set of proteins which otherwise fail to reach native state at a biologically relevant time scale.

5.1 Acceleration of substrate folding by GroEL/ES is biologically relevant

Acceleration of protein folding by GroEL/ES has been observed with a number of model substrates (Brinker 2001, Chakraborty 2010), however, the question remained whether the same would hold true for the natural substrates and whether accelerated folding of GroEL/ES substrates is merely a result of aggregation prevention (Apetri and Horwich 2008). In this work, the observed that the rate enhancement of folding by chaperonin cannot be explained by prevention of aggregation that might otherwise slow down the spontaneous folding reaction. We observed an acceleration of DAPA folding up to 30-fold over a wide range of concentrations and at different temperatures. This data argues strongly that aggregation does not affect the refolding rate of DAPA and GroEL/ES is a protein folding catalyst.

As an additional control, refolding of DAPA was also measured at very dilute concentrations excluding intermolecular association (aggregation or assembly), such as 100 pM DAPA. No association or aggregation was detected by dual colour FCCS at this concentration. Moreover, in these conditions (at 25-37^{o}C), a 50- to 130-fold acceleration of folding was measured by FCS, demonstrating that protein folding acceleration is an intrinsic property of GroEL/ES. Experiments of this kind are among the first examples of how methods of single molecule fluorescence can be applied to studying protein folding and the method was further developed by Amit Gupta and other members of Hartl department in Gupta et al., JMB 2014.

The catalysis of protein folding is highly biologically relevant. About 30-50 % of the obligate GroEL substrates, including DAPA, are known to share the TIM-barrel domain fold (Fujiwara et al., 2010; Kerner et al., 2005) and many of these proteins aggregate or are degraded in *E. coli* cells when GroEL/ES is depleted (Calloni et al., 2012; Kerner et al., 2005). Among these, DAPA also aggregates upon depletion of DnaK/DnaJ/GrpE in *E.coli* (Calloni et al., 2012). Together

with the results obtained in this study *in vitro*, these results provide insight into the interplay between Hsp70 and the chaperonin system inside the cell. We suggest that upon its release from the ribosome, DAPA is protected from aggregation and maintained in a folding-competent state by DnaK/J/E, which then deliver DAPA to the downstream GroEL/ES. This indicates that unassisted folding of DAPA, which requires several minutes is too slow compared to its rate of synthesis in the cell (~14 s, assuming a translation rate of 20 amino acids/s) (Powers et al., 2012). Moreover, knowing that *E.coli* cells divide every ~20 minutes at physiological temperature (37^{o}C) (Pine 1972), the slow folding rate of Class III GroEL substrate is not compatible with the life time of the cell. The absence of chaperones would lead to accumulation of not-yet folded DAPA in the cell and its strong aggregation. Based on these data and our results, we believe that GroEL/ES allows folding of proteins like DAPA to be completed faster than the time of synthesis, thus avoiding the build-up of unfolded protein and making efficient use of available chaperonin capacity.

5.2 Protein folding pathway is modified inside the chaperonin cage

Attempts have been made previously to characterize the folding pathway of several small proteins, such as malate dehydrogenase (Chen et al., 2001) or DM-MBP (Walters et al., 2013). However, none of these studies used a natural substrate of GroEL/ES or focused on comparing spontaneous and chaperonin-assisted folding. In this work, we found that the chaperonin mainly catalyses folding of individual subunits of DAPA, and for the first time, we were able to compare the folding pathway of a substrate protein inside and outside the chaperonin cage at peptide resolution.

Employing different methods, we found that in spontaneous refolding a folding intermediate represented as an ensemble of collapsed states is populated. The intermediate contains only

~20% secondary structure as measured by CD and has virtually no H/DX protection compared to the unfolded state in denaturant. This may imply that elements of secondary structure so far have not yet formed stable tertiary interactions, which corresponds to the idea of a framework folding mechanism, where elements of secondary structure appear simultaneously with or are promoted by a formation of long-range interactions (Gianni et al., 2003). Our analysis by H/DX coupled to MS showed that the spontaneous folding of DAPA proceeds in an apparent two-state manner with only kinetically trapped folding intermediate and largely folded subunits being significantly populated. The two-state character of folding is further supported by the kinetics as the spontaneous (as well as GroEL/ES-assisted) refolding reaction of DAPA shows an exponential time-course (Oliveberg and Wolynes, 2005). The Arrhenius plot of the temperature dependence of the refolding rates suggests the presence of important entropic component of the energy barrier on the spontaneous refolding pathway of DAPA, whereas the energy barrier in the GroEL-ES-assisted reaction has a more important enthalpic component and the entropic part appears to be reduced.

Performing H/DX coupled to LC and MS at peptide resolution allowed us to investigate the DAPA refolding pathway in detail. We found that nearly all peptides within the $(\beta\alpha)_8$ TIM-barrel acquire protection with an identical half-time of ~9.5 min at 10^{o}C, equivalent to the rate of subunit folding measured by enzymatic activity. Protection of the C-terminal domain was slower, with a half-time of ~20 min, which corresponds to the measured half-time of DAPA tetramer assembly. This data indicates that the TIM-barrel domain folds in a highly concerted manner without forming a compact detectable folding nucleus. Considering a large amount of distant polar and hydrophobic interactions that need to be formed, this result can explain the high entropic barrier and thus the slow kinetics of DAPA folding. Moreover, as some contacts in the

protein native state have been shown to be destabilizing (Shoemaker et al., 1999), random formation of such contacts would further slow the folding of DAPA. It was also predicted that a highly symmetrical $(\beta\alpha)_8$ barrel structure contains many long-range interactions and at a length of 224 amino acids exceeds the theoretical size limit for productive hydrophobic collapse (Lin and Zewail, 2012). Considering that the spontaneous reaction was stopped by addition of GroEL into the mix, the above mentioned data indicates that the chaperonin specifically recognizes the TIM-barrel domain of the substrate protein.

We found that GroEL/ES predominantly affects the folding of the TIM-barrel domain, accelerating structure formation at peptide level 20- to 50-fold compared to spontaneous folding (at 10^oC) (Figure 4.17). Different secondary structure elements of the TIM-barrel acquire H/DX protection at up to 3-fold different rates, with the fastest speed being measured for structure formation in α-helices. Interestingly, the difference in protection rate between TIM-barrel segments and peptides located in the C-terminal domain is magnified to more than 15-fold in assisted folding (only ~2-fold in spontaneous folding), reducing possible effects of inter-domain interference that may retard spontaneous folding. Thus, without GroEL/ES the TIM-barrel domain folds virtually all at once with a long search time and a high entropic penalty, while in the GroEL/ES catalysed reaction folding nucleates locally (Hu et al., 2013), building up the structure in a segmental manner and thereby reducing the entropic penalty (Figure 4.27).

Examples of segmental folding have been previously described in literature, e.g. for ribosomal protein S6 (α/β-protein of 96 amino acids, Haglund et al., 2011) where two independent folding nuclei appear to drive its folding. Attempts have also been made to identify sub-domain modules in amylase TIM-barrel proteins based on their structure and sequence alignment (Hleap et al., 2013). The segments of DAPA driving TIM-barrel folding inside GroEL/ES may represent such

"foldons" driving protein structure evolution and may be conserved among different bacterial species. Further studies characterizing refolding pathways of chaperonin-dependent DAPA homologues would be required to test this hypothesis.

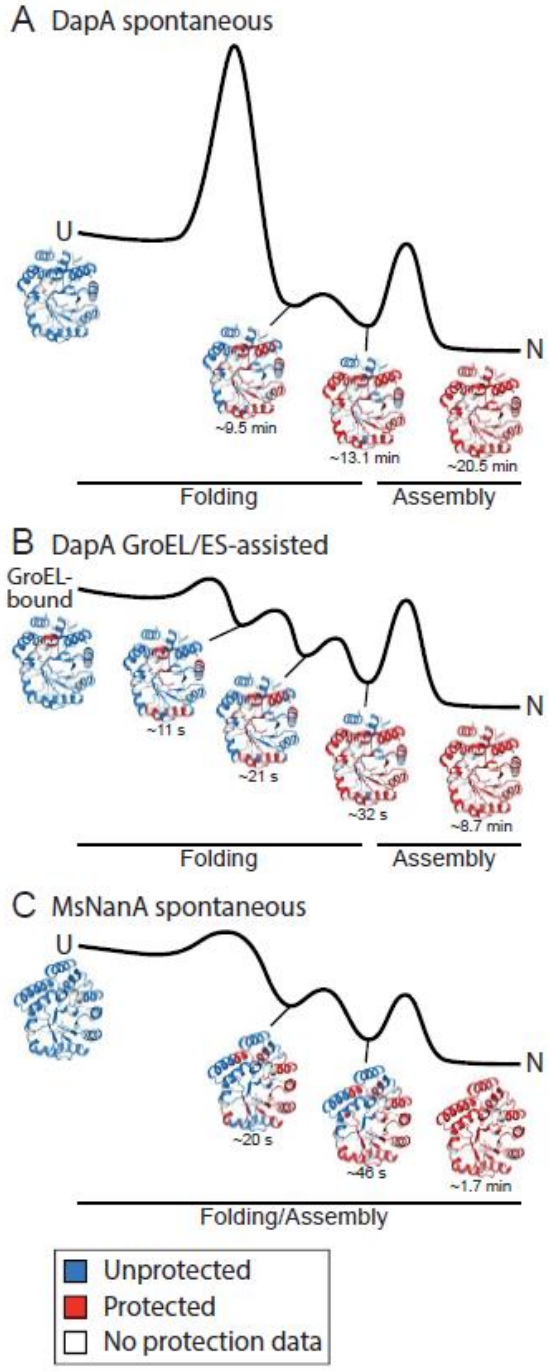


Figure 4.27. Mechanisms of Spontaneous and GroEL/ES-Assisted TIM-Barrel Folding

Free energy diagrams, schematically summarizing the salient features of spontaneous (A) folding and GroEL/ES-assisted folding (B) of DAPA and spontaneous folding of MsNANA (C). Intermediate states populated during folding with the approximate half-times indicated, as determined by H/DX-MS at peptide resolution, are tentatively assigned to different phases of the energy diagrams. Ribbon diagrams show acquisition of H/DX protection during folding in red. U, unfolded state; N, native tetramer.

How does GroEL/ES catalyse the folding of the DAPA TIM-barrel? According to theoretical predictions, steric confinement of unfolded protein in a repulsive (net-negatively charged) cage can accelerate folding by 1-2 orders of magnitude by restricting the entropic freedom of folding intermediates and making the formation of local and long-range contacts, including those present in the transition state, more favourable (Baumketner et al., 2003; Hayer-Hartl and Minton, 2006; Sirur and Best, 2013). The negative net charge of the GroEL/ES cavity wall likely plays an additional role through an ordering effect on water structure that may enhance hydrophobic core packing of encapsulated protein (England and Pande, 2008; Tang et al., 2006). Arrest of DAPA folding, observed in our study with SR-KKK2 mutant of GroEL, is in agreement with this theory and proves the critical role of the charges on GroEL cavity wall for the proper refolding of GroEL substrates. Further studies would be required to elucidate the exact mechanism of how charge pattern of GroEL affects protein folding.

Our results with the single ring variant of GroEL, SREL, show that folding catalysis is achieved upon a single round of protein encapsulation within the SREL/ES cage. This excludes repetitive binding and unfolding of misfolded states by GroEL (Lin et al., 2008; Sharma et al., 2008; Thirumalai and Lorimer, 2001) as a requirement of accelerated folding, at least for DAPA. Additionally, the lack of protection in the regions of DAPA forming assembly interfaces proves that substrate is stably encapsulated inside the SR-EL cage during the time of the experiment.

However, accelerated folding is not generally observed for all substrates. For example, in the case of the model substrate rhodanese, spontaneous and GroEL/ES-assisted refolding occur at similar rates (Brinker et al., 2001; Hofmann et al., 2010). We suggest that, as a result of co-evolution, the physical properties of the GroEL/ES-cage are particularly suited to achieve folding

catalysis for a subset of TIM-barrel domain proteins, which are known to occupy ~45 % of GroEL capacity *in vivo* (Kerner et al., 2005).

5.3 Escape from the chaperonin dependence

The genome of some parasitic bacteria, such as *Mycolplasma synoviae* and *Ureaplasma urealyticum* (Vanconcelos et al., 2005; Azia et al., 2012), is significantly reduced and contains only a subset of proteins necessary to maintain their basic survival. These bacteria have lost a number of proteins, including elaborate DNA repair systems and some chaperones (May et al., 2008). Interestingly, *M.synoviae* has also lost the GroEL/ES system, implying that homologs of strictly GroEL-dependent substrates in *E.coli* have either acquired specific strategies to become chaperone-independent or have been deleted. Our analysis showed that the genome of *M.synoviae* contains an ortholog of EcNANA and a close structural homolog of DAPA, MsNANA, which allowed us to gain insights into the strategies rendering this protein GroEL-independent. Studying MsNANA folding also has a significant biological impact as proteins of N-acetylneuraminate catabolism pathway have been described to be key players in pathogenicity of *Mycoplasma* strains (Bercic et al., Vet Microbiol. 2008).

In this study, we found that unlike DAPA, MsNANA has gained the ability to refold spontaneously with a high yield and folding rate. Moreover, MsNANA has evolved to become fully chaperonin-independent and does not bind GroEL upon renaturation. During renaturation, MsNANA does not form a largely unstructured intermediate but rather populates a number of intermediate states with various degrees of protection, as shown by H/DX on a full-length protein. The hydrophobic collapse is closely coupled with the gain of secondary structure during refolding, as measured by CD spectroscopy. To our knowledge, no similar studies of refolding pathway of *M.synoviae* have been performed to date. However, a recent computational study

employing Bayesian models of sequence evolution showed MsNANA to be under significant diversifying selection (May et al., 2009). This could imply that MsNANA, having already acquired GroEL-independency, is currently evolving further to improve its spontaneous folding efficiency.

We observed at the peptide level that the secondary structural elements of the TIM-barrel of MsNANA form sequentially, propagating "as a wave" from a nucleus initiating at α-helices H1-H3. The crystal structure of MsNANA showed that these amphiphilic helices are enriched in solvent exposed lysines which confer strong α-helical propensity (Burley and Petsko, 1985; Singh and Thornton, 1985).This observation is in good agreement with findings by May et al., 2008 and 2009 who reported that the residues of MsNANA under significant diversifying selection map to the regions of polar or cationic residues on the exterior of the protein. Moreover, burial of hydrophobic residues coupled with the formation of native structure is apparently facilitated by a hydrophobic core that is enriched in aromatic residues. These structural properties may explain the ability of MsNANA to nucleate TIM-barrel folding in distinct segments, a feature otherwise induced by the confining environment of the GroEL/ES cage.

Additionally, folding and assembly of MsNANA appears to be coupled, suggesting a mechanism of 'self-chaperoning' developed by MsNANA. The protein's independence of the chaperonin cage for folding would have facilitated the evolution of such a mechanism. Thus, our study demonstrates the role of GroEL/ES as protein folding catalyst, suggesting that the chaperonin modifies the folding pathway of its substrates and renders them to resemble the refolding of chaperonin-independent proteins such as MsNANA.

6 REFERENCES

Aisaka K1, Igarashi A, Yamaguchi K, Uwajima T. (1991) Purification, crystallization and characterization of N-acetylneuraminate lyase from Escherichia coli. *Biochem J*. 1991 Jun 1;276 (Pt 2):541-6.

Albanèse V, Yam AY, Baughman J, Parnot C, Frydman J. (2006) Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. *Cell*. Jan 13;124(1):75-88.

Alm E, Baker D. (1999) Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci U S A.* Sep 28;96(20):11305-10.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science (New York, N.Y.), 181*(96), 223-30. The American Association for the Advancement of Science. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4124164.

Apetri, A. C., & Horwich, A.L. (2008). Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. *Proceedings of the National Academy of Sciences*, *105*(45), 17351. National Acad Sciences.

Azia A, Unger R, Horovitz A. (2012) What distinguishes GroEL substrates from other Escherichia coli proteins? *FEBS J.* 2012 Feb;279(4):543-50. doi: 10.1111/j.1742-4658.2011.08458.x.

Balch WE, Morimoto RI, Dillin A, Kelly JW. (2008) Adapting proteostasis for disease intervention. *Science*. Feb 15;319(5865):916-9. doi: 10.1126/science.

Ballew RM, Sabelko J, Gruebele M. (1996) Observation of distinct nanosecond and microsecond protein folding events. Nat Struct Biol. Nov;3(11):923-6.

Banner DW, Bloomer AC, Petsko GA, Phillips DC, Pogson CI, Wilson IA, Corran PH, Furth AJ, Milman JD, Offord RE, Priddle JD, Waley SG. (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature*. 1975 Jun 19;255(5510):609-14.

Barbosa, J.A., Smith, B.J., DeGori, R., Ooi, H.C., Marcuccio, S.M., Campi, E.M., Jackson, W.R., Brossmer, R., Sommer, M., and Lawrence, M.C. (2000). Active site modulation in the

Nacetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed Haemophilus influenzae enzyme. *J. Mol. Biol*. 303, 405-421.

Basak S, Chattopadhyay K. (2014) Studies of protein folding and dynamics using single molecule fluorescence spectroscopy. *Phys Chem Chem Phys*. Jun 21;16(23):11139-49

Baumketner, A., Jewett, A., & Shea, J. E. (2003). Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape. *J Mol Biol*, *332*(3), 701-713.

Berg JM, Tymoczko JL, and Lubert Stryer. (2002) Biochemistry, 5th edition. *New York: W H Freeman*; ISBN-10: 0-7167-3051-0

Bicout, D. J., & Szabo, A. (2000). Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model. *Protein Science, 9*(3), 452-465. Cold Spring Harbor Laboratory Press.

Bingel-Erlenmeyer R, Kohler R, Kramer G, Sandikci A, Antolić S, Maier T, Schaffitzel C, Wiedmann B, Bukau B, Ban N. (2008) A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. *Nature*. Mar 6;452(7183):108-11. doi: 10.1038/nature06683.

Blickling S, Beisel HG, Bozic D, Knäblein J, Laber B, Huber R. (1997) Structure of dihydrodipicolinate synthase of Nicotiana sylvestris reveals novel quaternary structure. *J Mol Biol.* 1997 Dec 12;274(4):608-21.

Bracher A, Starling-Windhof A, Hartl FU, Hayer-Hartl M. (2011) Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat Struct Mol Biol.* 2011 Jul 17;18(8):875-80. doi: 10.1038/nsmb.2090.

Brandt, F., Etchells, S. A., Ortiz, J. O., Elcock, A. H., Hartl, F Ulrich, & Baumeister, W. (2009). The native 3D organization of bacterial polysomes. *Cell, 136*(2), 261-271.

Braselmann E, Chaney JL, Clark PL. (2013) Folding the proteome. *Trends Biochem Sci.* Jul;38(7):337-44. doi: 10.1016/j.tibs.2013.05.001.

Brinker, A., Pfeifer, G., Kerner, M J, Naylor, D J, Hartl, F U, & Hayer-Hartl, M. (2001). Dual function of protein confinement in chaperonin-assisted protein folding. *Cell*, *107*(2), 223-33.

Brockwell DJ, Smith DA, Radford SE. (2000) Protein folding mechanisms: new methods and emerging ideas. *Curr Opin Struct Biol.* Feb;10(1):16-25.

Brocchieri L, Karlin S. (2000) Conservation among Hsp60 sequences in relation to structure, function, and evolution. *Protein Sci*. 2000 Mar;9(3):476-86.

Bukau B, Horwich AL. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell*. 1998 Feb 6;92(3):351-66.

Bukau B, Deuerling E, Pfund C, Craig EA. (2000) Getting newly synthesized proteins into shape. *Cell*. Apr 14;101(2):119-22.

Burkitt, W., and O'Connor, G. (2008). Assessment of the repeatability and reproducibility of hydrogen/deuterium exchange mass spectrometry measurements. *Rapid Commun. Mass Spectrom.* 22, 3893-3901.

Burley SK, Petsko GA. (1985) Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science*. 1985 Jul 5;229(4708):23-8.

Calloni G, Chen T, Schermann SM, Chang HC, Genevaux P, Agostini F, Tartaglia GG, Hayer-Hartl M, Hartl FU. (2012) DnaK functions as a central hub in the E. coli chaperone network. *Cell Rep*. Mar 29;1(3):251-64. doi: 10.1016/j.celrep.2011.12.007.

Camacho, C. J., & Thirumalai, D. (1995). Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(5), 1277-1281.

Cavagnero S, Dyson HJ, Wright PE. (1999) Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J Mol Biol.* Jan 8;285(1):269-82.

Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, Sikor M, Jiang G, Lamb DC, Hartl FU, Hayer-Hartl M. (2010) Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell*. Jul 9;142(1):112-22.

Chapman E, Farr GW, Usaite R, Furtak K, Fenton WA, Chaudhuri TK, Hondorp ER, Matthews RG, Wolf SG, Yates JR, Pypaert M, Horwich AL. (2006) Global aggregation of newly translated proteins in an Escherichia coli strain deficient of the chaperonin GroEL. *Proc Natl Acad Sci U S A.* 2006 Oct 24;103(43):15800-5.

Chaudhry, C., Farr, George W, Todd, M. J., Rye, Hays S, Brunger, A. T., Adams, P. D., et al. (2003a). Role of the g-phosphate of ATP in triggering protein folding by GroEL±GroES: function, structure and energetics. *EMBO Journal*, *22*(19), 4877-4887.

Chaudhry, C., Farr, George W, Todd, M. J., Rye, Hays S, Brunger, A. T., Adams, P. D., et al. (2003b). Role of the γ-phosphate of ATP in triggering protein folding by GroEL–GroES: function, structure and energetics. *the The European Molecular Biology Organization Journal*, *22*(19), 4877-4887. Oxford University Press.

Chen J, Walter S, Horwich AL, Smith DL. (2001) Folding of malate dehydrogenase inside the GroEL-GroES cavity. *Nat Struct Biol.* 2001 Aug;8(8):721-8.

Cheung MS. (2013) Where soft matter meets living matter--protein structure, stability, and folding in the cell. *Curr Opin Struct Biol.* Apr;23(2):212-7. doi: 10.1016/j.sbi.2013.02.005.

Clare DK, Vasishtan D, Stagg S, Quispe J, Farr GW, Topf M, Horwich AL, Saibil HR. (2012) ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell*. 2012 Mar 30;149(1):113-23. doi: 10.1016/j.cell.2012.02.047.

Coyle JE, Texter FL, Ashcroft AE, Masselos D, Robinson CV, Radford SE. (1999) GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism. *Nat Struct Biol.* 1999 Jul;6(7):683-90.

del Alamo M, Hogan DJ, Pechmann S, Albanese V, Brown PO, Frydman J. (2011) Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. *PLoS Biol.* Jul;9(7):e1001100. doi: 10.1371/journal.pbio.1001100.

Dill, K A, & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat Str Biol 4*(1), 10-9. New York, NY: Nature Pub. Co., c1994-c2003. doi: 10.1038/nsb0197-10.

Dobson, C. M., and Karplus, M. (1999). The fundamentals of protein folding: bringing together theory and experiment. *Curr Opin Struct Biol* 9, 92-101.

Dobson RC, Valegård K, Gerrard JA. (2004) The crystal structure of three site-directed mutants of Escherichia coli dihydrodipicolinate synthase: further evidence for a catalytic triad. *J Mol Biol.* 2004 Apr 23;338(2):329-39.

Doolittle, R.F. (1995) The Origins and Evolution of Eukaryotic Proteins. Phil. Trans. R. Soc. Lond. B September 29, 1995 349 1329 235-240; 1471-2970

Elad N, Farr GW, Clare DK, Orlova EV, Horwich AL, Saibil HR. (2007) Topologies of a substrate protein bound to the chaperonin GroEL. *Mol Cell.* 2007 May 11;26(3):415-26.

Ellis, R. J., and Hartl, F. U. (1999). Principles of protein folding in the cellular environment. *Curr Opin Struct Biol* 9, 102-110.

Ellis RJ. (2001) romolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol.* Feb;11(1):114-9.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2126-2132.

England, J. L., & Pande, V. S. (2008). Potential for modulation of the hydrophobic effect inside chaperonins. *Biophysical Journal*, *95*(7), 3391-3399.

Englander SW. (2000) Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct.* 29:213-38.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. Sect. D Biol. Crystallogr. 62, 72-82.

Evans, P.R. (1997). *Scala. Joint CCP4 + ESF-EACBM Newsletter on Prot Crystallogr* 33, 22-24.

Ewalt KL, Hendrick JP, Houry WA, Hartl FU. (1997) In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell*. 1997 Aug 8;90(3):491-500.

Falke S, Tama F, Brooks CL 3rd, Gogol EP, Fisher MT. (2005) The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. J Mol Biol. 2005 Apr 22;348(1):219-30.

Fang J, Nevin P, Kairys V, Venclovas C, Engen JR, Beuning PJ. (2014) Conformational analysis of processivity clamps in solution demonstrates that tertiary structure does not correlate with protein dynamics. *Structure.* Apr 8;22(4):572-81. doi: 10.1016/j.str.2014.02.001.

Fani R1, Liò P, Chiarelli I, Bazzicalupo M. (1994) The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the hisA and hisF genes. *J Mol Evol.* 1994 May;38(5):489-95.

Farr, George W, Furtak, Krystyna, Rowland, M. B., Ranson, Neil A, Saibil, Helen R, Kirchhausen, T., et al. (2000). Multivalent Binding of Nonnative Substrate Proteins by the Chaperonin GroEL. *Cell*, *100*, 561-573.

Fenton, W A, Kashi, Y., Furtak, K, & Horwich, A L. (1994). Residues in chaperonin GroEL required for polypeptide binding and release. *Nature*, *371*(6498), 614-9. doi: 10.1038/371614a0.

Fenton WA, Horwich AL. (2003) Chaperonin-mediated protein folding: fate of substrate polypeptide. *Q Rev Biophys*. 2003 May;36(2):229-56.

Fersht, A. R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224, 771-782.

Fink AL. (1999) Chaperone-mediated protein folding. *Physiol Rev.* Apr;79(2):425-49

Frydman J. (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem*. 2001;70:603-47.

Frydman J, Erdjument-Bromage H, Tempst P, Hartl FU. (1999) Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol.* Jul;6(7):697-705.

Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H. (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J.* 2010 May 5;29(9):1552-64. doi: 10.1038/emboj.2010.52.

Gautschi M, Mun A, Ross S, Rospert S. (2002) A functional chaperone triad on the yeast ribosome. *Proc Natl Acad Sci U S A.* Apr 2;99(7):4209-14.

Gerlt JA, Babbitt PC. (2001) Barrels in pieces? *Nat Struct Biol.* 2001 Jan;8(1):5-7.

Gianni S1, Guydosh NR, Khan F, Caldas TD, Mayor U, White GW, DeMarco ML, Daggett V, Fersht AR. (2003) Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A.* 2003 Nov 11;100(23):13286-91

Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH. (2000) The complete sequence of the mucosal pathogen Ureaplasma urealyticum. *Nature*. 2000 Oct 12;407(6805):757-62.

Goloubinoff, P., Christeller, J. T., Gatenby, A. A., & Lorimer, G. H. (1989). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfoleded state depends on two chaperonin proteins and Mg-ATP. *Nature*, *342*(6252), 884-9. doi: 10.1038/342884a0.

Goloubinoff, P., Gatenby, A. A., & Lorimer, G. H. (1989). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature*, *337*(6202), 44-47.

Gorovits BM, McGee WA, Horowitz PM. (1998) Rhodanese folding is controlled by the partitioning of its folding intermediates. *Biochim Biophys Acta*. 1998 Jan 15;1382(1):120-8.

Gregor, I., and Enderlein, J. (2007). Time-resolved methods in biophysics. 3. Fluorescence lifetime correlation spectroscopy. *Photochem. Photobiol*. Sci. 6, 13-18.

Griffin MD, Dobson RC, Gerrard JA, Perugini MA. (2010) Exploring the dihydrodipicolinate synthase tetramer: how resilient is the dimer-dimer interface? *Arch Biochem Biophys*. 2010 Feb 1;494(1):58-63. doi: 10.1016/j.abb.2009.11.014.

Groebke K1, Renold P, Tsang KY, Allen TJ, McClure KF, Kemp DS. (1996) Template-nucleated alanine-lysine helices are stabilized by position-dependent interactions between the lysine side chain and the helix barrel. *Proc Natl Acad Sci U S A.* 1996 Apr 30;93(9):4025-9.

Gupta AJ, Haldar S, Miličić G, Hartl FU, Hayer-Hartl M. (2014) Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *J Mol Biol.* 2014 Jul 29;426(15):2739-54. doi: 10.1016/j.jmb.2014.04.018.

Haber, E, Anfinsen, CB (1962) Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J Biol Chem. 1962 Jun;*237:1839-44.

Haglund E, Danielsson J, Kadhirvel S, Lindberg MO, Logan DT, Oliveberg M. (2012) Trimming down a protein structure to its bare foldons: spatial organization of the cooperative unit. *J Biol Chem.* 2012 Jan 20;287(4):2731-8. doi: 10.1074/jbc.M111.312447.

Hamada D, Segawa S, Goto Y. (1996) Non-native alpha-helical intermediate in the refolding of beta-lactoglobulin, a predominantly beta-sheet protein. *Nat Struct Biol.* 1996 Oct;3(10):868-73.

Hartl FU. (1996) Molecular chaperones in cellular protein folding. *Nature*. Jun 13;381(6583):571-9.

Hartl FU, Bracher A, Hayer-Hartl M. (2011) Molecular chaperones in protein folding and proteostasis. *Nature*. Jul 20;475(7356):324-32. doi: 10.1038/nature10317.

Hartl FU, Hayer-Hartl M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*. 2002 Mar 8;295(5561):1852-8.

Hartl, F Ulrich, & Hayer-Hartl, Manajit. (2009). Converging concepts of protein folding *in vitro* and *in vivo*. *Nature Str Mol Biol*, *16*(6), 574-81. doi: 10.1038/nsmb.1591.

Hegyi H, Gerstein M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J Mol Biol.* 1999 Apr 23;288(1):147-64.

Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Dennis DT, Georgopoulos CP, Hendrix RW, Ellis RJ. (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature*. 1988 May 26;333(6171):330-4.

Hleap JS, Susko E, Blouin C. (2013) Defining structural and evolutionary modules in proteins: a community detection approach to explore sub-domain architecture. *BMC Struct Biol.* 2013 Oct 16;13:20. doi: 10.1186/1472-6807-13-20.

Höcker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R. (2001) Dissection of a (betaalpha)8-barrel enzyme into two folded halves. *Nat Struct Biol*. 2001 Jan;8(1):32-6.

Hofmann, H., Hillger, F., Pfeil, S.H., Hoffmann, A., Streich, D., Haenni, D., Nettels, D., Lipman, E.A., and Schuler, B. (2010). Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proc. Natl. Acad. Sci. U.S.A. 107*, 11793-11798.

Horst R, Fenton WA, Englander SW, Wüthrich K, Horwich AL. (2007) Folding trajectories of human dihydrofolate reductase inside the GroEL GroES chaperonin cavity and free in solution. *Proc Natl Acad Sci U S A.* 2007 Dec 26;104(52):20788-92.

Horwich AL, Fenton WA, Chapman E, Farr GW. (2007) Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol.* 2007;23:115-45.

Houde, D., Berkowitz, S.A., and Engen, J.R. (2011). The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. J Pharm Sci 100, 2071-2086. Kabsch, W. (2010). *XDS. Acta Crystallogr. Sect. D Biol. Crystallogr*. 66, 125-132.

Houry WA1, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU. (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature*. 1999 Nov 11;402(6758):147-54.

Hu, W., Walters, B.T., Kan, Z.Y., Mayne, L., Rosen, L.E., Marqusee, S., and Englander, S.W. (2013). Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A. 110*, 7684-7689.

Hutton CA, Perugini MA, Gerrard JA. (2007) Inhibition of lysine biosynthesis: an evolving antibiotic strategy. *Mol Biosyst.* 2007 Jul;3(7):458-65.

Izard T, Lawrence MC, Malby RL, Lilley GG, Colman PM. (1994) The three-dimensional structure of N-acetylneuraminate lyase from Escherichia coli. *Structure*. 1994 May 15;2(5):361-9.

Jaenicke, R. (1991). Protein stability and protein folding. *Ciba Found Symp* 161, 206-216; discussion 217-221.

Jennings PA, Wright PE. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*. 1993 Nov 5;262(5135):892-6.

Jewett AI, Baumketner A, Shea JE. (2004) Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: creation of an alternate fast folding pathway. *Proc Natl Acad Sci U S A.* 2004 Sep 7;101(36):13192-7.

Jewett AI, Shea JE. (2008) Do chaperonins boost protein yields by accelerating folding or preventing aggregation? *Biophys J.* 2008 Apr 15;94(8):2987-93. doi: 10.1529/biophysj.107.113209.

Juers DH, Huber RE, Matthews BW. (1999) Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between beta-galactosidase and other glycohydrolases. *Protein Sci.* 1999 Jan;8(1):122-36.

Jürgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R. (2000) Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A.* 2000 Aug 29;97(18):9925-30.

Kabsch W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr*. 2010 Feb;66(Pt 2):133-44. doi: 10.1107/S0907444909047374.

Kampinga HH1, Craig EA. (2010) The Hsp70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol.* 2010 Aug;11(8):579-92. doi: 10.1038/nrm2941.

Katta V, Chait BT (1991) Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Commun Mass Spectrom*. Apr;5(4):214-7.

Kerner, Michael J, Naylor, Dean J, Ishihama, Y., Maier, Tobias, Chang, H.-C., Stines, A. P., et al. (2005). Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell*, *122*(2), 209-220.

Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem.*;82:323-55. doi: 10.1146/annurev-biochem-060208-092442.

Kleywegt, G.T., and Jones, T.A. (1994). A super position. *CCP4/ESF-EACBM Newsletter on Protein Crystallogr*. 31, 9-14.

Koculi E, Horst R, Horwich AL, Wüthrich K. (2011) Nuclear magnetic resonance spectroscopy with the stringent substrate rhodanese bound to the single-ring variant SR1 of the E. coli chaperonin GroEL. *Protein Sci.* 2011 Aug;20(8):1380-6. doi: 10.1002/pro.665.

Koplin A, Preissler S, Ilina Y, Koch M, Scior A, Erhardt M, Deuerling E. (2010) A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. *J Cell Biol.* Apr 5;189(1):57-68. doi: 10.1083/jcb.200910074.

Krichevsky AM, Kosik KS. (2002) RNAi functions in cultured mammalian neurons. *Proc Natl Acad Sci U S A.* Sep 3;99(18):

Kuwajima K, Yamaya H, Sugai S. (1996) burst-phase intermediate in the refolding of beta-lactoglobulin studied by stopped-flow circular dichroism and absorption spectroscopy. *J Mol Biol.* Dec 13;264(4):806-22.

Kuwata K, Shastry R, Cheng H, Hoshino M, Batt CA, Goto Y, Roder H. (2001) Structural and kinetic characterization of early folding events in beta-lactoglobulin. *Nat Struct Biol.* 2001 Feb;8(2):151-5.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc 3, 1171-1179.

Lawrence, M.C., and Colman, P.M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* 234, 946-950.

Lanman J, Lam TT, Emmett MR, Marshall AG, Sakalian M, Prevelige PE Jr. (2004) Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nat Struct Mol Biol.* Jul;11(7):676-7.

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

Lin, Z., Madan, D., & Rye, Hays S. (2008). GroEL stimulates protein folding through forced unfolding. *Nat Str Mol Biol*, *15*(3), 303-311. doi: 10.1038/nsmb.1394.

Lin MM, Zewail AH. (2012) Hydrophobic forces and the length limit of foldable protein domains. *Proc Natl Acad Sci U S A.* 2012 Jun 19;109(25):9851-6. doi: 10.1073/pnas.1207382109.

Lorch M, Mason JM, Clarke AR, Parker MJ. (1999) Effects of core mutations on the folding of a beta-sheet protein: implications for backbone organization in the I-state. *Biochemistry*. Jan 26;38(4):1377-85.

Lorimer GH. (1996) A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. *FASEB J.* 1996 Jan;10(1):5-9.

Lucent D1, England J, Pande V. (2009) Inside the chaperonin toolbox: theoretical and computational models for chaperonin mechanism. *Phys Biol.* 2009 Feb 10;6(1):015003. doi: 10.1088/1478-3975/6/1/015003.

Luger K, Hommel U, Herold M, Hofsteenge J, Kirschner K. (1989) Correct folding of circularly permuted variants of a beta alpha barrel enzyme in vivo. *Science*. 1989 Jan 13;243(4888):206-10.

Makino Y, Taguchi H, Yoshida M. (1993) Truncated GroEL monomer has the ability to promote folding of rhodanese without GroES and ATP. *FEBS Lett.* 1993 Dec 27;336(2):363-7.

Marcsisin SR, Engen JR. (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us? *Anal Bioanal Chem*. Jun;397(3):967-72. doi: 10.1007/s00216-010-3556-4.

May M, Brown DR. (2008) Genetic variation in sialidase and linkage to N-acetylneuraminate catabolism in Mycoplasma synoviae. *Microb Pathog.* 2008 Jul;45(1):38-44. doi: 10.1016/j.micpath.2008.02.002.

May M, Brown DR. (2009) Diversifying and stabilizing selection of sialidase and N-acetylneuraminate catabolism in Mycoplasma synoviae. *J Bacteriol*. 2009 Jun;191(11):3588-93. doi: 10.1128/JB.00142-09.

Mayer MP. (2010) Gymnastics of molecular chaperones. *Mol Cell.* Aug 13;39(3):321-31. doi: 10.1016/j.molcel.2010.07.012.

Meyer AS, Gillespie JR, Walther D, Millet IS, Doniach S, Frydman J. (2003) Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell*. 2003 May 2;113(3):369-81.

McGuffee SR, Elcock AH. (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol.* Mar 5;6(3):e1000694. doi: 10.1371/journal.pcbi.1000694.

Mendes HF, van der Spuy J, Chapple JP, Cheetham ME. (2005) Mechanisms of cell death in rhodopsin retinitis pigmentosa: implications for therapy. *Trends Mol Med.* Apr;11(4):177-85.

Minton AP, Wilf J. (1981) Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry*. Aug 18;20(17):4821-6.

Miranker A, Radford SE, Karplus M, Dobson CM. (1991) Demonstration by NMR of folding domains in lysozyme. *Nature*. 1991 Feb 14;349(6310):633-6.

Mitsakos, V., Devenish, S.R., O'Donnell, P.A., Gerrard, J.A., and Hutton, C.A. (2011). LC-MS and NMR characterization of the purple chromophore formed in the o-aminobenzaldehyde assay of dihydrodipicolinate synthase. *Bioorg. Med. Chem*. 19, 1535-1540.

Moens PD, Gratton E, Salvemini IL (1972) Fluorescence correlation spectroscopy, raster image correlation spectroscopy, and number and brightness on a commercial confocal laser scanning microscope with analog detectors (Nikon C1). *Microsc Res Tech*. Apr;74(4):377-88.

Mogk A, Tomoyasu T, Goloubinoff P, Rüdiger S, Röder D, Langen H, Bukau B. (1999) Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* 1999 Dec 15;18(24):6934-49.

Morozova-Roche LA, Jones JA, Noppe W, Dobson CM. (1999) Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme. *J Mol Biol.* Jun 18;289(4):1055-73.

Mukhopadhyay, S., Krishnan, R., Lemke, E.A., Lindquist, S., and Deniz, A.A. (2007). A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2649-2654.

Muñoz V, Henry ER, Hofrichter J, Eaton WA. (1998) A statistical mechanical model for beta-hairpin kinetics. *Proc Natl Acad Sci U S A.* 1998 May 26;95(11):5872-9.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D Biol. Crystallogr*. 53, 240-255.

Muscroft-Taylor AC, Catchpole RJ, Dobson RC, Pearce FG, Perugini MA, Gerrard JA. (2010) Disruption of quaternary structure in Escherichia coli dihydrodipicolinate synthase (DHDPS) generates a functional monomer that is no longer inhibited by lysine. *Arch Biochem Biophys*. 2010 Nov 15;503(2):202-6. doi: 10.1016/j.abb.2010.08.009.

Müller, C.B., Loman, A., Pacheco, V., Koberling, F., Willbold, D., Richtering, W., and Enderlein, J. (2008). Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *EPL (Europhysics Letters)* 83, 46001.

Nagano N, Hutchinson EG, Thornton JM. (1999) Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels. *Protein Sci.* 1999 Oct;8(10):2072-84.

Nagi AD, Anderson KS, Regan L. (1999) Using loop length variants to dissect the folding pathway of a four-helix-bundle protein. *J Mol Biol.* Feb 12;286(1):257-65.

Niwa T, Ying BW, Saito K, Jin W, Takada S, Ueda T, Taguchi H. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia

coli proteins. *Proc Natl Acad Sci U S A.* 2009 Mar 17;106(11):4201-6. doi: 10.1073/pnas.0811922106.

Oh E, Becker AH, Sandikci A, Huber D, Chaba R, Gloge F, Nichols RJ, Typas A, Gross CA, Kramer G, Weissman JS, Bukau B. (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*. Dec 9;147(6):1295-308. doi: 10.1016/j.cell.2011.10.044.

Oliveberg M, Wolynes PG. (2005) The experimental survey of protein-folding energy landscapes. *Q Rev Biophys.* Aug;38(3):245-88.

Onuchic JN, Luthey-Schulten Z, Wolynes PG. (1997) Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem.* 1997;48:545-600.

Onuchic JN, Wolynes PG. (2004) Theory of protein folding. *Curr Opin Struct Biol*. 2004 Feb;14(1):70-5.

Pande VS, Grosberg AY, Tanaka T. (1994) Thermodynamic procedure to synthesize heteropolymers that can renature to recognize a given target molecule. *Proc Natl Acad Sci U S A*. Dec 20;91(26):12976-9.

Paul S, Singh C, Mishra S, Chaudhuri TK. (2007) The 69 kDa Escherichia coli maltodextrin glucosidase does not get encapsulated underneath GroES and folds through trans mechanism during GroEL/GroES-assisted folding. *FASEB J*. 2007 Sep;21(11):2874-85.

Pauling, L., and Corey, R. B. (1951a). Atomic coordinates and structure factors for two helical configurations of polypeptide chains. Proc Natl Acad Sci U S A *37*, 235-240.

Pauling, L., and Corey, R. B. (1951b). The pleated sheet, a new layer configuration of polypeptide chains. Proc Natl Acad Sci U S A *37*, 251-256.

Pine MJ. (1972) Turnover of intracellular proteins. *Annu Rev Microbiol*. 1972;26:103-26.

Plotkin SS, Onuchic JN. (2002) Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q Rev Biophys.* May;35(2):111-67.

Plumb, R.S., Johnson, K.A., Rainville, P., Smith, B.W., Wilson, I.D., Castro-Perez, J.M., and Nicholson, J.K. (2006). UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom*. 20, 1989-1994.

Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE. (2009) cal and chemical approaches to diseases of proteostasis deficiency. *Annu Rev Biochem*.;78:959-91. doi: 10.1146/annurev.biochem.052308.114844.

Preissler S, Deuerling E. (2012) Ribosome-associated chaperones as key players in proteostasis. *Trends Biochem Sci.* Jul;37(7):274-83. doi: 10.1016/j.tibs.2012.03.002.

Radford SE, Dobson CM, Evans PA. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature*. 1992 Jul 23;358(6384):302-7.

Radford, S. E. (2000). Protein folding: progress made and promises ahead. *Trends Biochem Sci* 25, 611-618.

Raineri E, Ribeca P, Serrano L, Maier T. (2010) A more precise characterization of chaperonin substrates. *Bioinformatics*. 2010 Jul 15;26(14):1685-9. doi: 10.1093/bioinformatics/btq287.

Saven JG1, Wolynes PG. (1996) Local conformational signals and the statistical thermodynamics of collapsed helical proteins. *J Mol Biol.* Mar 22;257(1):199-216.

Shtilerman M, Lorimer GH, Englander SW. (1999) Chaperonin function: folding by forced unfolding. *Science*. Apr 30;284(5415):822-5.

Ramachandran, G. N., and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. Adv Protein Chem *23*, 283-438.

Ranson, N a, Farr, G W, Roseman, a M., Gowen, B., Fenton, W. a, Horwich, a L., et al. (2001). ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, *107*(7), 869-79.

Reboul, C.F., Porebski, B.T., Griffin, M.D., Dobson, R.C., Perugini, M.A., Gerrard, J.A., and Buckle, A.M. (2012). Structural and dynamic requirements for optimal activity of the essential bacterial enzyme dihydrodipicolinate synthase. *PLOS Comput. Biol. 8*, e1002537.

Reissmann S, Parnot C, Booth CR, Chiu W, Frydman J. (2007) Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. *Nat Struct Mol Biol*. 2007 May;14(5):432-40.

Roseman AM, Chen S, White H, Braig K, Saibil HR. (1996) The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell*. 1996 Oct 18;87(2):241-51.

Rothnie A1, Clarke AR, Kuzmic P, Cameron A, Smith CJ. (2011) A sequential mechanism for clathrin cage disassembly by 70-kDa heat-shock cognate protein (Hsc70) and auxilin. *Proc Natl Acad Sci U S A. 2011* Apr 26;108(17):6927-32. doi: 10.1073/pnas.1018845108.

Rozema D, Gellman SH. (1996) Artificial chaperone-assisted refolding of denatured-reduced lysozyme: modulation of the competition between renaturation and aggregation. *Biochemistry*. 1996 Dec 10;35(49):15760-71.

Rye HS, Roseman AM, Chen S, Furtak K, Fenton WA, Saibil HR, Horwich AL. (1999) GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell*. 1999 Apr 30;97(3):325-38.

Saibil H. (2013) Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol.* Oct;14(10):630-42. doi: 10.1038/nrm3658.

Saibil HR, Fenton WA, Clare DK, Horwich AL. (2013) tructure and allostery of the chaperonin GroEL. *J Mol Biol*. 2013 May 13;425(9):1476-87. doi: 10.1016/j.jmb.2012.11.028.

Shoemaker BA, Wang J, Wolynes PG. (1999) Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. *J Mol Biol*. 1999 Apr 2;287(3):675-94.

Schultz SG, Wilson NL, Epstein W. (1962) Cation transport in Escherichia coli. II. Intracellular chloride concentration. *J Gen Physiol*. 1962 Sep;46:159-66.

Selvaraj S, Gromiha MM. (1998) Importance of long-range interactions in (alpha/beta)8 barrel fold. *J Protein Chem.* 1998 Oct;17(7):691-7.

Sharma, S., Chakraborty, K., Müller, B. K., Astola, N., Tang, Y.-chi, Lamb, D. C., et al. (2008). Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell*, *133*(1), 142-53. doi: 10.1016/j.cell.2008.01.048.

Shorter J. (2011) The mammalian disaggregase machinery: Hsp110 synergizes with Hsp70 and Hsp40 to catalyze protein disaggregation and reactivation in a cell-free system. *PLoS One*. 2011;6(10):e26319. doi: 10.1371/journal.pone.0026319.

Sirur A1, Best RB. (2013) Effects of interactions with the GroEL cavity on protein folding rates. *Biophys J.* 2013 Mar 5;104(5):1098-106. doi: 10.1016/j.bpj.2013.01.034.

Singh, J., and Thornton, J.M. (1985). The interaction between phenylalanine rings in proteins. *FEBS Lett. 191*, 1-6.

Smith DL, Deng Y, Zhang Z. (1997) Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J Mass Spectrom.* Feb;32(2):135-46.

Stagg L, Zhang SQ, Cheung MS, Wittung-Stafshede P. (2007) Molecular crowding enhances native structure and stability of alpha/beta protein flavodoxin. *Proc Natl Acad Sci U S A.* Nov 27;104(48):18976-81.

Stan G, Thirumalai D, Lorimer GH, Brooks BR. (2003) Annealing function of GroEL: structural and bioinformatic analysis. *Biophys Chem*. 2003;100(1-3):453-67.

Tang, Y.-C., Chang, H.-C., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, Michael J, et al. (2006). Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell*, *125*(5), 903-14. doi: 10.1016/j.cell.2006.04.027.

Tang YC, Chang HC, Hayer-Hartl M, Hartl FU. (2007) SnapShot: molecular chaperones, Part II. *Cell*. 2007 Jan 26;128(2):412.

Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. (2007) Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci.* 2007 May;32(5):204-6.

Tartaglia GG, Vendruscolo M. (2010) Proteome-level interplay between folding and aggregation propensities of proteins. *J Mol Biol.* 2010 Oct 8;402(5):919-28. doi: 10.1016/j.jmb.2010.08.013.

Tyagi NK, Fenton WA, Deniz AA, Horwich AL. (2011) Double mutant MBP refolds at same rate in free solution as inside the GroEL/GroES chaperonin chamber when aggregation in free solution is prevented. *FEBS Lett.* 2011 Jun 23;585(12):1969-72. doi: 10.1016/j.febslet.2011.05.031.

Ugrinov KG, Clark PL. (2010) Cotranslational folding increases GFP folding yield. *Biophys J.* Apr 7;98(7):1312-20. doi: 10.1016/j.bpj.2009.12.4291.

Ullers RS, Houben EN, Raine A, ten Hagen-Jongman CM, Ehrenberg M, Brunner J, Oudega B, Harms N, Luirink J. (2003) Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the Escherichia coli ribosome. *J Cell Biol.* May 26;161(4):679-84.

Urfer R, Kirschner K. (1992) The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. *Protein Sci.* 1992 Jan;1(1):31-45.

Vagin, A.A., and Isupov, M.N. (2001). Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 57, 1451-1456.

van den Berg, B., Ellis, R. J., and Dobson, C. M. (1999). Effects of macromolecular crowding on protein folding and aggregation. *EMBO J* 18, 6927-6933.

van der Vies SM, Viitanen PV, Gatenby AA, Lorimer GH, Jaenicke R. (1992) Conformational states of ribulosebisphosphate carboxylase and their interaction with chaperonin 60. *Biochemistry*. 1992 Apr 14;31(14):3635-44.

Vasconcelos AT1, Ferreira HB, Bizarro CV, Bonatto SL, Carvalho MO, Pinto PM, Almeida DF, Almeida LG, Almeida R, Alves-Filho L, Assunção EN, Azevedo VA, Bogo MR, Brigido MM, Brocchi M, Burity HA, Camargo AA, Camargo SS, Carepo MS, Carraro DM, de Mattos Cascardo JC, Castro LA, Cavalcanti G, Chemale G, Collevatti RG, Cunha CW, Dallagiovanna B, Dambrós BP, Dellagostin OA, Falcão C, Fantinatti-Garboggini F, Felipe MS, Fiorentin L, Franco GR, Freitas NS, Frías D, Grangeiro TB, Grisard EC, Guimarães CT, Hungria M, Jardim SN, Krieger MA, Laurino JP, Lima LF, Lopes MI, Loreto EL, Madeira HM, Manfio GP, Maranhão AQ, Martinkovics CT, Medeiros SR, Moreira MA, Neiva M, Ramalho-Neto CE, Nicolás MF, Oliveira SC, Paixão RF, Pedrosa FO, Pena SD, Pereira M, Pereira-Ferrari L, Piffer I, Pinto LS, Potrich DP, Salim AC, Santos FR, Schmitt R, Schneider MP, Schrank A, Schrank IS, Schuck AF, Seuanez HN, Silva DW, Silva R, Silva SC, Soares CM, Souza KR, Souza RC, Staats CC, Steffens MB, Teixeira SM, Urmenyi TP, Vainstein MH, Zuccherato LW, Simpson AJ, Zaha A. (2005) Swine and poultry pathogens: the complete genome sequences of two strains of Mycoplasma hyopneumoniae and a strain of Mycoplasma synoviae. *J Bacteriol*. 2005 Aug;187(16):5568-77.

Viitanen PV, Lubben TH, Reed J, Goloubinoff P, O'Keefe DP, Lorimer GH. (1990) Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (groEL) are K+ dependent. *Biochemistry*. 1990 Jun 19;29(24):5665-71.

Wales, T. E., & Engen, J. R. (2006). Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spec Rev, 25*(1), 158-170.

Wales, T.E., Fadgen, K.E., Gerhardt, G.C., and Engen, J.R. (2008). High-speed and high-resolution UPLC separation at zero degrees Celsius. *Anal. Chem.* 80, 6815-6820.

Walters BT, Mayne L, Hinshaw JR, Sosnick TR, Englander SW. (2013) Folding of a large protein at high structural resolution. *Proc Natl Acad Sci U S A.* 2013 Nov 19;110(47):18898-903. doi: 10.1073/pnas.1319482110.

Wang L, Lane LC, Smith DL. (2001) Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry. *Protein Sci.* Jun;10(6):1234-43.

Weber F, Hayer-Hartl M. (2000) Refolding of bovine mitochondrial rhodanese by chaperonins GroEL and GroES. *Methods Mol Biol.* 2000;140:117-26.

Weis DD, Engen JR, Kass IJ. (2006) Semi-automated data processing of hydrogen exchange mass spectra using HX-Express. *J Am Soc Mass Spectrom.* Dec;17(12):1700-3.

Weissman, J. S., Rye, H S, Fenton, W A, & And A L Horwich, J. M. B. (1996). Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell*, *84*(3), 481-490. References 135

Wirth AJ, Platkov M, Gruebele M. (2013) Temporal variation of a protein folding energy landscape in the cell. *J Am Chem Soc*. Dec 26;135(51):19215-21. doi: 10.1021/ja4087165.

Xu, Z., Horwich, A L, & Sigler, P B. (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature*, *388*(6644), 741-50. doi: 10.1038/41944.

Yang Z, Májek P, Bahar I. (2009) Allosteric transitions of supramolecular systems explored by network models: application to chaperonin GroEL. *PLoS Comput Biol.* 2009 Apr;5(4):e1000360. doi: 10.1371/journal.pcbi.1000360.

Yébenes H, Mesa P, Muñoz IG, Montoya G, Valpuesta JM. (2011) Chaperonins: two rings for folding. *Trends Biochem Sci.* 2011 Aug;36(8):424-32. doi: 10.1016/j.tibs.2011.05.003

Yifrach, O., & Horovitz, A. (1995). Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. *Biochemistry*, *34*(16), 5303-5308.

Zahn R, Buckle AM, Perrett S, Johnson CM, Corrales FJ, Golbik R, Fersht AR. (1996) Chaperone activity and structure of monomeric polypeptide binding domains of GroEL. *Proc Natl Acad Sci U S A.* 1996 Dec 24;93(26):15024-9.

Zhang, Z., and Smith, D.L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci. 2*, 522-531.

Zhang J1, Adrián FJ, Jahnke W, Cowan-Jacob SW, Li AG, Iacob RE, Sim T, Powers J, Dierks C, Sun F, Guo GR, Ding Q, Okram B, Choi Y, Wojciechowski A, Deng X, Liu G, Fendrich G, Strauss A, Vajpai N, Grzesiek S, Tuntland T, Liu Y, Bursulaya B, Azam M, Manley PW, Engen JR, Daley GQ, Warmuth M, Gray NS. (2010) Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature*. 2010 Jan 28;463(7280):501-6. doi: 10.1038/nature08675.

Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E., et al. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science*, *272*(5268), 1606-14.

Zimmerman, S. B., & Minton, A P. (1993). Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu Rev Biophys Biomol Str*, *22*(1), 27-6