MECHANISM OF AGGREGATE REACTIVATION BY THE MOLECULAR CHAPERONE CLPB

by

TING ZHANG

B.S., Zhejiang University, Hangzhou, China, 2006

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Graduate Biochemistry Group
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Abstract

ClpB, a bacterial chaperone that belongs to the AAA+ protein family, cooperates with the Hsp70/40 system (DnaK, DnaJ and GrpE in *E.coli*) in the reactivation of aggregated substrates by translocating them through the central channel of its hexameric form. ClpB is essential for survival of bacteria under heat shock and plays an important role in the infectivity of pathogenic microorganisms. However the detailed mechanism of ClpB disaggregation activity is still not clear.

ClpB is a multi-domain protein, which consists of two nucleotide binding domains (NBD1 and NBD2) connected by the middle domain (M domain), and the N-terminal domain connected to the rest of the protein by a flexible linker. In this work, mutations were introduced into the linker region to modify the mobility of the N-terminal domain. It was found that without altering the proper folding and oligomerization of ClpB, all the mutants had deficiencies in aggregate reactivation, possibly due to the weaker binding to aggregated substrates in the initial step of disaggregation. This led to the conclusion that the flexible attachment of the N-terminal domain supports substrate binding and controls the disaggregation by ClpB. Moreover, partial inhibition of the ClpB chaperone activity was observed for all the linker variants, suggesting that the linker sequence might have been optimized by selective pressure to maintain the optimal efficiency of aggregate reactivation.

To study the substrate translocation of ClpB, a BAP (ClpB-ClpA P-loop) variant that binds to the protease ClpP was constructed. A FRET-based experiment was designed and the fluorescently-labeled ClpB substrates were produced. This work sets the stage for further studies on the mechanism of aggregate recognition by ClpB.

ClpB also plays important roles in pathogenic bacteria invasion and virulence. Recombinant ClpB from *Ehrlichia chaffeensis*, a pathogenic bacterium that causes human monocytic ehrlichiosis, was purified to study its biochemical properties. *Ehrlichia* ClpB (Eh_B) and *E.coli* ClpB (Ec_B) sequences are highly conserved in the nucleotide binding region and poorly conserved in the N-terminal and M domain. The

oligomerization, ATPase activity, chaperone activity and substrate binding of the recombinant Eh_B were tested. Recombinant Eh_B was able to reactivate aggregated proteins in the presence of HSP70 from *E.coli* with equal efficiency as Ec_B. However, the mechanism of Eh_B interactions with substrates and/or substrate specificity may be different from that of *E. coli* ClpB.

MECHANISM OF AGGREGATE REACTIVATION BY THE MOLECULAR CHAPERONE CLPB

by

TING ZHANG

B.S., Zhejiang University, Hangzhou, China, 2006

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Graduate Biochemistry Group
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Approved by:

Major Professor
Michal Zolkiewski

Copyright

Abstract

ClpB, a bacterial chaperone that belongs to the AAA+ protein family, cooperates with the Hsp70/40 system (DnaK, DnaJ and GrpE in *E.coli*) in the reactivation of aggregated substrates by translocating them through the central channel of its hexameric form. ClpB is essential for survival of bacteria under heat shock and plays an important role in the infectivity of pathogenic microorganisms. However the detailed mechanism of ClpB disaggregation activity is still not clear.

ClpB is a multi-domain protein, which consists of two nucleotide binding domains (NBD1 and NBD2) connected by the middle domain (M domain), and the N-terminal domain connected to the rest of the protein by a flexible linker. In this work, mutations were introduced into the linker region to modify the mobility of the N-terminal domain. It was found that without altering the proper folding and oligomerization of ClpB, all the mutants had deficiencies in aggregate reactivation, possibly due to the weaker binding to aggregated substrates in the initial step of disaggregation. This led to the conclusion that the flexible attachment of the N-terminal domain supports substrate binding and controls the disaggregation by ClpB. Moreover, partial inhibition of the ClpB chaperone activity was observed for all the linker variants, suggesting that the linker sequence might have been optimized by selective pressure to maintain the optimal efficiency of aggregate reactivation.

To study the substrate translocation of ClpB, a BAP (ClpB-ClpA P-loop) variant that binds to the protease ClpP was constructed. A FRET-based experiment was designed and the fluorescently-labeled ClpB substrates were produced. This work sets the stage for further studies on the mechanism of aggregate recognition by ClpB.

ClpB also plays important roles in pathogenic bacteria invasion and virulence. Recombinant ClpB from *Ehrlichia chaffeensis*, a pathogenic bacterium that causes human monocytic ehrlichiosis, was purified to study its biochemical properties. *Ehrlichia* ClpB (Eh_B) and *E.coli* ClpB (Ec_B) sequences are highly conserved in the nucleotide binding region and poorly conserved in the N-terminal and M domain. The

oligomerization, ATPase activity, chaperone activity and substrate binding of the recombinant Eh_B were tested. Recombinant Eh_B was able to reactivate aggregated proteins in the presence of HSP70 from *E.coli* with equal efficiency as Ec_B. However, the mechanism of Eh_B interactions with substrates and/or substrate specificity may be different from that of *E. coli* ClpB.

Table of Contents

List of Figures

Acknowledgements

First of all, I would like to express my sincere gratitude to my advisor, Dr Michal Zolkiewski, for his encouragement and guidance during my gradate study. This thesis would not be completed without his insightful criticism and patient instruction. I thank him for helping through the difficulties with great patience and for showing me how to become a scientist.

I would like to thank my committee members, Dr Jianhan Chen, Dr Rollie J Clem, Dr Subbarat Muthukrishnan, and Dr Tiruvoor G Nagaraja, for their suggestions during my PhD study, and for putting their considerable time and effort into the valuable comments on this thesis.

I thank the former and current members of the Zolkiewski lab, Dr Maria Nagy, Huichuan (Arleen) Wu, Yan Wu, Sam Molina, and Fabrice Ngansop for their help and friendship. Special thanks to Dr Maria Nagy, who taught me so much when I started to work in the lab and has been an effective teacher ever since.

I would like to thank all the faculties and graduate students of the Biochemistry Department at Kansas State University, for their assistance and friendship.

Finally, I would like to thank my friends and family. I thank Danqiong Sun, Ying Ding, Xiaoxi Zhu, Huaien Dai, Xiangming Li, Chen Lin, Sheng Yi, and Chenchen Wang for their help and friendship. I express my deepest gratitude for my family for their love, support and encouragement though this wonderful journey.

Dedication

I dedicate this work to my beloved parents, Xiujiang Zhang and Suping Zhao, for their love and support.

Chapter 1

Introduction

1.1 Protein folding and proteostasis

Proteins are delicate macromolecules that carry out most biological functions in a living cell. A newly synthesized linear polypeptide chain needs to be correctly folded into its "native" conformation in order to be biologically functional (Bartlett and Radford 2009; Hartl, Bracher et al. 2011). In theory, a polypeptide can have countless different conformations and the thermodynamic hypothesis of protein folding states that the native conformation is the one with minimum free energy. This hypothesis was enunciated by the folding of ribonuclease A *in vitro* (Anfinsen 1973) and remains popular today.

For proteins with more than 100 amino acid residues (~90% of all proteins), partially folded intermediates form during the protein folding process (Hartl, Bracher et al. 2011). An important force that drives polypeptide chain to collapse and bury the non-polar amino residues inside the tertiary structure is the hydrophobic force (Skach 2009). Thus in the intermediate state, a lot of the hydrophobic residues could be exposed and misfolding commonly happens. These exposed hydrophobic residues also tend to interact with each other and form aggregates. Cells have evolved strategies like network of chaperones to facilitate folding and to remove misfolded proteins and aggregates in order to gain protein quality control and to maintain proteostasis (Fig. 1). However if the misfolded proteins and aggregates accumulate, for example when cell is under stress, they may overwhelm the proteostasis capacity and cell death will eventually happen.

1.2 Aggregation related diseases

The proteome of a cell is a highly conserved biological network that keeps the cell functioning properly. Mutations and environmental stress can lead to protein

misfolding and cause related diseases. For example, mutations of lysozyme make its folding in the ER inefficient, causing lysosomal storage diseases (LSDs) (Sawkar, D'Haeze et al. 2006); another famous example is cystic fibrosis which is caused by misfolding of the mutated cystic fibrosis transmembrane conductance regulator (CFTR) (Denning, Anderson et al. 1992). However, there is a group of protein misfolding related diseases that draw the most attention: amyloid deposition diseases (the term "amyloidosis" is often used).

Amyloid fibrils are defined as polypeptide aggregates with a cross-β conformation (Fandrich 2007). They arise from normally soluble proteins or protein fragments and accumulate in human organs (Dobson 1999). There are about 20 diseases in this category including Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis (ALS), type II diabetes mellitus and prion related diseases. It is believed that among these diseases, the neurodegenerative ones are caused by interaction between the aggregates and normal functioning components, usually in the brain; whereas the non-neurodegenerative ones are widely believed to be the result of large amount of accumulation (even kilograms) of aggregates in the organ (Stefani and Dobson 2003). The initial trigger of these diseases is abnormal folding or aggregation of mutated protein, followed by environmental factors such as Ca^{2+} and oxidative stress (Zerovnik, Stoka et al. 2011).

The mechanism of amyloid fibril formation is still unknown. Although all amyloid fibrils share similar architectures, their precursors lack similarity in sequence or structure (Stefani 2010). There have been several different models proposed: templating and nucleation models, linear colloid-like assembly of spherical oligomers and domain swapping models. In all these models, amyloid fibrils start from a usually soluble nucleation unit formation first and oligomers of that unit are built up into long fibrils. The rate limiting step is always the formation of ordered structure (Zerovnik, Stoka et al. 2011). These unstable, oligomeric nucleation units, or pre-fibrillar aggregates, are considered most toxic to cells. Actually, the very stable and inert mature fibrils have the potential to protect cells from the attack of those precursors

(Stefani 2010). This explains why there is no obvious correlation between Alzheimer's' disease and the size of fibril plaque (Dickson, Crystal et al. 1995).

Many prion related diseases also include the formation of amyloid like aggregates. Prions are inheritable, infectious protein agents that cause many life threatening neurological diseases such as Creutzfeldt–Jakob disease in humans and bovine spongiform encephalopathy (also known as mad cow disease) in cattle (Colby and Prusiner 2011). Prions exist both in mammals and fungi. In mammal prion propagation, normal prion protein PrP(C) converts into the disease causing isoform PrP(Sc) and PrP(Sc) continuously recruits PrP(C) and converts it into the β rich PrP(Sc), and the mechanism is more likely to be a template-assisted replication (Colby and Prusiner 2011). For yeast prions, seeded polymerization similar to amyloid fibrils formation with the help of protein chaperones is favored (Wickner 1994; Speransky, Taylor et al. 2001).

1.3 Features of protein aggregation

Protein aggregation has also been a problem in many biological studies such as protein over expression and protein folding *in vitro*. Protein aggregation can be classified into *in vivo* and *in vitro*, and ordered and disordered (Fink 1998). Protein aggregation *in vitro* impairs protein stability; and protein aggregation *in vivo* is responsible for the formation of inclusion bodies (disordered aggregates) and amyloid fibrils (ordered aggregates). Disordered protein aggregation produced during protein folding such as inclusion bodies, thermal aggregates and refolding aggregates in *in vitro* experiments arise from partially folded intermediates, whereas amyloid fibrils arise from the polymerization of native-like conformations (Wetzel 1994).

For amorphous aggregates, hydrophobic surfaces of the intermediates interact with each other and form oligomers; eventually the size of the oligomers becomes too large for solubility and aggregates form. This mechanism has been demonstrated in many cases of protein aggregation (Speed, Wang et al. 1995; Kim and Yu 1996; Liu and Wang 2010).

For amyloids, people used to believe that only amyloidoses (amyloid fibrils and their precursors) with particular conformational characters are responsible until several proteins that are not related to amyloidoses were found to have the potential to form amyloid fibrils *in vitro* (Dobson 1999). Despite their different origins, all amyloid fibrils share a β-sheet-rich core with a supramolecular organization (Stefani 2010). The ability to form amyloid aggregates seems to lie in the sequence of any polypeptides, that is to say, it is generic (Dobson 2003), although the rates of aggregation still correlate with the physicochemical properties of the protein such as charge, secondary structure and hydrophobicity (Chiti, Stefani et al. 2003). If any protein has the potential to form amyloid aggregates, then cells must have developed certain strategies to keep proteins in their native structure for proper function during evolution. Actually, it is thought that preventing the formation of amyloid like aggregates is a driving force of protein evolution (Monsellier and Chiti 2007). Many strategies such as correctly folding, limiting β-propensity, hydrophobicity and low net charge, and protein quality control have been developed to maintain the proteostasis of a living cell (Monsellier and Chiti 2007).

1.4 Protein aggregation prediction

The propensity to form ordered β sheet aggregates lies in every protein (Dobson 2003). However, the primary sequence of a protein still is a key factor to modulate protein aggregation (Ventura 2005). Prediction of protein aggregation is a useful tool to improve recombinant protein production in bacteria and could provide information for therapeutical targets of anti-amyloid drugs.

Many computational methods have been developed to predict the propensity of aggregation based on a protein's sequence. Many of them rely on the β sheet conformation possibility in the sequence: the sequences that appear to be α helical but have the propensity to form β sheet; the sequences that are highly predicted to be both α helix and β sheet; and repeated β sheet stack in structure (Yoon and Welsh 2004; Hamodrakas, Liappa et al. 2007; Zhang, Chen et al. 2007). There has also been

investigation of short amino acid stretches known as “hot spots” that are aggregation prone (Ventura, Zurdo et al. 2004; Sanchez de Groot, Pallares et al. 2005). For example, mutation studies have found that amino acid residues that reduce the net charge of a protein are highly aggregation prone (Chiti, Calamai et al. 2002; Chiti, Taddei et al. 2002). Many computational algorithms are multi-parameter based.

The following is a list of popular aggregation prediction programs (Hamodrakas 2011):

TANGO: http://tango.crg.es/

PASTA: http://protein.cribi.unipd.it/pasta/

AGGRESCAN: http://bioinf.uab.es/aggrescan/

ZYGGREGATOR: http://www-vendruscolo.ch.cam.ac.uk/zyggregator_test.php

AMYLPRED: http://biophysics.biol.uoa.gr/AMYLPRED/

PAFIG: http://www.mobioinfor.cn/pafig/

NETCSSP: http://cssp2.sookmyung.ac.kr/

BETASCAN: http://groups.csail.mit.edu/cb/betascan/

FOLDAMYLOID: http://antares.protres.ru/fold-amyloid/oga.cgi

WALTZ: http://waltz.switchlab.org/

PRE-AMYL: ftp://mdl.ipc.pku.edu.cn/pub/software/pre-amyl

1.5 Chaperones

Newly synthesized polypeptides need to fold into proper 3D structures to gain function. However in the cell it is highly possible that a polypeptide misfolds or aggregates. In order to maintain the proteostasis of the cellular environment, cells have evolved a complex network of chaperones to facilitate protein folding, unfold misfolded proteins and disaggregate protein aggregation.

Protein misfolding and aggregation are triggered under stress. Since heat is a universal stress factor, many chaperones respond to heat stress. There are a group of highly conserved chaperone proteins called heat shock proteins (Hsps) that are up-regulated under heat-shock. Heat shock proteins are universal. They exist in every

organism from archebacteria to mammals and in all types of cells and tissues (Lindquist 1986). Major members of Hsps include Hsp100, Hsp90, Hsp70, Hsp60 (chaperonins), Hsp40 and small Hsps (typically 20 to 25 kDa), sorted according to their approximate molecular weight (Smith, Whitesell et al. 1998). These chaperones are involved in many different processes such as *de novo* protein folding, refolding of denatured proteins, protein trafficking and proteolytic degradation (Hartl, Bracher et al. 2011). Most chaperones bind to hydrophobic regions in a misfolded protein (Fink 1999).

The classic chaperonin GroEL-GroES in *E. coli* is the one of the best described chaperones. The whole machinery of GroEL is composed of a stack of rings of subunits and binds to the exposed hydrophobic surface of nonnative proteins in the binding-active state to facilitate protein folding in the central cavity. The folding-active state of GroEL is achieved by the binding of ATP and a lid like co-chaperone GroES (Bukau and Horwich 1998). The native form of a protein can be achieved by several rounds of binding, refolding and release of GroEL-GroES chaperonin (Horwich 2002). Since GroEL-GroES recognizes hydrophobic region through the hydrophobic walls of the cavity, it competes with the binding between misfolded proteins and prevents aggregation (Grantcharova, Alm et al. 2001).

The Hsp70s are ubiquitous chaperones involved in a wide range of protein folding and refolding functions. Their function is illustrated by the *E. coli* Hsp70 protein DnaK. DnaK contains an N-terminal ATPase domain and a C-terminal substrate binding domain. DnaK binds to substrates in its ADP bound state and releases them in the ATP bound state. The whole DnaK machinery works as a "holdase": it captures the substrate and releases it during the ATP hydrolysis cycle. DnaK never works alone. The whole machinery always consists of co-chaperones such as Hsp40 DnaJ and nucleotide exchange factor GrpE. DnaJ stimulates ATP hydrolysis, facilitating substrate binding, whereas GrpE stimulates the disassociation of ADP, facilitating substrate release. ATP cycling controls the opening and closing of the substrate binding domain, and the ATPase domain and the substrate binding

domain are coupled (Bukau and Horwich 1998; Mayer and Bukau 2005; Kampinga and Craig 2010).

Another important function of the Hsp70 family is their cooperation with Hsp100 family in protein disaggregation. Hsp104 is a member of Hsp100 family from yeast. It is required for thermo-tolerance of yeast (Sanchez and Lindquist 1990; Sanchez, Taulien et al. 1992). Lindquist and co-workers found the disaggregation of luciferase by Hsp104 *in vivo* (Parsell, Kowal et al. 1994) and later disaggregation *in vitro* by Hsp104 and co-chaperone Ydj1 and Ssa1 was proved (Glover and Lindquist 1998). In 1999, the same bichaperone machinery was found in Hsp104 homolog ClpB/DnaK/DnaJ/GrpE in *E. coli* (Goloubinoff, Mogk et al. 1999; Motohashi, Watanabe et al. 1999; Zolkiewski 1999). These bichaperone networks disaggregate seemingly irreversible aggregates and their mechanism is still unclear today. Notably this remarkable disaggregation machinery only exists in bacteria, fungi and plants, not in animals (Sanchez, Taulien et al. 1992).

Besides protein folding and refolding chaperones, another important member of the chaperone network is the proteolytic system including ClpXP, ClpAP, ClpCP, HslUV, Lon, FtsH, PAN/20S, and the 26S proteasome (Sauer and Baker 2011). Instead of refolding nonnative proteins into their native conformation, these machines simply degrade them for amino acid recycling.

1.6 AAA+ super family

The proteolytic machines and the disaggregase Hsp104 and ClpB all belong to the AAA+ super family of ATPases (AAA stands for ATPases associated with various cellular activities) (Neuwald, Aravind et al. 1999). The members of AAA+ family share the common character of utilizing the energy from ATP hydrolysis to carry out functions such as protein folding and refolding, membrane fusion, DNA replication and transcription, assembly of DNA-protein complex and organelle biogenesis (Ogura and Wilkinson 2001). Most of them form ring-shaped oligomers (mostly hexamers) with a central channel (Sauer and Baker 2011).

All AAA+ proteins have conserved AAA domain (nucleotide binding domain) and according to the number of AAA domains they can be classified as class I (two AAA domains) or class II (one AAA domain) (Neuwald, Aravind et al. 1999). The AAA module consists of two sub-domains: an N terminal domain with one α/β fold and nucleotide binding pocket; and a C-terminal domain that is mainly α-helical. There are several conserved motifs in the AAA module: Walker A motif (GxxxxGKT, x=any residue) which forms a loop (the P-loop), Walker B motif (hhhhDExx, h=hydrophobic residue), sensor 1 (T/N) and sensor-2 (R/K) motifs. Lysine in the P-loop of Walker A and hydrophobic residues in Walker B are crucial for ATPase activity (Iyer, Leipe et al. 2004; Hanson and Whiteheart 2005). Other domains may vary to adapt to various functions (Ogura and Wilkinson 2001).

1.7 Structural and functional information about ClpB

The unique disaggregation function of ClpB has attracted much attention. The crystal structure of ClpB from *Thermus thermophilus* (Lee, Sowa et al. 2003) has been solved and provides a useful tool to better understand the function of ClpB. However, knowing a protein's structure does not necessarily reveal everything about its function (Skolnick and Fetrow 2000). Thus, numerous structure-function studies of each domain of ClpB have been done in recent years.

The ClpB monomer is about 95kDa and consists of two NBD domains separated by a middle domain, a C-terminal domain and an N-terminal domain connected to the rest part of the protein by a 17 amino acid long flexible linker (Fig. 2). ClpB forms hexamers, or heptamer at high concentration. The oligomer is stabilized by the presence of ATP (Akoev, Gogol et al. 2004). Cryo-EM study showed that ClpB hexamer has a two-tiered ring structure with the M domain reaching out as propellers (Lee, Sowa et al. 2003). Homology modeling of *E. coli* ClpB (unpublished data by Dr Maria Nagy, Yale University) showed that *E. coli* ClpB has a very similar structure to *Thermus thermophilus* ClpB. The N-terminal domain is on the top of the hexamer and M domain is at the outer surface. There is a narrow channel with conserved pore loops

protruding into the channel (see Fig. 2).

1.7.1 Mechanism of ClpB/KJE machinery

ClpA, ClpC, ClpX, and HsIU are the closest relatives of ClpB in the AAA+ family (Iyer, Leipe et al. 2004). The mechanism of ClpAP and ClpXP proteolytic machines has been well established. Both ClpA and ClpX translocate their substrates through the central channel of their oligomers into the chamber of protease ClpP (Kim, Burton et al. 2000; Reid, Fenton et al. 2001). A similar mechanism for ClpB was found by constructing BAP, a mutated ClpB that could bind to ClpP (Weibezahn, Tessarz et al. 2004). As shown in Fig. 3, aggregates bind to ClpB hexamers in the presence of ATP. However before the translocation begins, there is an additional step to make a conformational change of substrate and ClpB to convert the substrate binding from uncommitted state to committed state. Co-chaperone KJE and the M domain may be involved in this commitment step to facilitate interaction of aggregates and the pore loop of NBD1 (Haslberger, Weibezahn et al. 2007). After substrate is committed, it is translocated through the central channel of ClpB using the energy of ATP hydrolysis and becomes unfolded. When the aggregates are too difficult to unfold, the dynamic formation of ClpB hexamer allows it to give up and try another position to pull.

1.7.2 N-terminal domain of ClpB

The N-terminal domain of ClpB resides on top of the ClpB hexamer and is relatively more mobile comparing to the rest of the protein (Lee, Sowa et al. 2003). The N-terminal domain contains two repeated domains of about 75 amino acid residues (Lo, Baker et al. 2001). Several studies have shown that the N-terminal domain in the ClpB analog Hsp104 is dispensable in protein disaggregation of some substrates *in vitro* and *in vivo* (Hung and Masison 2006; Sielaff and Tsai 2010). Also in *E. coli*, ClpB has two transcription isoforms: the full length ClpB95 and the N-terminal truncated ClpB80 (Woo, Kim et al. 1992). The N-terminal domain seems

inconsequential but why has it not been discarded during evolution? And what role is it playing in disaggregation?

It was found that deletion or mutation of the N-terminal domain inhibited the activation of ATPase of ClpB by α-casein (Park, Kim et al. 1993; Barnett, Zolkiewska et al. 2000; Liu, Tek et al. 2002) and impair disaggregation, especially mutations on the surface of the N-terminal domain (Liu, Tek et al. 2002; Tanaka, Tani et al. 2004). *In vivo* studies also found that the truncated isoform ClpB80 was less efficient to deal with protein mis-folding and ClpB80 function was impaired while DnaK was deleted (Chow, Barnett et al. 2005). All these data suggest the importance of the N-terminal domain. A recent study proved that the N-terminal domain supported substrate binding with large aggregates and several amino acid residues in the N-terminal domain were found to be important for substrate binding (Barnett, Nagy et al. 2005), consistent with the structural study of the N-terminal domain that there was a putative peptide binding groove in the N-terminal domain (Li and Sha 2003) . Considering the mobility of the N-terminal domain and its relationship with DnaK, it is possible that the N-terminal domain interacts with substrates and, together with DnaK, helps the insertion of substrate into the central channel of ClpB. The N-terminal domain may also play a role in substrate specificity of ClpB considering the fact that it is only essential for the reactivation of selected substrates.

Since the N-terminal domain of ClpB is important for biochemical function, what is the purpose of the truncated isoform ClpB80? Actually there is a synergistic cooperation between the two isoforms: interaction of two isoforms activated disaggregation activity (Nagy, Guenther et al. 2010). This cooperation is also essential in vivo for *E. coli* to survive under heat shock (Chow and Baneyx 2005). The mechanism of this cooperation is still unclear, although it is not coming from a faster ATP hydrolysis rate or more efficient substrate binding (Nagy, Guenther et al. 2010). The N-terminal domain is attached to the rest of ClpB by an unstructured and flexible linker. It is possible that in the hetero-oligomer of ClpB95/ClpB80, the N-terminal domain has more space to move and this increase in mobility may result in superior

chaperone activity.

All in all, the N-terminal domain supports aggregate binding to ClpB and its mobility may also be important for protein disaggregation.

1.7.3 M-domain of ClpB

The M domain of ClpB is located between two nucleotide binding domains and reaches out in the ClpB hexamer like a propeller (Fig. 2). It is a long coiled-coil with four helices. The presence of the M domain stabilizes ClpB oligomers (del Castillo, Alfonso et al. 2011) and is also essential for the disaggregation machinery of ClpB. Deletion of the M domain will abolish reactivation of aggregates by ClpB (Kedzierska, Akoev et al. 2003; Mogk, Schlieker et al. 2003).

The M domain is also mobile, although less than the N-terminal domain. In the crystal structure of *TClpB* in the AMPPNP bound state, the M domain can swing as much as 15 degrees in different conformations and this mobility of M domain is essential for ClpB disaggregation function (Lee, Sowa et al. 2003). Later studies showed that the mobility of the M domain is actually greater than that shown in the crystal structure, especially helix 3, which not only changes its position but also changes its conformation during ATPase cycle (Haslberger, Weibezahn et al. 2007; Lee, Choi et al. 2007). A full-atom hexameric model study of *E. coli* ClpB also showed that in the hexamer ring, M domains from different subunits were in different orientations (Zietkiewicz, Slusarz et al. 2010), again proving the flexibility and plasticity of the M domain.

The precise function of the M domain is still unclear. One early assumption is that the propeller shape of the M domain in the hexamer can serve as crow bars and breaks aggregates into pieces (Glover and Lindquist 1998). This model is not favored because the crow bar model implies direct contact between aggregates and the M domain, which has not been found. Another possibility is that the M domain facilitates the interaction between KJE and ClpB, helping KJE to insert polypeptides into the central channel of ClpB in the initial step of substrate translocation (Weibezahn,

Tessarz et al. 2004; Zietkiewicz, Lewandowska et al. 2006). Failure in cooperation between mutated M domain and KJE abolished reactivation of aggregates (Haslberger, Weibezahn et al. 2007). Unexpectedly, insertion of T4 lysozyme into helix 2 of M domain allows Hsp104 to function without co-chaperone Hsp70 (Lee, Sielaff et al. 2010; Sielaff and Tsai 2010), which adds to the mystery of M domain/Hsp70 cooperation.

Recent studies have revealed another important feature of the M domain: it is responsible for the species specific cooperation with the Hsp70 system (Sielaff and Tsai 2010; Miot, Reidy et al. 2011). Switching the M domain of Hsp104 and ClpB abolished their collaboration with corresponding Hsp70 system. This species specificity relies on the helix 2 of the M domain (Miot, Reidy et al. 2011).

1.7.4 C-terminal domain of ClpB

The small C-terminal domain contains mainly α-helices and contacts with the second nucleotide binding domain. It is proposed to bind to some substrates and therefore is called the sensor and substrate discrimination (SSD) domains (Lo, Baker et al. 2001). The C-terminal domain is essential for oligomerization (Barnett, Zolkiewska et al. 2000; Mogk, Schlieker et al. 2003; Mackay, Helsen et al. 2008) and studies found that the C-terminal domain of another Hsp100 member HsIU could sense the nucleotide status of the ATPase core domain (Bochtler, Hartmann et al. 2000; Sousa, Trame et al. 2000). In HSP104, there is a long acidic C-terminal extension that is not present in ClpB which may be important for the functional differences of Hsp104 and ClpB (Mackay, Helsen et al. 2008).

1.7.5 NBD1 and NBD2

The two nucleotide binding domains are the core part of the ClpB machine. During substrate translocation, both NBD1 and NBD2 undergo a series of still unknown conformational changes to unfold aggregates. The translocation is possibly driven by the movement of pore loops in NBD1 and NBD2 (Lum, Tkach et al. 2004).

It has been proposed that the central channel of ClpB hexamer is so narrow that it only allows unfolded polypeptides to pass through (Schlieker, Tews et al. 2004).

The two NBDs are highly coupled. However it is still unclear exactly how the signal is transmitted through different parts of ClpB. The coupling is not covalent since mixing of isolated NBD1 with NBD2 showed WT chaperone activity. It has also been found that NBD2 has a high intrinsic ATPase activity that is inhibited by NBD1, suggesting a regulatory role of NBD1 (Beinker, Schlee et al. 2005). This is further shown by activation of NBD2 through the movement of M domain triggered by nucleotide binding to NBD1 (Watanabe, Takano et al. 2005). Also, the nucleotide binding property of NBD2 is controlled by NBD1 (Werbeck, Kellner et al. 2009).

To better understand the allosteric modulation of two NBD upon nucleotide binding, many experiments including mixture of active and inactive subunits of ClpB hexamer have been done. Study of mixed oligomers showed that ClpB has an intrinsic chaperone activity. When wild type ClpB was mixed with an inactive mutant, it gained chaperone activity in the absence of co-chaperones with some substrates; for other substrates, this chaperone activity was blocked in the presence of co-chaperone, indicating the need for cooperation between subunits of ClpB hexamer (Hoskins, Doyle et al. 2009). Actually not all 12 NBDs are needed for chaperone activity. As long as one ring in the hexamer contains an active nucleotide binding site, four intact subunits are sufficient to carry out chaperone function (del Castillo, Fernandez-Higuero et al. 2010).

1.8 The biological role of ClpB

The Hsp100 chaperones play important roles in different species. ClpB as well as its ortholog Hsp104 in yeast and Hsp101 in *Arabidopsis*, are essential for cell survival under heat shock (Squires, Pedersen et al. 1991; Sanchez, Taulien et al. 1992; Queitsch, Hong et al. 2000). Hsp104 plays a unique function in prion replication in yeast and replacement of hsp104 by bacterial ClpB *in vivo* abolished that function (Shorter and Lindquist 2004; Tipton, Verges et al. 2008).

In the past decade, evidence of ClpB involved in pathogenic microorganism virulence has been reported. ClpB is implicated in *Francisella tularensis* virulence (Grall, Livny et al. 2009) and inactivation of *clpb* gene in the pathogen *Leptospira interrogans* reduces virulence and resistance to stress conditions (Lourdault, Cerqueira et al. 2011). ClpB is also required for intracellular multiplication of *Staphylococcus aureus* within bovine mammary epithelial cells (Frees, Chastanet et al. 2004). ClpB is usually up-regulated under stress and makes pathogens resistant to treatment (Yukitake, Naito et al. 2011). Since ClpB does not exist in humans, it might be a good therapeutical target. For example, antibodies specific for ClpB in *Flavobacterium psychrophilum* may be important for protective immunity from bacterial coldwater disease (LaFrentz, LaPatra et al. 2011).

1.9 Goal of current study

The goal of the current study was to better understand the mechanism of disaggregation by ClpB/KJE. Different aspects related to this topic have been investigated: The role of the flexible linker that connects the N-terminal domain to the rest of the protein, substrate recognition and translocation, and the characterization of ClpB from the pathogenic bacterium *Ehrlichia chaffeensis*.

References

Akoev, V., E. P. Gogol, et al. (2004). "Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB." Protein Sci **13**(3): 567-574.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(96): 223-230.

Barnett, M. E., M. Nagy, et al. (2005). "The amino-terminal domain of ClpB supports binding to strongly aggregated proteins." J Biol Chem **280**(41): 34940-34945.

Barnett, M. E., A. Zolkiewska, et al. (2000). "Structure and activity of ClpB from Escherichia coli. Role of the amino-and -carboxyl-terminal domains." J Biol Chem **275**(48): 37565-37571.

Bartlett, A. I. and S. E. Radford (2009). "An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms." Nat Struct Mol Biol **16**(6): 582-588.

Beinker, P., S. Schlee, et al. (2005). "Biochemical coupling of the two nucleotide binding domains of ClpB: covalent linkage is not a prerequisite for chaperone activity." J Biol Chem **280**(45): 37965-37973.

Bochtler, M., C. Hartmann, et al. (2000). "The structures of HsIU and the ATP-dependent protease HsIU-HsIV." Nature **403**(6771): 800-805.

Bukau, B. and A. L. Horwich (1998). "The Hsp70 and Hsp60 chaperone machines." Cell **92**(3): 351-366.

Chiti, F., M. Calamai, et al. (2002). "Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases." Proc Natl Acad Sci U S A **99 Suppl 4**: 16419-16426.

Chiti, F., M. Stefani, et al. (2003). "Rationalization of the effects of mutations on peptide and protein aggregation rates." Nature **424**(6950): 805-808.

Chiti, F., N. Taddei, et al. (2002). "Kinetic partitioning of protein folding and aggregation." Nat Struct Biol **9**(2): 137-143.

Chow, I. T. and F. Baneyx (2005). "Coordinated synthesis of the two ClpB isoforms improves the ability of Escherichia coli to survive thermal stress." FEBS Lett **579**(20): 4235-4241.

Chow, I. T., M. E. Barnett, et al. (2005). "The N-terminal domain of Escherichia coli ClpB enhances chaperone function." FEBS Lett **579**(20): 4242-4248.

Colby, D. W. and S. B. Prusiner (2011). "De novo generation of prion strains." Nat Rev Microbiol **9**(11): 771-777.

Colby, D. W. and S. B. Prusiner (2011). "Prions." Cold Spring Harb Perspect Biol **3**(1): a006833.

del Castillo, U., C. Alfonso, et al. (2011). "A quantitative analysis of the effect of nucleotides and the M domain on the association equilibrium of ClpB." Biochemistry **50**(12): 1991-2003.

del Castillo, U., J. A. Fernandez-Higuero, et al. (2010). "Nucleotide utilization requirements that render ClpB active as a chaperone." FEBS Lett **584**(5): 929-934.

Denning, G. M., M. P. Anderson, et al. (1992). "Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive." Nature **358**(6389): 761-764.

Dickson, D. W., H. A. Crystal, et al. (1995). "Correlations of synaptic and pathological markers with cognition of the elderly." Neurobiol Aging **16**(3): 285-298; discussion 298-304.

Dobson, C. M. (1999). "Protein misfolding, evolution and disease." Trends Biochem Sci **24**(9): 329-332.

Dobson, C. M. (2003). "Protein folding and misfolding." Nature **426**(6968): 884-890.

Fandrich, M. (2007). "On the structural definition of amyloid fibrils and other polypeptide aggregates." Cell Mol Life Sci **64**(16): 2066-2078.

Fink, A. L. (1998). "Protein aggregation: folding aggregates, inclusion bodies and amyloid." Fold Des **3**(1): R9-23.

Fink, A. L. (1999). "Chaperone-mediated protein folding." Physiol Rev **79**(2):

425-449.

Frees, D., A. Chastanet, et al. (2004). "Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in Staphylococcus aureus." Mol Microbiol **54**(5): 1445-1462.

Glover, J. R. and S. Lindquist (1998). "Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins." Cell **94**(1): 73-82.

Goloubinoff, P., A. Mogk, et al. (1999). "Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network." Proc Natl Acad Sci U S A **96**(24): 13732-13737.

Grall, N., J. Livny, et al. (2009). "Pivotal role of the Francisella tularensis heat-shock sigma factor RpoH." Microbiology **155**(Pt 8): 2560-2572.

Grantcharova, V., E. J. Alm, et al. (2001). "Mechanisms of protein folding." Curr Opin Struct Biol **11**(1): 70-82.

Hamodrakas, S. J. (2011). "Protein aggregation and amyloid fibril formation prediction software from primary sequence: towards controlling the formation of bacterial inclusion bodies." FEBS J **278**(14): 2428-2435.

Hamodrakas, S. J., C. Liappa, et al. (2007). "Consensus prediction of amyloidogenic determinants in amyloid fibril-forming proteins." Int J Biol Macromol **41**(3): 295-300.

Hanson, P. I. and S. W. Whiteheart (2005). "AAA+ proteins: have engine, will work." Nat Rev Mol Cell Biol **6**(7): 519-529.

Hartl, F. U., A. Bracher, et al. (2011). "Molecular chaperones in protein folding and proteostasis." Nature **475**(7356): 324-332.

Haslberger, T., J. Weibezahn, et al. (2007). "M domains couple the ClpB threading motor with the DnaK chaperone activity." Mol Cell **25**(2): 247-260.

Horwich, A. (2002). "Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions." J Clin Invest **110**(9): 1221-1232.

Hoskins, J. R., S. M. Doyle, et al. (2009). "Coupling ATP utilization to protein

remodeling by ClpB, a hexameric AAA+ protein." Proc Natl Acad Sci U S A **106**(52): 22233-22238.

Hung, G. C. and D. C. Masison (2006). "N-terminal domain of yeast Hsp104 chaperone is dispensable for thermotolerance and prion propagation but necessary for curing prions by Hsp104 overexpression." Genetics **173**(2): 611-620.

Iyer, L. M., D. D. Leipe, et al. (2004). "Evolutionary history and higher order classification of AAA+ ATPases." J Struct Biol **146**(1-2): 11-31.

Kampinga, H. H. and E. A. Craig (2010). "The HSP70 chaperone machinery: J proteins as drivers of functional specificity." Nat Rev Mol Cell Biol **11**(8): 579-592.

Kedzierska, S., V. Akoev, et al. (2003). "Structure and function of the middle domain of ClpB from Escherichia coli." Biochemistry **42**(48): 14242-14248.

Kim, D. and M. H. Yu (1996). "Folding pathway of human alpha 1-antitrypsin: characterization of an intermediate that is active but prone to aggregation." Biochem Biophys Res Commun **226**(2): 378-384.

Kim, Y. I., R. E. Burton, et al. (2000). "Dynamics of substrate denaturation and translocation by the ClpXP degradation machine." Mol Cell **5**(4): 639-648.

LaFrentz, B. R., S. E. LaPatra, et al. (2011). "Identification of immunogenic proteins within distinct molecular mass fractions of Flavobacterium psychrophilum." J Fish Dis **34**(11): 823-830.

Lee, S., J. M. Choi, et al. (2007). "Visualizing the ATPase cycle in a protein disaggregating machine: structural basis for substrate binding by ClpB." Mol Cell **25**(2): 261-271.

Lee, S., B. Sielaff, et al. (2010). "CryoEM structure of Hsp104 and its mechanistic implication for protein disaggregation." Proc Natl Acad Sci U S A **107**(18): 8135-8140.

Lee, S., M. E. Sowa, et al. (2003). "The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state." Cell **115**(2): 229-240.

Li, J. and B. Sha (2003). "Crystal structure of the E. coli Hsp100 ClpB N-terminal domain." Structure **11**(3): 323-328.

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

Liu, T. and X. Wang (2010). "Zinc induces unfolding and aggregation of dimeric arginine kinase by trapping reversible unfolding intermediate." Acta Biochim Biophys Sin (Shanghai) **42**(11): 779-786.

Liu, Z., V. Tek, et al. (2002). "Conserved amino acid residues within the amino-terminal domain of ClpB are essential for the chaperone activity." J Mol Biol **321**(1): 111-120.

Lo, J. H., T. A. Baker, et al. (2001). "Characterization of the N-terminal repeat domain of Escherichia coli ClpA-A class I Clp/HSP100 ATPase." Protein Sci **10**(3): 551-559.

Lourdault, K., G. M. Cerqueira, et al. (2011). "Inactivation of clpB in the pathogen Leptospira interrogans reduces virulence and resistance to stress conditions." Infect Immun **79**(9): 3711-3717.

Lum, R., J. M. Tkach, et al. (2004). "Evidence for an unfolding/threading mechanism for protein disaggregation by Saccharomyces cerevisiae Hsp104." J Biol Chem **279**(28): 29139-29146.

Mackay, R. G., C. W. Helsen, et al. (2008). "The C-terminal extension of Saccharomyces cerevisiae Hsp104 plays a role in oligomer assembly." Biochemistry **47**(7): 1918-1927.

Mayer, M. P. and B. Bukau (2005). "Hsp70 chaperones: cellular functions and molecular mechanism." Cell Mol Life Sci **62**(6): 670-684.

Miot, M., M. Reidy, et al. (2011). "Species-specific collaboration of heat shock proteins (Hsp) 70 and 100 in thermotolerance and protein disaggregation." Proc Natl Acad Sci U S A **108**(17): 6915-6920.

Mogk, A., C. Schlieker, et al. (2003). "Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity." J Biol Chem **278**(20): 17615-17624.

Monsellier, E. and F. Chiti (2007). "Prevention of amyloid-like aggregation as a

driving force of protein evolution." EMBO Rep **8**(8): 737-742.

Motohashi, K., Y. Watanabe, et al. (1999). "Heat-inactivated proteins are rescued by the DnaK.J-GrpE set and ClpB chaperones." Proc Natl Acad Sci U S A **96**(13): 7184-7189.

Nagy, M., I. Guenther, et al. (2010). "Synergistic cooperation between two ClpB isoforms in aggregate reactivation." J Mol Biol **396**(3): 697-707.

Neuwald, A. F., L. Aravind, et al. (1999). "AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes." Genome Res **9**(1): 27-43.

Ogura, T. and A. J. Wilkinson (2001). "AAA+ superfamily ATPases: common structure--diverse function." Genes Cells **6**(7): 575-597.

Pahl, A., K. Brune, et al. (1997). "Fit for life? Evolution of chaperones and folding catalysts parallels the development of complex organisms." Cell Stress Chaperones **2**(2): 78-86.

Park, S. K., K. I. Kim, et al. (1993). "Site-directed mutagenesis of the dual translational initiation sites of the clpB gene of Escherichia coli and characterization of its gene products." J Biol Chem **268**(27): 20170-20174.

Parsell, D. A., A. S. Kowal, et al. (1994). "Protein disaggregation mediated by heat-shock protein Hsp104." Nature **372**(6505): 475-478.

Queitsch, C., S. W. Hong, et al. (2000). "Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis." Plant Cell **12**(4): 479-492.

Reid, B. G., W. A. Fenton, et al. (2001). "ClpA mediates directional translocation of substrate proteins into the ClpP protease." Proc Natl Acad Sci U S A **98**(7): 3768-3772.

Sanchez de Groot, N., I. Pallares, et al. (2005). "Prediction of "hot spots" of aggregation in disease-linked polypeptides." BMC Struct Biol **5**: 18.

Sanchez, Y. and S. L. Lindquist (1990). "HSP104 required for induced thermotolerance." Science **248**(4959): 1112-1115.

Sanchez, Y., J. Taulien, et al. (1992). "Hsp104 is required for tolerance to many

forms of stress." EMBO J **11**(6): 2357-2364.

Sauer, R. T. and T. A. Baker (2011). "AAA+ proteases: ATP-fueled machines of protein destruction." Annu Rev Biochem **80**: 587-612.

Sawkar, A. R., W. D'Haeze, et al. (2006). "Therapeutic strategies to ameliorate lysosomal storage disorders--a focus on Gaucher disease." Cell Mol Life Sci **63**(10): 1179-1192.

Schlieker, C., I. Tews, et al. (2004). "Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides." FEBS Lett **578**(3): 351-356.

Shorter, J. and S. Lindquist (2004). "Hsp104 catalyzes formation and elimination of self-replicating Sup35 prion conformers." Science **304**(5678): 1793-1797.

Sielaff, B. and F. T. Tsai (2010). "The M-domain controls Hsp104 protein remodeling activity in an Hsp70/Hsp40-dependent manner." J Mol Biol **402**(1): 30-37.

Skach, W. R. (2009). "Cellular mechanisms of membrane protein folding." Nat Struct Mol Biol **16**(6): 606-612.

Skolnick, J. and J. S. Fetrow (2000). "From genes to protein structure and function: novel applications of computational approaches in the genomic era." Trends Biotechnol **18**(1): 34-39.

Smith, D. F., L. Whitesell, et al. (1998). "Molecular chaperones: biology and prospects for pharmacological intervention." Pharmacol Rev **50**(4): 493-514.

Sousa, M. C., C. B. Trame, et al. (2000). "Crystal and solution structures of an HslUV protease-chaperone complex." Cell **103**(4): 633-643.

Speed, M. A., D. I. Wang, et al. (1995). "Multimeric intermediates in the pathway to the aggregated inclusion body state for P22 tailspike polypeptide chains." Protein Sci **4**(5): 900-908.

Speransky, V. V., K. L. Taylor, et al. (2001). "Prion filament networks in [URE3] cells of Saccharomyces cerevisiae." J Cell Biol **153**(6): 1327-1336.

Squires, C. L., S. Pedersen, et al. (1991). "ClpB is the Escherichia coli heat

shock protein F84.1." J Bacteriol **173**(14): 4254-4262.

Stefani, M. (2010). "Protein aggregation diseases: toxicity of soluble prefibrillar aggregates and their clinical significance." Methods Mol Biol **648**: 25-41.

Stefani, M. and C. M. Dobson (2003). "Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution." J Mol Med (Berl) **81**(11): 678-699.

Tanaka, N., Y. Tani, et al. (2004). "Interaction of the N-terminal domain of Escherichia coli heat-shock protein ClpB and protein aggregates during chaperone activity." Protein Sci **13**(12): 3214-3221.

Tipton, K. A., K. J. Verges, et al. (2008). "In vivo monitoring of the prion replication cycle reveals a critical role for Sis1 in delivering substrates to Hsp104." Mol Cell **32**(4): 584-591.

Ventura, S. (2005). "Sequence determinants of protein aggregation: tools to increase protein solubility." Microb Cell Fact **4**(1): 11.

Ventura, S., J. Zurdo, et al. (2004). "Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case." Proc Natl Acad Sci U S A **101**(19): 7258-7263.

Watanabe, Y. H., M. Takano, et al. (2005). "ATP binding to nucleotide binding domain (NBD)1 of the ClpB chaperone induces motion of the long coiled-coil, stabilizes the hexamer, and activates NBD2." J Biol Chem **280**(26): 24562-24567.

Weibezahn, J., P. Tessarz, et al. (2004). "Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB." Cell **119**(5): 653-665.

Werbeck, N. D., J. N. Kellner, et al. (2009). "Nucleotide binding and allosteric modulation of the second AAA+ domain of ClpB probed by transient kinetic studies." Biochemistry **48**(30): 7240-7250.

Wetzel, R. (1994). "Mutations and off-pathway aggregation of proteins." Trends Biotechnol **12**(5): 193-198.

Wickner, R. B. (1994). "[URE3] as an altered URE2 protein: evidence for a prion

analog in Saccharomyces cerevisiae." Science **264**(5158): 566-569.

Woo, K. M., K. I. Kim, et al. (1992). "The heat-shock protein ClpB in Escherichia coli is a protein-activated ATPase." J Biol Chem **267**(28): 20429-20434.

Yoon, S. and W. J. Welsh (2004). "Detecting hidden sequence propensity for amyloid fibril formation." Protein Sci **13**(8): 2149-2160.

Yukitake, H., M. Naito, et al. (2011). "Effects of non-iron metalloporphyrins on growth and gene expression of Porphyromonas gingivalis." Microbiol Immunol **55**(3): 141-153.

Zerovnik, E., V. Stoka, et al. (2011). "Mechanisms of amyloid fibril formation--focus on domain-swapping." FEBS J **278**(13): 2263-2282.

Zhang, Z., H. Chen, et al. (2007). "Identification of amyloid fibril-forming segments based on structure and residue-based statistical potential." Bioinformatics **23**(17): 2218-2225.

Zietkiewicz, S., A. Lewandowska, et al. (2006). "Hsp70 chaperone machine remodels protein aggregates at the initial step of Hsp70-Hsp100-dependent disaggregation." J Biol Chem **281**(11): 7022-7029.

Zietkiewicz, S., M. J. Slusarz, et al. (2010). "Conformational stability of the full-atom hexameric model of the ClpB chaperone from Escherichia coli." Biopolymers **93**(1): 47-60.

Zolkiewski, M. (1999). "ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from Escherichia coli." J Biol Chem **274**(40): 28083-28086.

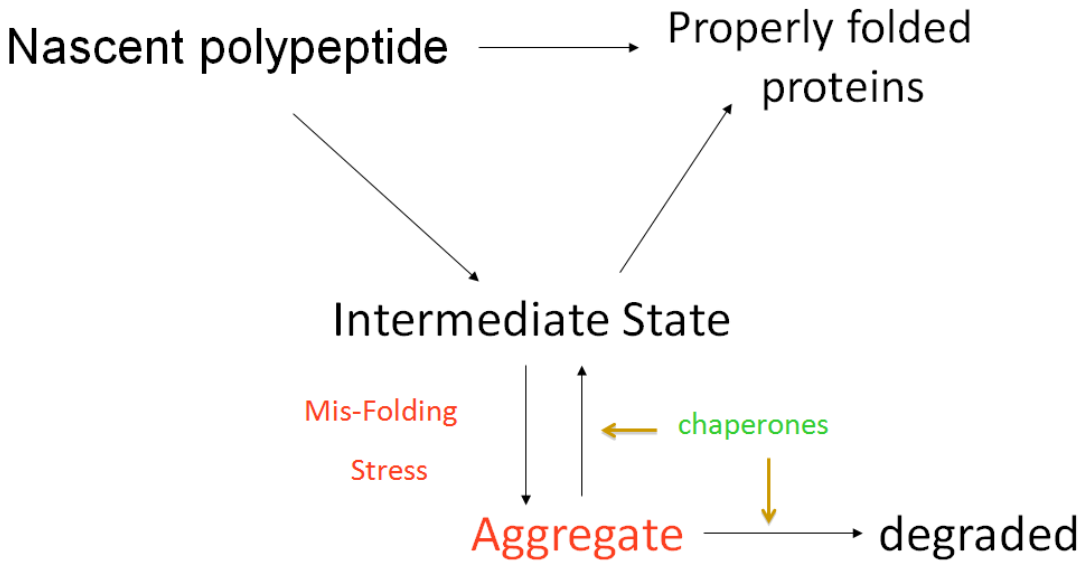
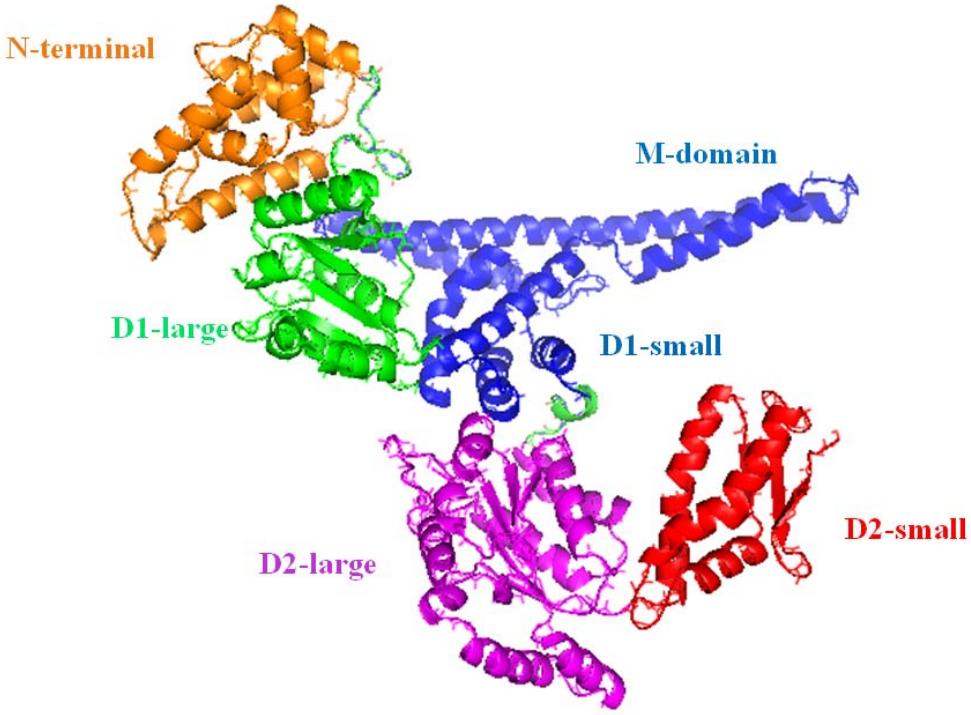


Fig. 1 Protein folding and aggregation.

Large polypeptide chains may form partly folded intermediates which tend to misfold and interact with each other under stress conditions such as heat shock. As a result, aggregates are formed. Cells have evolved chaperones that can promote proper folding, thus avoiding accumulation of aggregates or reversing the aggregation process.



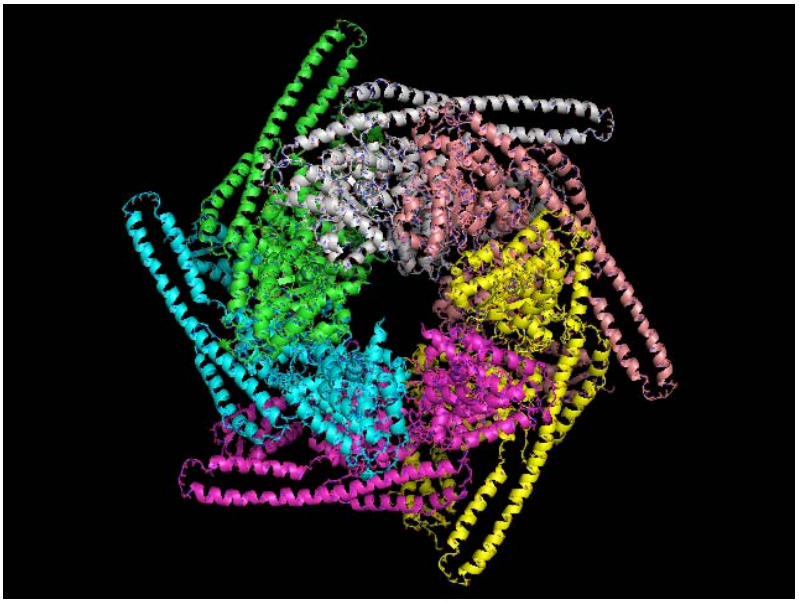


Fig. 2 ClpB structure.

Top: A ribbon drawing of the *Thermophilus* ClpB (PDB code: 1qvr) (Lee et al., 2003). The N-terminal domain is colored in orange and is located on the top of the molecule. The coiled coil middle domain is inserted into the small sub-domain of the first AAA+ module (D1-small).

Bottom: Top view of ClpB hexamer.

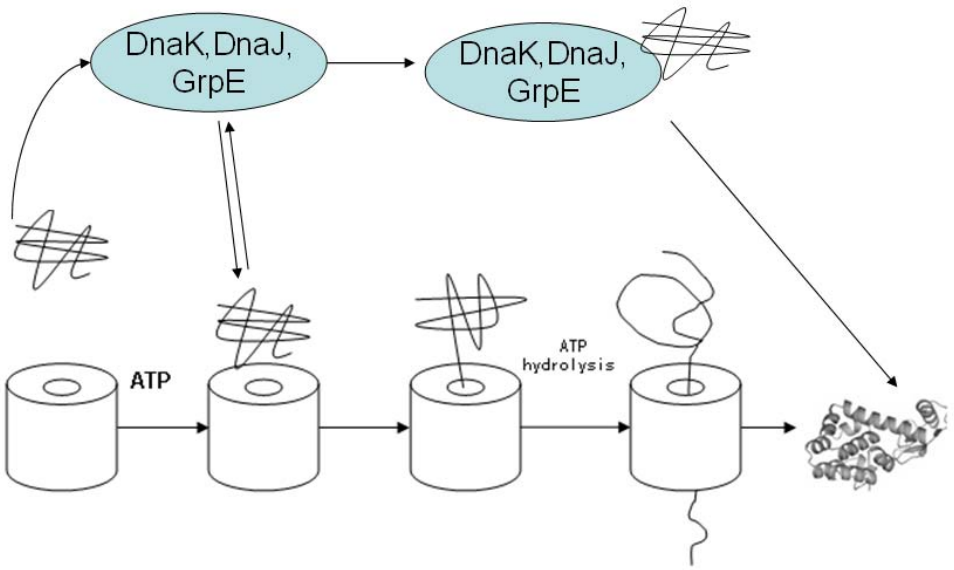


Fig. 3 Mechanism of ClpB disaggregation.

A ClpB hexamer is shown as a cylinder with a central channel. DnaK/DnaJ/GrpE co-chaperones can reactivate smaller aggregates. Larger aggregates bind to the ClpB hexamer and DnaK/DnaJ/GrpE may facilitate the insertion of aggregates into the central channel of ClpB hexamer. Polypeptide chains from aggregates are threaded through the channel using energy provided by ATP hydrolysis.

Chapter 2

Flexible connection of the N-terminal domain in ClpB supports substrate binding and controls the aggregate reactivation efficiency

2.1 Introduction

ClpB, as most other AAA+ family members, forms oligomers to carry out proper function. In cooperation with Hsp70 system (DnaK, DnaJ and GrpE in *E. coli*), ClpB reactivates aggregated substrates by translocating them through the central channel of its hexamer. This disaggregation function is also found in other bacteria and yeast, but not in higher organisms like mammals (Sharma, Christen et al. 2009).

ClpB is a multi-domain protein which consists of two nucleotide binding domains (NBD1 and NBD2) connected by a middle domain (M domain), and an N-terminal domain connected to the rest of the protein by a flexible linker (Fig. 1A). The effective disaggregation function of ClpB requires the precise cooperation of all the domains. For example, ATP binding to NBD1 will cause the movement of M domain, resulting in stabilization of ClpB hexamer and stimulation of ATP hydrolysis in NBD2 (Watanabe, Takano et al. 2005). Also, impairing nucleotide binding in NBD1 will decrease the ATP hydrolysis activity in NBD2 of the same monomer (Werbeck, Schlee et al. 2008). The N-terminal domain seems less important since it is dispensable in some in vitro disaggregation studies (Sielaff and Tsai 2010). However there is also evidence showing that the N-terminal domain is required for binding and reactivating strongly aggregated substrates (Barnett, Nagy et al. 2005) and direct interaction between N-domain and substrate was also observed (Tanaka, Tani et al. 2004). The N-domain resides on the top surface of ClpB hexamer and shows a high mobility with 120 degree difference is observed in crystal structure (Lee, Sowa et al. 2003), suggesting its possible role in substrate recognition and initial processing.

There are two isoforms of ClpB in *E. coli*, full length ClpB95 and the N-terminal

truncated ClpB80 due to different translation initiation sites. The coordination between these two isoforms improves the survival ability of *E. coli* under stress, possibly because the ClpB95/ClpB80 hetero-oligomer is a more effective disaggregation machine than the homo-oligomer of full length ClpB95 (Barnett, Nagy et al. 2005; Chow and Baneyx 2005; Chow, Barnett et al. 2005). It is obvious that in the ClpB95/ClpB80 hetero-oligomer, there is more space on the top surface of the hexamer thus the N-terminal domain has more freedom to move, which may lead to a more efficient substrate binding and recognition. The only connection between the N-domain and the rest of ClpB is the 17 aa long unstructured linker since no physical contact is observed between N-domain and the rest of ClpB (Tek and Zolkiewski 2002). In this study, we focused on how the linker may influence the mobility of N-domain and if it could have an impact on the whole reactivation machinery.

By sequence alignment we found that there is not very much similarity among the linker regions from different species but there are two conserved glycines (Gly149, Gly150 in *E. coli* ClpB) that might be important for the flexibility of the linker (Fig. 1B). We hypothesized that by inserting more glycines we could increase the flexibility of the linker and thus improve the mobility of N-domain, and by deleting the glycines or substituting them with prolines we could decrease the flexibility of the linker and thus decrease the mobility of N-domain. Seven mutants (Fig. 1C) were made and their biochemical properties and activities were tested. We found that although the linker mutants had similar biochemical properties as the wild type, all of them showed lower chaperone activity than the wild type, which may be caused by lower substrate binding. Thus it appears that the linker of wild type ClpB is optimized during evolution.

2.2 Materials and methods.

Proteins and Aggregates. Chaperones (ClpB, DnaK, DnaJ, GrpE) were produced or obtained as previously described (Barnett, Nagy et al. 2005). G6PDH from *Leuconostoc mesenteroides* and α-casein were obtained from Sigma (St. Louis, MO).

Firefly luciferase was obtained from Promega (Madison, WI). GFP was produced as described previously (Doyle, Shorter et al. 2007). Protein concentrations were determined spectrophotometrically. Mutations in the ClpB linker DNA sequence were introduced using the QuikChange site-directed mutagenesis kit (Stratagene/Agilent Technologies).

To produce aggregates of G6PDH, the protein stock (324 μM) was diluted 2-fold with unfolding buffer (10 M urea, 16% glycerol and 40 mM DTT) and was incubated at 47 °C for 5 min. The mixture was then diluted 10-fold with refolding buffer 1 (50 mM Tris/HCl pH 7.5, 20 mM $Mg(OAc)_2$, 30 mM KCl, 1 mM EDTA, and 1 mM β-mercaptoethanol) and was incubated at 47 °C for 15 min and then on ice for 2 min. To produce aggregates of luciferase, 216 μM luciferase stock was diluted 300-fold with PBS containing 1 mg/ml BSA and then was incubated at 45 °C for 12 min. To produce aggregated GFP, 4.5 μM protein was heated for 10 min at 80 ˚C.

Differential Scanning Calorimetry. DSC experiments were performed with a VP-DSC calorimeter (MicroCal Inc., Northampton, Massachusetts) at a scan-rate of 1 K/min. For each protein sample, the instrument baseline was obtained first by measuring the thermogram of the dialysis buffer. Subsequently, the buffer in the sample cell was replaced with a ClpB solution and the protein thermogram was measured. The thermal unfolding of ClpB was irreversible, as shown by the lack of endotherms in the repetitive scans of each protein sample.

Analytical Ultracentrifugation. A Beckman XL-I analytical ultracentrifuge was used in sedimentation velocity experiments with two-channel analytical cells. The data were analyzed using the time-derivative method (Stafford 1992; Stafford III 1994) and the software distributed with the instrument.

ClpB ATPase activity. The ClpB variants were incubated in assay buffer (100 mM Tris/HCl pH 8.0, 1 mM DTT, 1 mM EDTA, 10 mM $MgCl_2$, and 5 mM ATP) at 37 °C for 15 min without or with 0.1 mg/ml α-casein or 0.04 mg/ml poly-lysine. The

concentration of ClpB was 0.05 mg/ml for the basal activity and in the presence of α-casein or 0.005 mg/ml in the presence of poly-lysine. The phosphate concentration generated by ClpB was measured as described previously (Zolkiewski 1999).

Aggregate Reactivation Assays. Aggregated G6PDH (16.2 μM) was diluted 10-fold with refolding buffer 1 containing 1.5 μM ClpB, 1 μM DnaK, 1 μM DnaJ, 0.5 μM GrpE and 6 mM ATP. The mixture was incubated at 30 °C and aliquots of the mixture were withdrawn to test the recovery of the G6PDH enzymatic activity. Aggregates diluted with refolding buffer without the chaperones were used as control. To measure the G6PDH activity, aliquots from the refolding reaction were incubated in 50 mM Tris/HCl pH 7.8, 5 mM $MgCl_2$, 1.5 mM G6P and 1 mM $NADP^+$ for 10 min followed by the measurement of absorption at 340 nm. Aggregated luciferase (0.7 μM) was diluted 20-fold with refolding buffer 2 (30 mM Hepes, pH 7.65, 120 mM KCl, 10 mM $MgCl_2$, 6 mM ATP, 1 mM EDTA, 10 mM DTT, 0.1 mg/ml BSA) containing 1.5 μM ClpB, 1 μM DnaK, 1 μM DnaJ, and 0.5 μM GrpE. The mixture was incubated at room temperature and aliquots were withdrawn to test the recovery of the luciferase activity using the luminescence assay system (Promega, Madison, WI). GFP reactivation assays (100 μL) were performed in 20 mM Tris-HCl, pH 7.5, 100 mM KCl, 5 mM DTT, 0.1 mM EDTA, and 10% glycerol (vol/vol) with 2 mM ATP and 2 mM ATPγS, an ATP regenerating system (20 mM creatine phosphate and 6 μg creatine kinase), 10 mM $MgCl_2$, 10 μL heat-aggregated GFP and 1.0 μM ClpB. Reactions were initiated by the addition of Mg-ATP and reactivation was monitored over time at 23 °C using a Perkin-Elmer LS50B luminescence spectrophotometer with a plate reader. Excitation and emission wavelengths were 395 nm and 510 nm, respectively.

ClpB-Aggregate Interaction Assay. Aggregated G6PDH (16.2 μM) was diluted 10-fold with the refolding buffer 1 containing 1.5 μM ClpB and 6 mM ATPγS. The mixture was incubated at 30 °C with 600 rpm shaking for 20 min and then was applied to the filter device (Millipore Ultrafree-MC Centrifugal Filter Unit with the

membrane, pore size 0.1 μm). After 5 min incubation at room temperature, the filter device was centrifuged at 13,000 rpm for 4 min to get the flow-through fractions. The filter device was washed with the refolding buffer 1 containing ATPγS at 30 °C for 5 min and then re-centrifuged. Next, 1x SDS loading buffer was added to the filter device and the filter device was incubated at 50 °C for 5 min with shaking. Then, it was centrifuged to get the eluate fractions, which were applied to SDS-PAGE. The Coomassie-stained band intensity was determined with the BandScan software (http://bandscan.software.informer.com).

2.3 Results.

2.3.1 Biochemical properties of the linker mutants

After purification of the linker mutants, their biochemical properties were tested. Differential Scanning Calorimetry (DSC) experiments were carried out to test the structural integrity of the linker mutants. Thermal denaturation of ClpB variants is shown in Fig. 2. As we can see, thermal denaturation of ClpB wild type is irreversible with two transition peaks around 52°C and 62°C, the higher peak representing the unfolding of N-domain shown in the previous study (Liu, Tek et al. 2002). All the linker mutants showed similar transition pattern as the wild type, suggesting the protein integrity is not affected by the linker mutation.

Another important feature of ClpB is its self-association, which could be enhanced by high protein concentration, the presence of nucleotide or lower salt level (Schlee, Groemping et al. 2001). Previous study showed that N-terminal truncated ClpB80 had a preference to from hexamer comparing to the wild type at low protein concentration (Barnett, Zolkiewska et al. 2000), which led us to consider the role that the linker mutants may play in the oligomerization of ClpB. Sedimentation velocity is a very efficient way to provide the information about the mass and shape of a molecule. As the equilibrium between ClpB monomer and hexamer is mainly dependent on the protein concentration, we managed to find a condition where both

monomers and hexamers existed in the buffer (Fig. 3). The most radically changed mutants were tested. Distributions of the sedimentation coefficients g(s*) showed both monomeric (4S) and oligomeric (14S) forms of ClpB in the solution. All the mutants had similar sedimentation coefficient for the hexamer, suggesting that the self-association of the mutants were not affected. Although the delG mutant and 2P mutant had a small shift in the distribution (around 12S), indicating lower affinity of the formed hexamer by these two mutants, this is not likely to have much influence on the activity of the chaperone machinery.

2.3.2 Molecular dynamics

Molecular dynamics was used to gain insights into the structural properties of the linker mutants (data obtained by Elizabeth Ploetz from Kansas State University, Department of Chemistry not shown). Since the mutations only took place in the two conserved glycine region, we assumed that the rest of the linker and the protein is irrelevant to the internal dynamics of the linker. Thus only the mutated region MRGGES was simulated. As we hypothesized, as the number of glycine increased, the movement scale of the segment expanded, showing a more flexible linker; on the contrary, the movement of 1P, 2P and DelG mutants was limited, indicating less flexible linker, especially the DelG mutant.

2.3.3 ATPase activity assay of linker mutants

Previous data showed that the oligomerization of the linker mutants is not affected. Since the ATP hydrolysis occurs in the interface of the subunits of the hexamer, we want to see if the ATPase activity of the linker mutants is also unaltered. As shown in Fig. 4, the basal ATPase activity of all the linker mutants is the same as wild type, which further proves that oligomerization is not affected. However, all the linker mutants failed to respond to activation by the pseudo-substrate alpha-casein, with only about half activity of the wild type. In the case of another ATPase activator poly-lysine, there was not much difference between the linker mutants and the wild type (Fig. 4). These results suggest there is certain deficiency in the linker mutants

and more experiments need to be done to reveal why.

2.3.4 Chaperone assay of linker mutants

We tested the reactivation by the linker mutants of different aggregated substrates. Fig. 5 shows the reactivation of large G6PDH aggregates made by urea unfolding followed by refolding in the presence of co-chaperone DnaK, DnaJ and GepE (KJE). All the mutants showed lower reactivation activity compared to wild type. The decrease of chaperone activity was most severe in the 2P mutant, which had 80% decrease; on the other hand, the less manipulated mutants 1G and 2G had only about 40% decrease.

In the reactivation of heat aggregated luciferase, no significant deficiency was observed in the 1G and 2G mutant. All the other mutants showed a 50% to 70% decrease in activity, 3G and 4G being the most inhibited (Fig. 6).

We also tested the reactivation of GFP. As a smaller aggregated substrate, GFP could be reactivated in the absence of co-chaperone when a mixture of ATP and ATPγS were provided (Doyle, Shorter et al. 2007). All the mutants showed undistinguishable levels of reactivation as wild type. (Fig. 7, experiment performed by Shannon Doyle, National Institutes of Health)

All these data suggested that the linker mutants had deficiencies in chaperone activity, especially when dealing with large aggregates that are difficult to reactivate.

2.3.5 Substrate binding of linker mutants

Next we tried to find out the reason why the linker mutants could not work as well as the wild type. We carried out a filter assay that was developed in our laboratory. Large aggregates of G6PDH were made and mixed with ClpB in the presence of ATPγS. ATPγS binds to ClpB but cannot be hydrolyzed, which keeps ClpB in the substrate binding state. G6PDH aggregates bound to ClpB will stay on the filter because they are too large to pass. The mixture was then analyzed by SDS-PAGE. From Fig. 8, we can see that trace amounts of ClpB stayed on the filter without aggregates (Fig. 8 left part) but the amount of ClpB significantly increased

when substrate was provided (Fig. 8 right part), showing binding of ClpB to the aggregated substrate. All the mutants had less binding to the substrate than the wild type. 3G, 4G and 2P only showed ~60% of substrate binding. This lack of substrate binding by the mutants mirrored the decrease of reactivation in Fig. 5, which could be an explanation for the deficiency of the linker mutants in chaperone activity.

2.4 Discussion

In this study we constructed linker variants with different flexibility and found that all the linker mutants had deficiency in disaggregation, which demonstrated the important role of the linker in proper disaggregation function of the ClpB chaperone machinery, even without a folded structure.

In the crystal structure of T. *thermophilus* ClpB monomer, the N-terminal domain orientation is rotated 120 degrees in two different conformations, suggesting the mobility of the N-terminal domain, which may support aggregate substrate binding (Lee, Sowa et al. 2003). Since there is no direct contact between the N-terminal domain and NBD1 (Tek and Zolkiewski 2002), the mobility of the N-terminal domain depends completely on the unstructured linker (Fig. 9). After we inserted additional glycines into the linker, it gained flexibility and supplied more space for the N-terminal domain movement. This was confirmed in the MD simulation results, from which we could see that 3G and 4G had much stronger movements than the wild type; on the other hand, movement of delG and 2P were strictly limited. The deletion of two conserved glycines also shortened the linker up to 30% of the wild type while all the inserted glycine mutants were slightly longer. Since the changes were not significant, we still focused on the flexibility instead of the length of the linker.

Despite the differences shown in the MD simulation, all the linker mutants had similar biochemical properties. All the mutants had similar oligomerization and protein structural integrity, all failed to respond to pseudo-substrate in ATPase activity assay, and all had deficiencies in protein disaggregation. It seems that although there

is so little conservation in the sequences of linkers from different species (Fig. 1A), the linker is still optimized during evolution. This is in agreement with a previous study of the ClpA linker, in which the length of the linker was optimized (Cranz-Mileva, Imkamp et al. 2008). Taken together, these results may be applied to many other AAA+ proteins which have unstructured but yet optimized linkers.

We also looked into why the linker mutants showed weaker disaggregation activity. One possible reason is that the modified linker inhibits the binding of substrate to ClpB. Weaker binding to substrates could cause weaker reactivation. However there is no evidence showing the direct interaction of the linker to the substrate and this hypothesis is highly unlikely. The linker has no stable structure, and the sequence of the linker contains mainly negatively charged residues which do not support binding. Previous studies showed that the N-terminal domain supports binding to strongly aggregated substrates (Barnett, Nagy et al. 2005), and even a small his-tag on the N-terminal domain will abolish its response to pseudo substrate in ATPase activity assay (Chow, Barnett et al. 2005). Thus, the mobility of the N-terminal domain is essential for substrate binding and controls the reactivation efficiency.

Different from the reactivation of luciferase and G6PDH, the linker mutants had similar reactivation efficiency as the wild type in the GFP reactivation. There are two possible reasons for this: firstly, GFP aggregates in this assay are rather small comparing to the other two. When working with smaller aggregates, the N-terminal domain of ClpB is dispensable (Clarke and Eriksson 2000; Beinker, Schlee et al. 2002; Mogk, Schlieker et al. 2003). Secondly, GFP reactivation took place in the absence of co-chaperones. It is possible that the mobility of the N-terminal domain may support interaction of ClpB and DnaK, thus when DnaK is absent, this effect may be absent.

The mechanism of the ClpB/KJE machinery is still unknown. We know that aggregated protein is recognized and then translocated by ClpB, but considering the variety of substrate types it is difficult to find one certain sequence in the substrates for ClpB binding. Instead, there may be many sequences which share some similar

properties such as charged and aromatic residues (Schlieker, Tews et al. 2004). From the structure of ClpB hexamer, we can imagine that the substrate binding should occur on the surface of the entrance to the central channel of ClpB. Thus the N-terminal domain may play an important role in facilitating the substrate entering the channel. From Fig. 9, we can see that the N-terminal domain sits like a crown of the rest of ClpB and is capable of blocking the entrance of the translocation channel. To make the chaperone machinery work effectively, the protein must make sure that the N-terminal domain has certain mobility that it could control the entrance of substrate to the channel as well as not block it. As a result of the modified linker, the orientation of the N-terminal domain is changed, and thus binding sites of ClpB could be covered or the entrance of the channel may be blocked, causing loss of reactivation efficiency.

In conclusion, our work demonstrates that the ClpB linker supports substrate binding and controls reactivation efficiency, and it is optimized during evolution. Even though the linker lacks a folded structure, it may still be able to transfer allosteric signals from the N-terminal domain to NBD1 and NBD2. It is important to not overlook the role of unstructured linkers between domains in multi-domain proteins.

Acknowledgements

This work was supported by the National Institutes of Health (GM079277, ARRA Supplement), the Terry C. Johnson Center for Basic Cancer Research, and the Kansas Agricultural Experiment Station. I thank Dr. Paul E. Smith and Elizabeth A. Ploetz for the MD simulation, Dr. Sue Wickner and Dr. Shannon Doyle for the GFP experiment, and Dr. Maria Nagy for making the 4G mutant.

References

Barnett, M. E., M. Nagy, et al. (2005). "The amino-terminal domain of ClpB supports binding to strongly aggregated proteins." J Biol Chem **280**(41): 34940-34945.

Barnett, M. E., M. Nagy, et al. (2005). "The amino-terminal domain of ClpB supports binding to strongly aggregated proteins." The Journal of biological chemistry **280**(41): 34940-34945.

Barnett, M. E., A. Zolkiewska, et al. (2000). "Structure and activity of ClpB from Escherichia coli. Role of the amino-and -carboxyl-terminal domains." J Biol Chem **275**(48): 37565-37571.

Beinker, P., S. Schlee, et al. (2002). "The N terminus of ClpB from Thermus thermophilus is not essential for the chaperone activity." J Biol Chem **277**(49): 47160-47166.

Chow, I. T. and F. Baneyx (2005). "Coordinated synthesis of the two ClpB isoforms improves the ability of Escherichia coli to survive thermal stress." FEBS Lett **579**(20): 4235-4241.

Chow, I. T., M. E. Barnett, et al. (2005). "The N-terminal domain of Escherichia coli ClpB enhances chaperone function." FEBS Lett **579**(20): 4242-4248.

Clarke, A. K. and M. J. Eriksson (2000). "The truncated form of the bacterial heat shock protein ClpB/HSP100 contributes to development of thermotolerance in the cyanobacterium Synechococcus sp. strain PCC 7942." J Bacteriol **182**(24): 7092-7096.

Cranz-Mileva, S., F. Imkamp, et al. (2008). "The flexible attachment of the N-domains to the ClpA ring body allows their use on demand." J Mol Biol **378**(2): 412-424.

Doyle, S. M., J. Shorter, et al. (2007). "Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity." Nature structural &

molecular biology **14**(2): 114-122.

Doyle, S. M., J. Shorter, et al. (2007). "Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity." Nat Struct Mol Biol **14**(2): 114-122.

Lee, S., M. E. Sowa, et al. (2003). "The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state." Cell **115**(2): 229-240.

Liu, Z., V. Tek, et al. (2002). "Conserved amino acid residues within the amino-terminal domain of ClpB are essential for the chaperone activity." J Mol Biol **321**(1): 111-120.

Mogk, A., C. Schlieker, et al. (2003). "Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity." J Biol Chem **278**(20): 17615-17624.

Schlee, S., Y. Groemping, et al. (2001). "The chaperone function of ClpB from Thermus thermophilus depends on allosteric interactions of its two ATP-binding sites." J Mol Biol **306**(4): 889-899.

Schlieker, C., I. Tews, et al. (2004). "Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides." FEBS Lett **578**(3): 351-356.

Sharma, S. K., P. Christen, et al. (2009). "Disaggregating chaperones: an unfolding story." Curr Protein Pept Sci **10**(5): 432-446.

Sielaff, B. and F. T. Tsai (2010). "The M-domain controls Hsp104 protein remodeling activity in an Hsp70/Hsp40-dependent manner." J Mol Biol **402**(1): 30-37.

Stafford III, W. F., Ed. (1994). Sedimentation boundary analysis of interacting systems: Use of the apparent sedimentation coefficient distribution function. Modern Analytical Ultracentrifugation. Boston, Birkhauser.

Stafford, W. F., 3rd (1992). "Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile." Analytical biochemistry **203**(2): 295-301.

Tanaka, N., Y. Tani, et al. (2004). "Interaction of the N-terminal domain of Escherichia coli heat-shock protein ClpB and protein aggregates during chaperone activity." Protein Sci **13**(12): 3214-3221.

Tek, V. and M. Zolkiewski (2002). "Stability and interactions of the amino-terminal domain of ClpB from Escherichia coli." Protein Sci **11**(5): 1192-1198.

Watanabe, Y. H., M. Takano, et al. (2005). "ATP binding to nucleotide binding domain (NBD)1 of the ClpB chaperone induces motion of the long coiled-coil, stabilizes the hexamer, and activates NBD2." J Biol Chem **280**(26): 24562-24567.

Werbeck, N. D., S. Schlee, et al. (2008). "Coupling and dynamics of subunits in the hexameric AAA+ chaperone ClpB." J Mol Biol **378**(1): 178-190.

Zolkiewski, M. (1999). "ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from Escherichia coli." The Journal of biological chemistry **274**(40): 28083-28086.

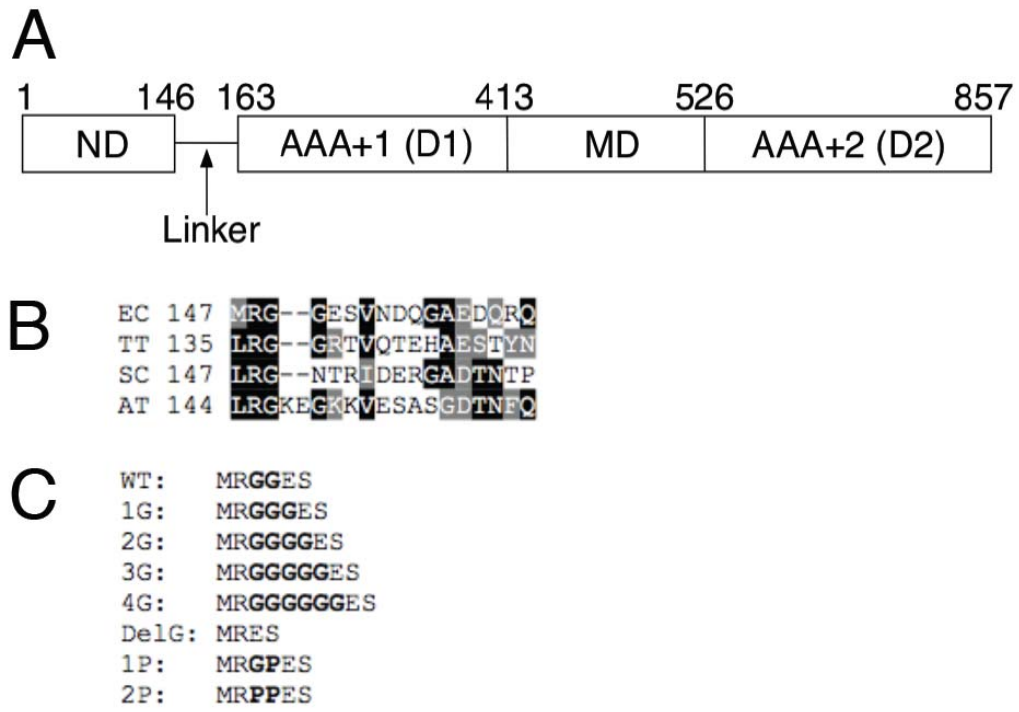


Fig. 1 (A) Domain structure of ClpB. The diagram shows the structural domains of ClpB as determined by X-ray crystallography (Lee, Sowa et al. 2003): the N-terminal domain (ND), D1 and D2 AAA+ modules, and the middle domain (MD). The residue numbers are given for *E. coli* ClpB and the position of the unstructured linker is indicated. (B) Sequence alignment of the ClpB N-terminal linkers from *Escherichia coli*, *Thermus thermophilus*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*. (C) Modifications of the ClpB linker sequences produced in this work.

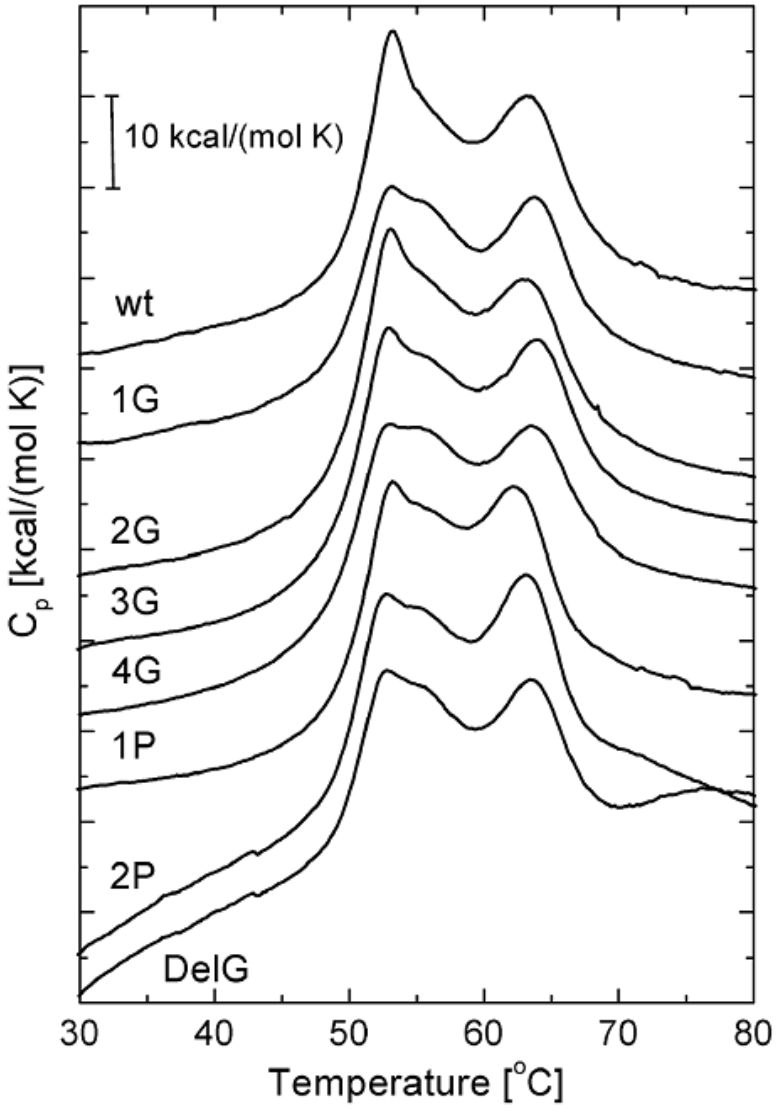


Fig. 2 Differential scanning calorimetry of ClpB and its modified linker variants. DSC thermograms were obtained at the 1 K/min scan rate for the protein samples at 0.7 mg/ml. A baseline measured with the dialysis buffer (50 mM Hepes/KOH, pH. 7.5) without the protein was subtracted from the protein scans. The DSC data were normalized for protein concentration. Constant offsets were applied to the data sets for the clarity of presentation.

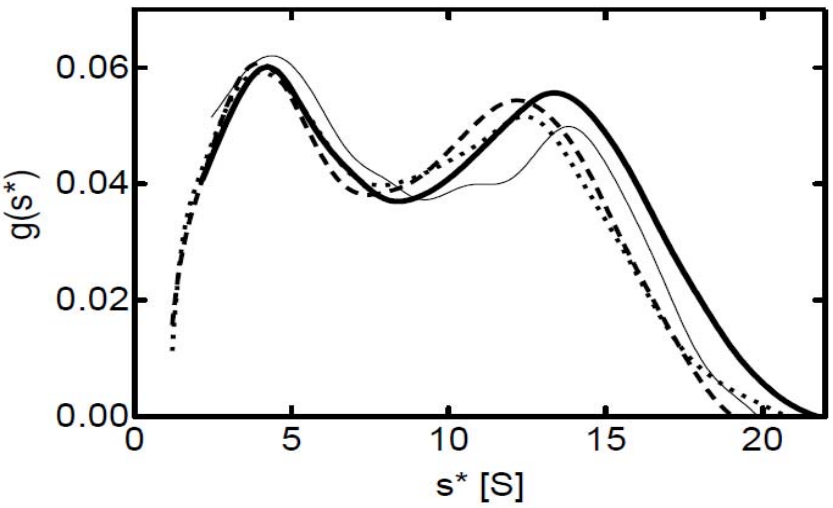


Fig. 3 Sedimentation velocity analysis of ClpB and its modified linker variants. Ultracentrifugation was performed at 48,000 rpm and 20 °C for the 0.3-mg/ml protein samples in 50 mM Hepes/KOH pH 7.5, 0.2 M KCl, 20 mM $MgCl_2$, 1 mM EDTA, 2 mM β-mercaptoethanol. The apparent sedimentation coefficient distributions g(s*) are shown for wild type ClpB (thick solid line), 4G (thin solid line), DelG (dotted line), and 2P (broken line).

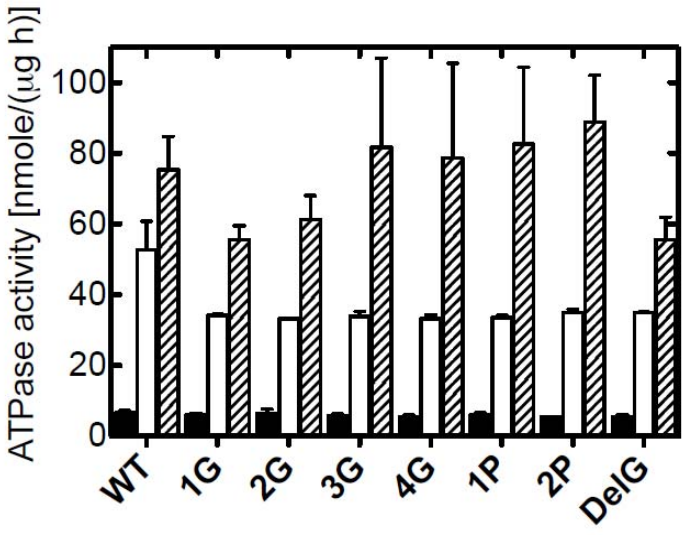


Fig. 4 ATPase activity of the modified linker variants of ClpB. The hydrolysis of ATP catalyzed by the ClpB variants was determined at 37 °C in the absence of other proteins (black bars), with 0.1 mg/ml α-casein (white bars), and with 0.04 mg/ml poly-lysine (hatched bars). The average values from three separate experiments are shown with the standard deviations.

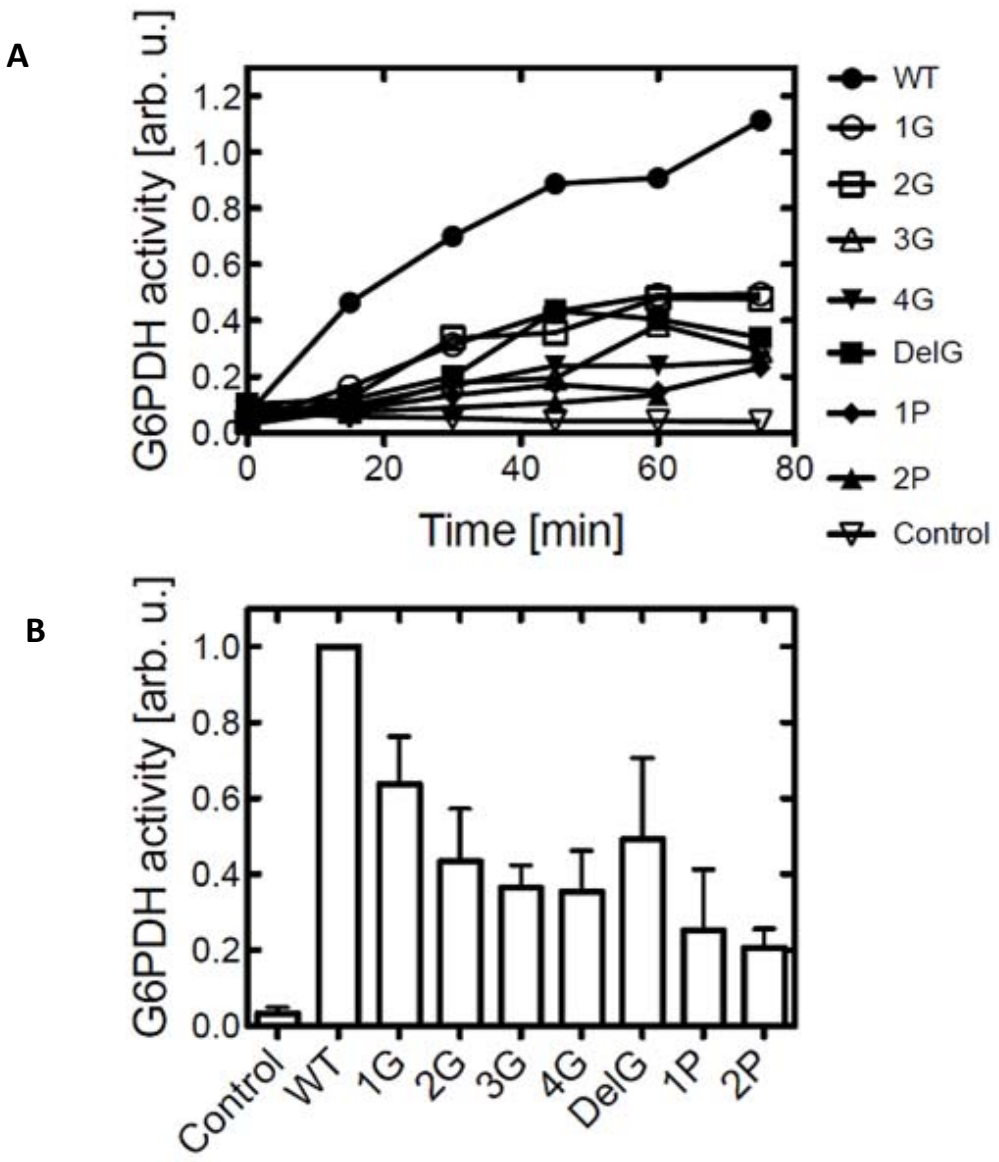


Fig. 5 Reactivation of aggregated glucose-6-phosphate dehydrogenase in the presence of ClpB and DnaK/DnaJ/GrpE. (A) A representative time-course of the reactivation of aggregated G6PDH without the chaperones (control) and with the indicated ClpB linker variants and DnaK/DnaJ/GrpE. (B) Relative changes in the G6PDH reactivation rates for the modified linker ClpB variants measured after 60 min of the aggregate reactivation. The average values from three independent experiments are shown with the standard deviations.

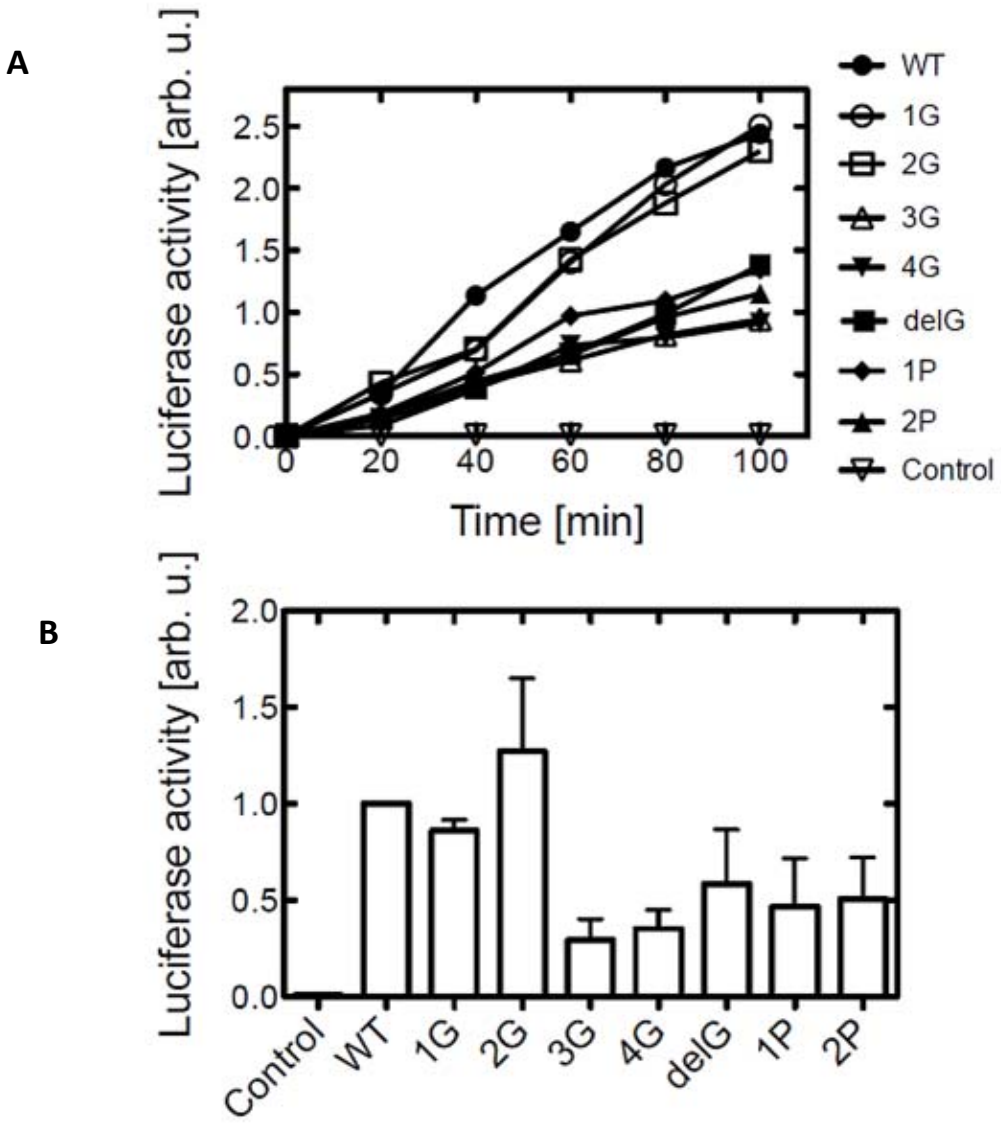


Fig. 6 Reactivation of aggregated firefly luciferase in the presence of ClpB and DnaK/DnaJ/GrpE. (A) A representative time-course of the reactivation of aggregated luciferase without the chaperones (control) and with the indicated ClpB linker variants and DnaK/DnaJ/GrpE. (B) Relative changes in the luciferase reactivation rates for the modified linker ClpB variants measured after 60 min of the aggregate reactivation. The average values from three independent experiments are shown with the standard deviations.

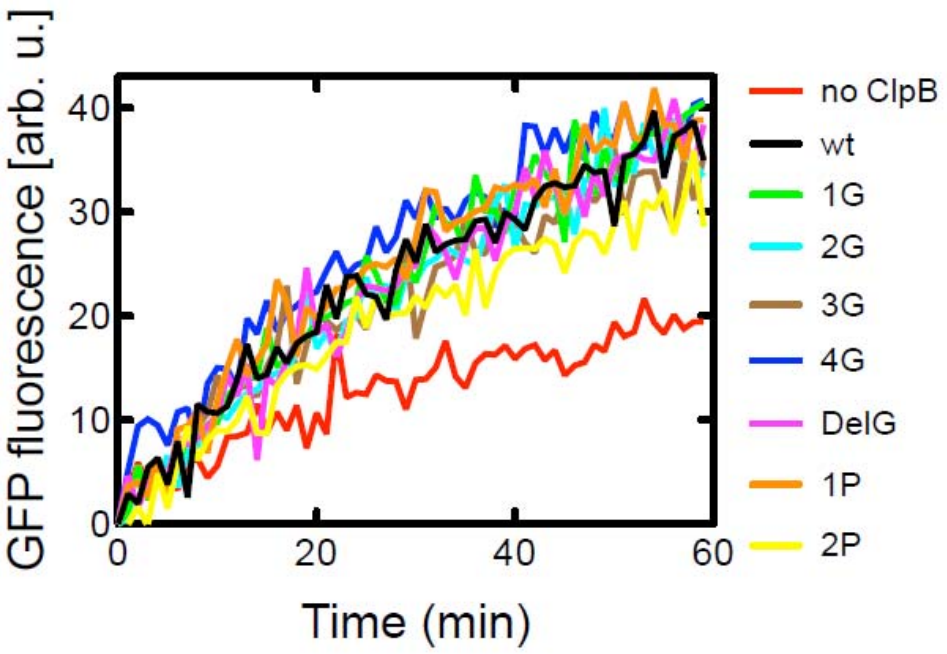


Fig. 7 Time-course of the reactivation of aggregated GFP in the presence of wt ClpB or its modified linker variants and in the absence of the co-chaperones.

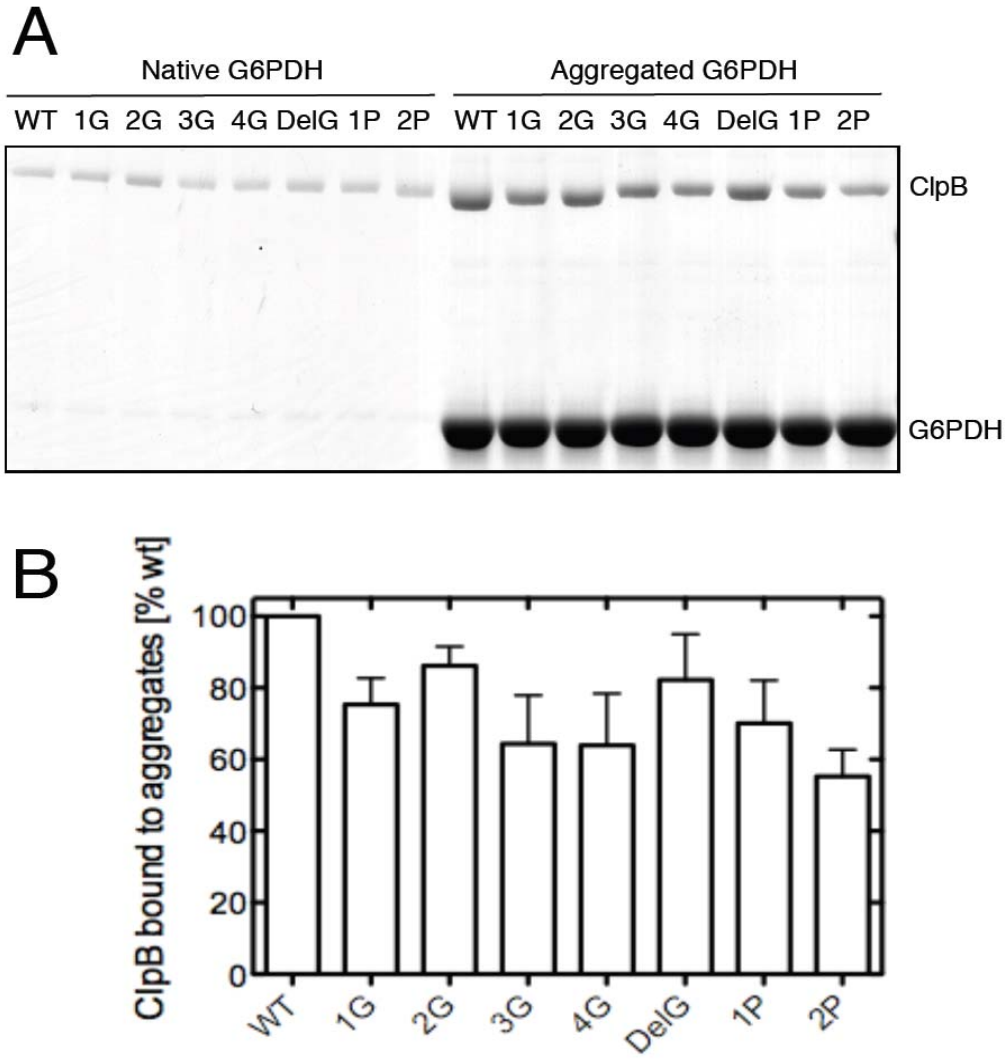


Fig. 8 Interactions of ClpB with aggregated G6PDH. Wt ClpB and its modified linker variants were incubated with native or aggregated G6PDH in the presence of ATPγS. The solutions were passed through a 0.1- μm filter and the fractions retained on the filter were analyzed by SDS-PAGE with Coomassie stain. (A) A representative gel from the filtration experiment. (B) Image-intensity analysis of the ClpB band in panel A normalized for the amount of G6PDH retained on the filter. The average relative amounts of ClpB bound to the aggregates from three independent experiments are shown with the standard deviations.

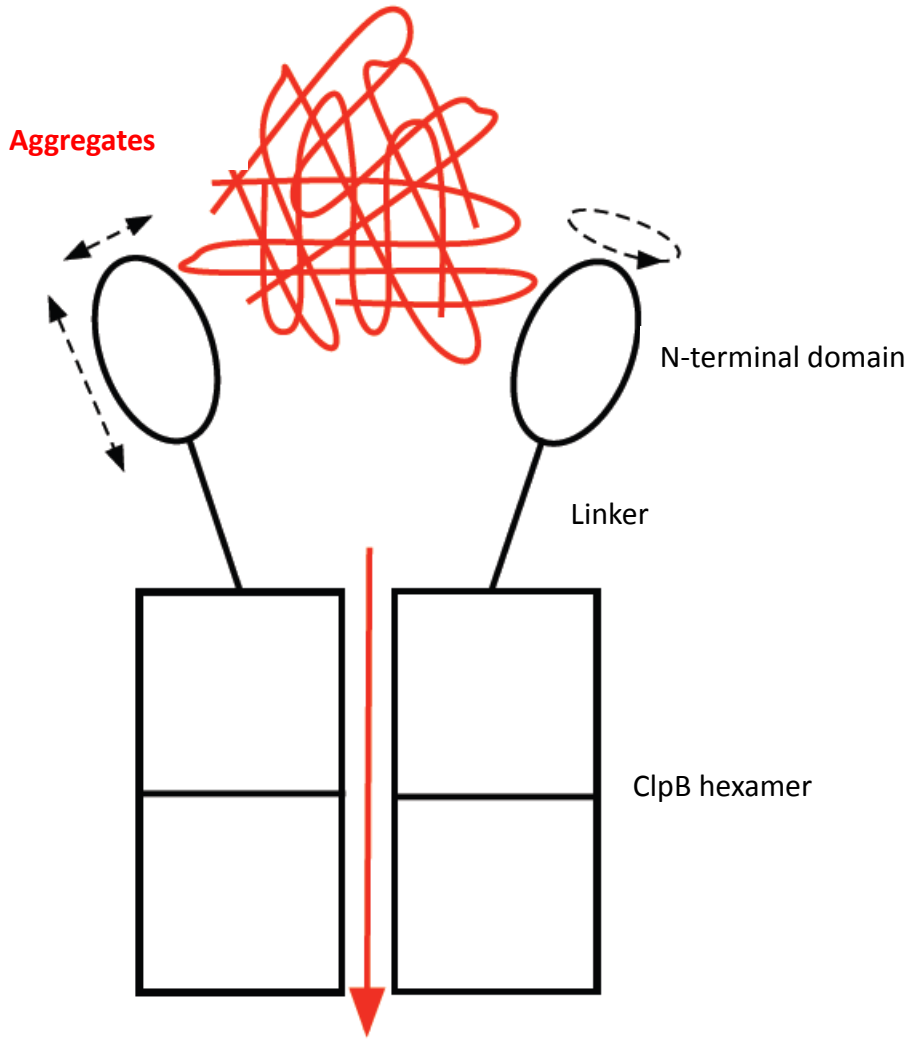


Fig. 9 Model of the role of the ClpB N-terminal domain mobility in aggregate reactivation. For simplicity, two out of six ClpB subunits of a hexamer are shown. D1 and D2 AAA+ modules are shown as squares and the N-terminal domain is shown as an oval. The interaction of an aggregated substrate (red) with the N-terminal domains of the hexameric ClpB may precede insertion of the substrate into the ClpB channel and its forced unfolding and extraction from the aggregate (red arrow). Motions of the N-terminal domains (black arrows) supported by the mobility of the linker may facilitate the search for ClpB-binding motifs at the aggregate surface and an efficient insertion of the substrate into the ClpB channel.

Chapter 3

Characterization of ClpB from *Ehrlichia chaffeensis*

3.1 Introduction

The molecular chaperone ClpB from *E. coli* (as well as HSP104 from yeast) has been studied intensively in recent years for its unique function in reactivating protein aggregation. However, the chaperone function of ClpB could be a double edged sword. On one hand, when cells are under severe stress and protein misfolding overwhelms the capacity of other unfolding chaperones, ClpB is up-regulated to reactivate aggregated proteins, and this machinery could give us insights into the possible treatment of aggregate related diseases such as Alzheimer's disease. On the other hand, ClpB is also up-regulated in many pathogenic bacteria during their invasion of the host cell, which helps the pathogen to overcome the stress from the immune system of the host cell.

A *clpb* mutant from *Leptospira interrogans* was shown to be attenuated in virulence in an animal model of acute leptospirosis, suggesting its important role in maintaining the virulence of the pathogen (Lourdault, Cerqueira et al. 2011). Also, mutated ClpB made the pathogen less virulent comparing to their WT in *Salmonella enterica* serovar Typhimurium, *Mycoplasma pneumoniae*, *Porphyromonas gingivalis*, *Listeria monocytogenes*, and *Francisella tularensis* (Turner, Lovell et al. 1998; Chastanet, Derre et al. 2004; Yuan, Rodrigues et al. 2007; Kannan, Musatovova et al. 2008; Meibom, Dubail et al. 2008). ClpB is also associated with intramacrophage growth with *Francisella novicida* (Havlasova, Hernychova et al. 2005). ClpB is not the only chaperone that is involved in pathogenic bacteria invasion. The highly conserved Clp family has been found important for virulence properties of many pathogenic bacteria. For example, ClpC and ClpXP are essential for survival under heat and entry into host epithelial cells in *Porphyromonas gingivalis* (Capestany, Tribble et al. 2008). All these data suggest that protein chaperones are playing

important roles in pathogenic bacteria invasion and virulence. Since ClpB only exists in bacteria, not in humans, and considering its role in pathogen invasion, it is reasonable to consider it as a potential therapeutic target.

In this chapter, we purified recombinant ClpB from *Ehrlichia chaffeensis* and studied its biochemical properties. *E. chaffeensis* is an obligatory intracellular endosomal gram-negative bacterium that causes human monocytic ehrlichiosis. A preliminary study in the laboratory of Dr Roman Ganta at KSU showed that ClpB was up-regulated during infection of *E. chaffeensis* (Fig. 1), which confirmed the possible role of ClpB in the invasion of the bacterium.

From the sequence alignment of ClpB from different species (Fig. 2), we can see two conserved NBDs in *E. chaffeensis* ClpB (Eh_B), making it a class II AAA+ protein. Apart from the conserved NBDs, the total sequence identity between Eh_B and Ec_B is only about 56%. The N-terminal domain and the M domain share a few conserved amino acid residues, and notably there is also a conserved glycine in the linker of Eh_B.

In this chapter, we studied the oligomerization, ATPase activity, chaperone activity and substrate binding of recombinant Eh_B. We found that Eh_B could reactivate aggregated protein in the presence of HSP70 from *E. coli* similar to Ec_B; however, the different substrate binding pattern of Eh_B may suggest a different disaggregation mechanism.

3.2 Materials and methods.

Proteins and Aggregates. Co-chaperones (DnaK, DnaJ, GrpE) were purchase from MBL International (Woburn, MA). G6PDH from *Leuconostoc mesenteroides* was purified as described in Chapter 2 and α-casein was obtained from Sigma (St. Louis, MO). Firefly luciferase was obtained from Promega (Madison, WI). Protein concentrations were determined spectrophotometrically.

To produce aggregates of G6PDH, the protein stock (324 μM) was diluted 2-fold with unfolding buffer (10 M urea, 16% glycerol and 40 mM DTT) and was incubated

at 47 °C for 5 min. The mixture was then diluted 10-fold with refolding buffer 1 (50 mM Tris/HCl pH 7.5, 20 mM $Mg(OAc)_2$, 30 mM KCl, 1 mM EDTA, and 1 mM β-mercaptoethanol) and was incubated at 47 °C for 15 min and then on ice for 2 min. To produce aggregates of luciferase, 216 μM luciferase stock was diluted 300-fold with PBS containing 1 mg/ml BSA and then was incubated at 45 °C for 12 min. To produce aggregated GFP, 4.5 μM protein was heated for 10 min at 80 ˚C.

Western Blot. Similar amount of Eh_B and Ec_B were applied to SDS–PAGE, proteins were transferred onto a nitrocellulose membrane and detected by Western blotting using rabbit polyclonal anti-Ec_ClpB antibody (produced by Cocalico Biologicals, Reamstown, PA) and an HRP-coupled goat anti-rabbit secondary antibody (Pierce, Rockford, IL), and visualized by SuperSignal West Pico Chemiluminescent Substrate from Pierce.

Analytical Ultracentrifugation. A Beckman XL-I analytical ultracentrifuge was used in sedimentation velocity experiments with two-channel analytical cells. The data were analyzed using the time-derivative method (Stafford 1992; Stafford III 1994) and the software distributed with the instrument.

ClpB ATPase activity The ClpB variants were incubated in assay buffer (100 mM Tris/HCl pH 8.0, 1 mM DTT, 1 mM EDTA, 10 mM $MgCl_2$, and 5 mM ATP) at 37 °C for 15 min without or with 0.1 mg/ml α-casein or 0.04 mg/ml poly-lysine or 1.6 μM aggregated G6PDH. The concentration of ClpB was 0.05 mg/ml for the basal activity and in the presence of α-casein or 0.005 mg/ml in the presence of poly-lysine. The phosphate concentration generated by ClpB was measured as described before (Zolkiewski 1999).

Aggregate Reactivation Assays Aggregated G6PDH (16.2 μM) was diluted 10-fold with refolding buffer 1 containing 1.5 μM ClpB, 1 μM DnaK, 1 μM DnaJ, 0.5

μM GrpE and 6 mM ATP. The mixture was incubated at 30 °C and aliquots of the mixture were withdrawn to test the recovery of the G6PDH enzymatic activity. Aggregates diluted with refolding buffer without the chaperones were used as control. To measure the G6PDH activity, aliquots from the refolding reaction were incubated in 50 mM Tris/HCl pH 7.8, 5 mM $MgCl_2$, 1.5 mM G6P and 1 mM $NADP^+$ for 10 min followed by the measurement of absorption at 340 nm. Aggregated luciferase (0.7 μM) was diluted 20-fold with refolding buffer 2 (30 mM Hepes, pH 7.65, 120 mM KCl, 10 mM $MgCl_2$, 6 mM ATP, 1 mM EDTA, 10 mM DTT, 0.1 mg/ml BSA) containing 1.5 μM ClpB, 1 μM DnaK, 1 μM DnaJ, and 0.5 μM GrpE. The mixture was incubated at room temperature and aliquots were withdrawn to test the recovery of the luciferase activity using the luminescence assay system (Promega, Madison, WI). GFP reactivation assays (100 μL) were performed in 20 mM Tris-HCl, pH 7.5, 100 mM KCl, 5 mM DTT, 0.1 mM EDTA, and 10% glycerol (vol/vol) with 2 mM ATP and 2 mM ATPγS, an ATP regenerating system (20 mM creatine phosphate and 6 μg creatine kinase), 10 mM $MgCl_2$, 10 μL heat-aggregated GFP and 1.0 μM ClpB. Reactions were initiated by the addition of Mg-ATP and reactivation was monitored over time at 23 °C using a Perkin-Elmer LS50B luminometer with a plate reader. Excitation and emission wavelengths were 395 nm and 510 nm, respectively.

ClpB-Aggregate Interaction Assay. Aggregated G6PDH or luciferase (same concentration as in the reactivation assay) was diluted 10 times with the refolding buffer 1 containing 1.5 μM ClpB and 6 mM nucleotides. The mixture was incubated at 30 °C with 600 rpm shaking for 20 min and then was applied to the filter device (Millipore Ultrafree-MC Centrifugal Filter Unit with the membrane, pore size 0.1 μm). After 5 min incubation at room temperature, the filter device was centrifuged at 13,000 rpm for 4 min to get the flow-through fractions. The filter device was washed with the refolding buffer 1 containing ATPγS at 30 °C for 5 min and then re-centrifuged. Next, 1x SDS loading buffer was added to the filter device and the filter device was incubated at 50 °C for 5 min with shaking. Then, it was centrifuged

to obtain the eluate fractions, which were applied to SDS-PAGE. The Coomassie-stained band intensity was determined with the BandScan software (http://bandscan.software.informer.com).

3.3 Results

3.3.1 Up-regulation of clpb gene during infection of macrophages with Ehrlichia chaffeensis

Semi-quantitative RT-PCR showed that after 6 hrs of infection, the level of *eh_b* mRNA level increased and reached maximum at about 30hrs post infection (Fig. 1), suggesting the involvement of ClpB function during infection.

3.3.2 Expression and purification of Eh_B protein

The *eh_b* gene was cloned into a pET28 vector with a his-tag on the N-terminus. The vector was first transformed into normal BL21(DE3) cells for protein expression. However, after induction by IPTG for about 2 hrs, the protein level significantly dropped and lowering the culture temperature did not improve the expression level (Fig. 3 top panel). We noticed that *E. chaffeensis* has a different codon usage preference comparing to *E. coli*, especially for leucine, isoleucine, and arginine (Fig. 3 bottom). Thus a special *E. coli* strain designed for rare codon usage protein expression was used and the successful expression of Eh_B was confirmed by SDS-PAGE (Fig. 3 top). The cells were harvested, sonicated and the soluble Eh_B protein was purified using nickel column and further purified by gel filtration. Previous work with Ec_B showed that a his tag on the N-terminal domain will inhibit the ATPase activity of the protein (Chow, Barnett et al. 2005), thus the his-tag was removed by thrombin after purification.

We tested the recognition of Eh_B by anti-Ec_B antibody. As shown in Fig. 4, the Eh_B cannot be recognized by Ec_B antibody, suggesting significant epitope differences between the two proteins.

3.3.3 Self-association of Eh_B

Many AAA+ family members share the common feature of assembly into a ring-shaped barrel with a central channel in the middle (Sauer and Baker 2011). The self-association of ClpB is tightly controlled by protein concentration and could be enhanced by the presence of nucleotide (Chow, Barnett et al. 2005).

The self-association of Eh_B was first analyzed by gel filtration. Low concentrations of the protein were analyzed with/without ATPγS. The elution peak of Eh_B was significantly shifted after ATPγS was provided, indicating the formation of hexamer; however the shift was not as great as for Ec_B. Considering the fact that the sizes of two proteins are very similar, this result suggested that Eh_B was forming a less stable hexamer.

We also tested the oligomerization of Eh_B by sedimentation velocity. In the condition that Ec_B will maintain a mixture of both monomer and hexamer, Eh_B stayed mainly monomeric, which was consistent with the gel filtration data that self-association of Eh_B was not as strong as that of Ec_B. When ATPγS was provided, distributions of the sedimentation coefficients g(s*) shifted from 4S (monomeric) to 14S (oligomeric), indicating hexamers were the main component in the solution (Fig. 5). Thus, we conclude that the oligomerization of Eh_B does occur but the self-association was weaker compared to Ec_B, suggesting a more dynamic hexamer.

3.3.4 ATPase activity of Eh_B

ATPase activity was tested under the same conditions as Ec_B. Eh_B had similar basal ATPase activity as Ec_B, which further confirmed the successful self-association (Fig. 6). However, although maintaining a higher basal ATPase activity than Ec_B, Eh_B failed to respond to activation by α-casein and poly-lysine (Fig. 6). α-casein activates ATPase activity of Ec_B by mimicking the structure conformation of a substrate. Thus we also tested whether a real aggregated substrate can activate the ATPase activity of both ClpB homologs. As we can see in Fig. 6,

unlike the pseudo substrate α-casein or poly-lysine, aggregated G6PDH activated both Ec_B and Eh_B to a similar extent.

3.3.5 Chaperone activity of Eh_B

Next we tested the chaperone activity of Eh_B. From Fig. 7, we can see that Eh_B could reactivate aggregated luciferase as effective as Ec_B in the presence of *E. coli* HSP70 system. Surprisingly, with the large aggregates of G6PDH, Eh_B worked much better than Ec_B. Studies with HSP104 indicated the species specificity of HSP70 (Miot, Reidy et al. 2011). In our case, it seemed that Eh_B can work with *E. coli* HSP70 just as well as Ec_B, or even better. This is consistent with the previous result of higher ATPase activity of Eh_B but also suggests differences in the working mechanism of Eh_B.

3.3.6 Substrate binding of Eh_B

Considering the significantly different reactivation rate of G6PDH by Eh_B, we carried out the filter assay described in Chapter 2 to monitor the binding of Eh_B to G6PDH aggregates. Ec_B bound to substrate in the presence of ATPγS (Fig. 8), which makes sense since ATPγS keeps ClpB in the ATP bound state/substrate binding state. However, Eh_B showed an unusual binding in the presence of ATP instead of ATPγS. It is possible that the dynamic feature of Eh_B requires an active ATP hydrolysis cycle to maintain the substrate binding. Thus, all the data suggested a different working model of the Eh_B machinery.

3.3.7 Chaperone activity of Eh_B in the absence of HSP70 co-chaperones

Previous studies have shown that when substrate can be held by Ec_B during the ATP turnover, which is induced by trap mutations or providing both ATP and ATPγS at the same time, Ec_B can reactivate aggregates without co-chaperones (Doyle, Shorter et al. 2007; Hoskins, Doyle et al. 2009). Since Eh_B binds to its substrates in the presence of ATP, we wondered if Eh_B could reactivate aggregates without

co-chaperones. From Fig. 7 panel B and C, we can see that Eh_B reactivated aggregated G6PDH but not aggregated luciferase without DnaK system. We hypothesize that, with G6PDH aggregate, Eh_B can hold the substrate while hydrolyzing ATP, which gives Eh_B the ability to pull the aggregate through the central channel. In this case, the *E. coli* DnaK system may serve as a chaperone itself to help the released unfolded polypeptide to refold. On the other hand we still cannot exclude the possibility that collaboration exists between Eh_B and *E. coli* DnaK system since Eh_B cannot reactivate aggregated luciferase without DnaK. This DnaK-independent chaperone activity of Eh_B may be substrate specific.

3.3.8 Eh_B cannot rescue clpb deletion E. coli strain under heat shock

ClpB in E. coli is essential for survival under heat shock. A plasmid containing *eh_b* gene was transformed into a *clpb* deletion *E. coli* strain. The *clpb* deletion *E. coli* strain could not be rescued by *eh_b* gene under heat shock as *ec_b* gene, which could be further prove of the difference between two ClpBs (experiment performed by Dr Sabina Kedzierska at University of Gdańsk, data not shown).

3.4 Discussion

In this Chapter, the *eh_b* gene was cloned and the recombinant protein was purified and its biochemical properties were tested. Eh_B was found to be an active ATPase that shares the similar function of reactivating aggregates as Ec_B, however, with a different working pattern.

When pathogen invasion takes place, both the pathogen and the host will up-regulate many stress-related genes. The host cell turns on the immune response to defend itself and up-regulates the stress-related genes, mostly chaperones, to fight the stress caused by the intruders; on the other hand, the pathogen has to pass through the guard line of the host's immune system, making its chaperones important for survival. ClpB is one of the many chaperones that are up-regulated in *E. chaffeensis* during infection and because of its lack of existence in mammals, ClpB may serve as a viable

target for treatment of Human monocytic erlichiosis.

E. chaffeensis would be expected to undergo more immune stress compared to *E. coli* considering the fact that it is pathogenic. Thus, despite the conservation in the NBDs, Eh_B and Ec_B are similar yet significantly different. In the purification we found the codon usage of two proteins was different. Rare codon usage has been suggested to regulate protein synthesis rate in order to allow the newly synthesized peptide to fold in to proper secondary and tertiary structure (Angov 2011). Most of the rare codon occurs in leucine, isoleucine and arginine, and these residues are pretty abundant in the conserved region. It is possible that due to the different living environment, synthesis of Eh_B protein is tightly controlled by rare codon usage.

The gel filtration and sedimentation velocity experiments further confirmed that the two proteins were different. The oligomerization of Eh_B was successful but suggested weaker association between monomers; in other words, the hexamer of Eh_B was very dynamic. The equilibrium between monomer and hexamer could be pulled to the direction of hexamer by the presence of nucleotide, which is consistent with many other members in the AAA+ family. This dynamic feature of the Eh_B hexamer may also explain the results in substrate binding experiment where Eh_B bound to substrate in the presence of ATP. It is possible that Eh_B hydrolyzes ATP very fast in the presence of substrate and blocking the ATP hydrolysis cycle will abandon the dynamic cycle of hexamerization, thus abolishing substrate binding. However, further experiments showed that this was not the case. ATP hydrolysis cycle of Eh_B was activated about two-fold when aggregates were provided. Importantly, the ability of holding substrate during ATP turnover provides Eh_B the ability of disaggregation without HSP70 system.

Eh_B had a much higher reactivation rate with G6PDH aggregates. This effect was not visible with the luciferase aggregate. In the previous linker study, the defect was also not so obvious in the luciferase reactivation compared to G6PDH reactivation. It is possible that the influence on aggregate reactivation was amplified with strongly aggregated G6PDH comparing the “weaker” luciferase aggregates. Also,

this could be explained by the latter experiment, where Eh_B could reactivate G6PDH aggregate but not aggregated luciferase without KJE.

It is notable that in all the reactivation assays, Eh_B was cooperating with HSP70 (KJE) from *E. coli*. Studies of ClpB and HSP104 have shown that the HSP70 system was not interchangeable, that is to say, the HSP70 system was species specific (Miot, Reidy et al. 2011). This species specificity is related to the middle domain of ClpB, which interacts with HSP70 (Miot, Reidy et al. 2011). In our case, this species specificity was not observed. From the sequence alignment we can see that sequential correlation between the middle domain of Ec_B, Eh_B and HSP104 are very similar (Fig. 2). So what is the mystery in the interaction of HSP100 and HSP70? It is possible that the species specificity rule only applies when the species are far away enough on the evolutionary tree. To confirm this hypothesis, reactivation with HSP70 from different species is required. Another possibility is that the HSP70 was not collaborating with Eh_B in the reactivation; instead, they were helping the refolding of released polypeptide by Eh_B since Eh_B can reactivate aggregates without the help of co-chaperones.

In summary, Eh_B shares similarity with Ec_B in their ability to hydrolyze ATP and reactivate aggregates, but their mechanisms appeared surprisingly different. The Eh_B hexamer was more dynamic and reactivation was much faster. For treatment of HME caused by *Ehrlichia chaffeensis*, interfering with the function of the *eh_b* gene will provide a new direction. Thus, further investigation is needed to better understand how the Eh_B machinery works.

References

Angov, E. (2011). "Codon usage: nature's roadmap to expression and folding of proteins." Biotechnol J **6**(6): 650-659.

Capestany, C. A., G. D. Tribble, et al. (2008). "Role of the Clp system in stress tolerance, biofilm formation, and intracellular invasion in *Porphyromonas gingivalis*." J Bacteriol **190**(4): 1436-1446.

Chastanet, A., I. Derre, et al. (2004). "clpB, a novel member of the *Listeria monocytogenes* CtsR regulon, is involved in virulence but not in general stress tolerance." J Bacteriol **186**(4): 1165-1174.

Chow, I. T., M. E. Barnett, et al. (2005). "The N-terminal domain of *Escherichia coli* ClpB enhances chaperone function." FEBS Lett **579**(20): 4242-4248.

Doyle, S. M., J. Shorter, et al. (2007). "Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity." Nat Struct Mol Biol **14**(2): 114-122.

Havlasova, J., L. Hernychova, et al. (2005). "Proteomic analysis of anti-*Francisella tularensis* LVS antibody response in murine model of tularemia." Proteomics **5**(8): 2090-2103.

Hoskins, J. R., S. M. Doyle, et al. (2009). "Coupling ATP utilization to protein remodeling by ClpB, a hexameric AAA+ protein." Proc Natl Acad Sci U S A **106**(52): 22233-22238.

Kannan, T. R., O. Musatovova, et al. (2008). "Characterization of a unique ClpB protein of *Mycoplasma pneumoniae* and its impact on growth." Infect Immun **76**(11): 5082-5092.

Lourdault, K., G. M. Cerqueira, et al. (2011). "Inactivation of clpB in the pathogen *Leptospira interrogans* reduces virulence and resistance to stress conditions." Infect Immun **79**(9): 3711-3717.

Meibom, K. L., I. Dubail, et al. (2008). "The heat-shock protein ClpB of *Francisella tularensis* is involved in stress tolerance and is required for multiplication

in target organs of infected mice." Mol Microbiol **67**(6): 1384-1401.

Miot, M., M. Reidy, et al. (2011). "Species-specific collaboration of heat shock proteins (Hsp) 70 and 100 in thermotolerance and protein disaggregation." Proc Natl Acad Sci U S A **108**(17): 6915-6920.

Sauer, R. T. and T. A. Baker (2011). "AAA+ proteases: ATP-fueled machines of protein destruction." Annu Rev Biochem **80**: 587-612.

Stafford III, W. F., Ed. (1994). Sedimentation boundary analysis of interacting systems: Use of the apparent sedimentation coefficient distribution function. Modern Analytical Ultracentrifugation. Boston, Birkhauser.

Stafford, W. F., 3rd (1992). "Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile." Analytical biochemistry **203**(2): 295-301.

Turner, A. K., M. A. Lovell, et al. (1998). "Identification of *Salmonella typhimurium* genes required for colonization of the chicken alimentary tract and for virulence in newly hatched chicks." Infect Immun **66**(5): 2099-2106.

Yuan, L., P. H. Rodrigues, et al. (2007). "The *Porphyromonas gingivalis* clpB gene is involved in cellular invasion in vitro and virulence in vivo." FEMS Immunol Med Microbiol **51**(2): 388-398.

Zolkiewski, M. (1999). "ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from *Escherichia coli*." The Journal of biological chemistry **274**(40): 28083-28086.

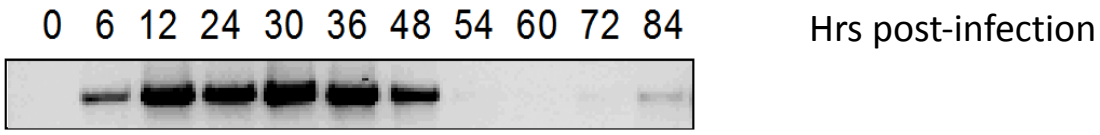
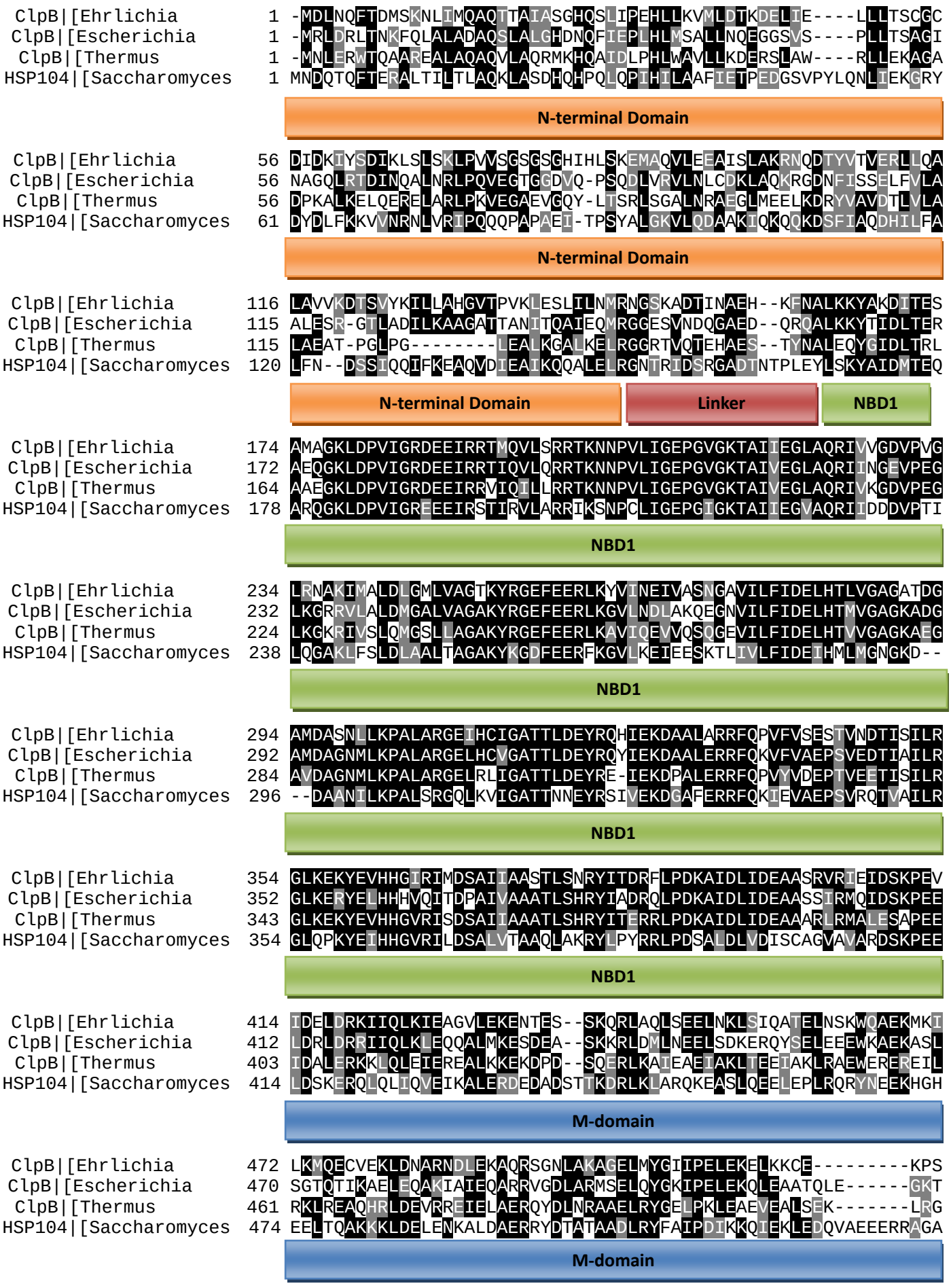


Fig. 1 Semi-quantitative RT-PCR of *Ehrlichia chaffeensis clpb* gene. ClpB mRNA level in *Ehrlichia chaffeensis* is found to increase after infection.

Experiment performed by Chuanmin Cheng at Dr Roman Ganta's laboratory at Kansas State University.


ClpB|[Ehrlichia 1 -MDLNQFTDMSKNLIMQAQTTAIASGHQSLIPEHLLKVMLDTKDELIE----LLLTSCGC
ClpB|[Escherichia 1 -MRLDRLTNKFQLALADAQSLALGHDNQFIEPLHLMSALLNQEGGSVS----PLLTSAGI
ClpB|[Thermus 1 -MNLERWTQAAREALAQAQVLAQRMKHQAIDLPHLWAVLLKDERSLAW----RLLEKAGA
HSP104|[Saccharomyces 1 MNDQTQFTERALTILTLAQKLASDHQHPQLQPIHILAAFIETPEDGSVPYLQNLIEKGRY
N-terminal Domain
ClpB|[Ehrlichia 56 DIDKIYSDIKLSLSKLPVVSGSGSGHIHLSKEMAQVLEEAISLAKRNQDTYVTVERLLQA
ClpB|[Escherichia 56 NAGQLRTDINQALNRLPQVEGTGGDVQ-PSQDLVRVLNLCDKLAQKRGDNFISSELFVLA
ClpB|[Thermus 56 DPKALKELQERELARLPKVEGAEVGQY-LTSRLSGALNRAEGLMEELKDRYVAVDTLVLA
HSP104|[Saccharomyces 61 DYDLFKKVVNRNLVRIPQQQPAPAEI-TPSYALGKVLQDAAKIQKQQKDSFIAQDHILFA
N-terminal Domain
ClpB|[Ehrlichia 116 LAVVKDTSVYKILLAHGVTPVKLESLILNMRNGSKADTINAEH--KFNALKKYAKDITES
ClpB|[Escherichia 115 ALESR-GTLADILKAAGATTANITQAIEQMRGGESVNDQGAED--QRQALKKYTIDLTER
ClpB|[Thermus 115 LAEAT-PGLPG--------LEALKGALKELRGGRTVQTEHAES--TYNALEQYGIDLTRL
HSP104|[Saccharomyces 120 LFN--DSSIQQIFKEAQVDIEAIKQQADLELRGNTRIDSRGADTNTPLEYLSKYAIDMTEQ
N-terminal Domain
Linker
NBD1
ClpB|[Ehrlichia 174 AMAGKLDPVIGRDEEIRRTMQVLSRRTKNNPVLIGEPGVGKTAIIEGLAQRIVVGDVPVG
ClpB|[Escherichia 172 AEQGKLDPVIGRDEEIRRTIQVLQRRTKNNPVLIGEPGVGKTAIVEGLAQRIINGEVPEG
ClpB|[Thermus 164 AAEGKLDPVIGRDEEIRRVIQILLRRTKNNPVLIGEPGVGKTAIVEGLAQRIVKGDVPEG
HSP104|[Saccharomyces 178 ARQGKLDPVIGREEEIRSTIRVLARRIKSNPCLIGEPGIGKTAIIEGVAQRIIDDDVPTI
NBD1
ClpB|[Ehrlichia 234 LRNAKIMALDLGMLVAGTKYRGEFEERLKYVINEIVASNGAVILFIDELHTLVGAGATDG
ClpB|[Escherichia 232 LKGRRVLALDMGALVAGAKYRGEFEERLKGVLNDLAKQEGNVILFIDELHTMVGAGKADG
ClpB|[Thermus 224 LKGKRIVSLQMGSLLAGAKYRGEFEERLKAVIQEVVQSQGEVILFIDELHTVVGAGKAEG
HSP104|[Saccharomyces 238 LQGAKLFSLDLAALTAGAKYKGDFEERFKGVLKEIEESKTLIVLFIDEIHMLMGNGKD--
NBD1
ClpB|[Ehrlichia 294 AMDASNLLKPALARGEIHCIGATTLDEYRQHIEKDAALARRFQPVFVSESTVNDTISILR
ClpB|[Escherichia 292 AMDAGNMLKPALARGELHCVGATTLDEYRQYIEKDAALERRFQKVFVAEPSVEDTIAILR
ClpB|[Thermus 284 AVDAGNMLKPALARGELRLIGATTLDEYRE-IEKDPALERRFQPVYVDEPTVEETISILR
HSP104|[Saccharomyces 296 --DAANILKPALSRGQLKVIGATTNNEYRSIVEKDGAFERRFQKIEVAEPSVRQTVAILR
NBD1
ClpB|[Ehrlichia 354 GLKEKYEVHHGIRIMDSAIIAASTLSNRYITDRFLPDKAIDLIDEAASRVRIEIDSKPEV
ClpB|[Escherichia 352 GLKERYELHHHVQITDPAIVAAATLSHRYIADRQLPDKAIDLIDEAASSIRMQIDSKPEE
ClpB|[Thermus 343 GLKEKYEVHHGVRISDSAIIAAATLSHRYITERRLPDKAIDLIDEAAARLRMALESAPEE
HSP104|[Saccharomyces 354 GLQPKYEIHHGVRILDSALVTAAQLAKRYLPYRRLPDSALDLVDISCAGVAVARDSKPEE
NBD1
ClpB|[Ehrlichia 414 IDELDRKIIQLKIEAGVLEKENTES--SKQRLAQLSEELNKLSIQATELNSKWQAEKMKI
ClpB|[Escherichia 412 LDRLDRRIIQLKLEQQALMKESDEA--SKKRLDMLNEELSDKERQYSELEEEWKAEKASL
ClpB|[Thermus 403 IDALERKKLQLEIEREALKKEKDPD--SQERLKAIEAEIAKLTEEIAKLRAEWERERETL
HSP104|[Saccharomyces 414 LDSKERQLQLIQVEIKALERDEDADSTTKDRLKLARQKEASLQEELEPLRQRYNEEKHGH
M-domain
ClpB|[Ehrlichia 472 LKMQECVEKLDNARNDLEKAQRSGNLAKAGELMYGIIPELEKELKKCE---------KPS
ClpB|[Escherichia 470 SGTQTIKAELEQAKIAIEQARRVGDLARMSELQYGKIPELEKQLEAATQLE------GKT
ClpB|[Thermus 461 RKLREAQHRLDEVRREIELAERQYDLNRAAELRYGELPKLEAEVEALSEK-------LRG
HSP104|[Saccharomyces 474 EELTQAKKKLDELENKALDAERRYDTATAADLRYFAIPDIKKQIEKLEDQVAEEERRAGA
M-domain

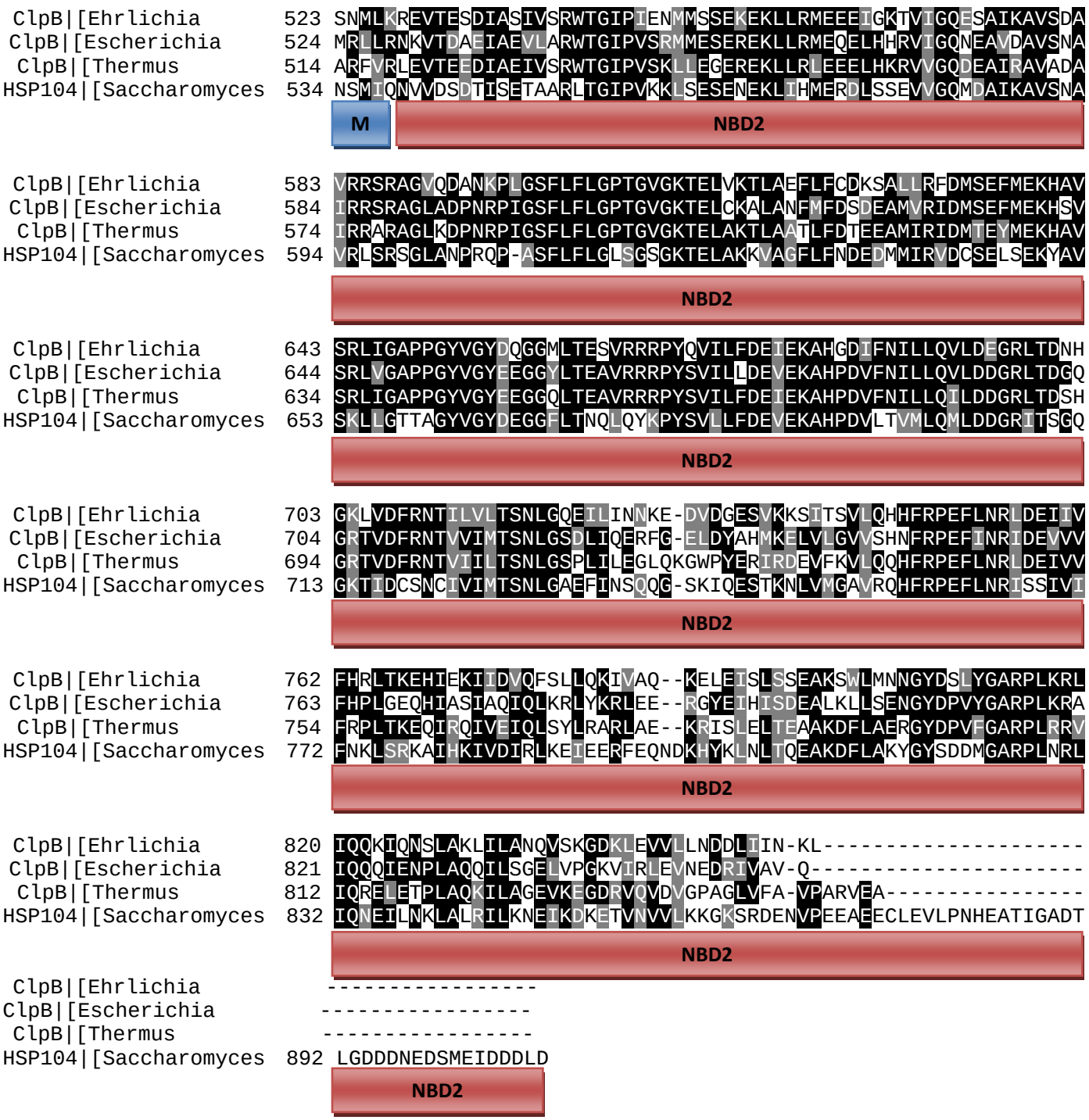


Fig. 2 Sequence alignment of ClpB from *Ehrlichia chaffeensis*, *Escherichia coli*, *Thermus thermophilus*, and HSP104 from *Saccharomyces cerevisiae*. The AAA domains of all ClpBs are strongly conserved. The main difference lies in the N-terminal domain and the Middle domain.

Sequences were aligned using Clustal Omega and alignment file was re-arranged using BOXSHADE. Black background represents identical amino acid residues and grey represents similar ones.

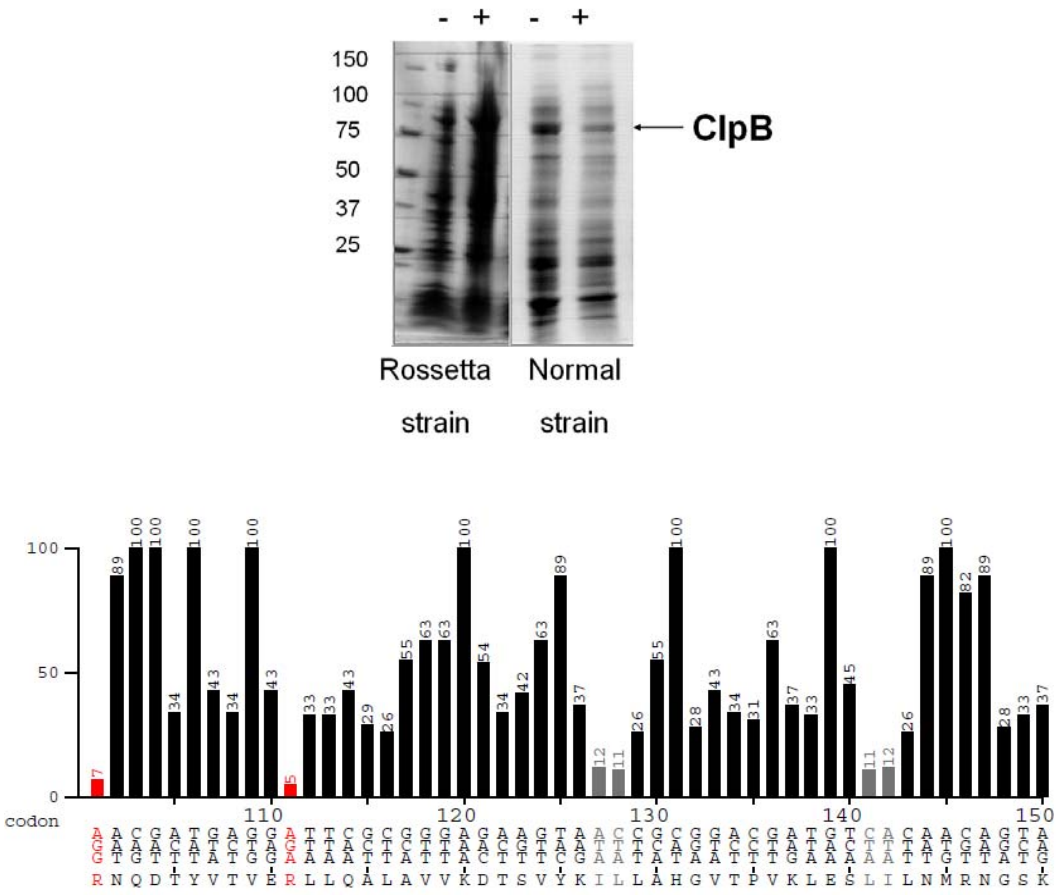


Fig. 3 Top: Expression of Eh_B protein in different *E. coli* strains. +: induced by IPTG for 2hrs; -: uninduced.

Bottom: Part of the condon usage graph created by graphical codon usage analyser (http://gcua.schoedl.de/index.html). The x axial shows the translation of *eh_b* gene; the y axial number shows the fraction of each codon used in *Ehrlichia* compared to the fraction of the same codon in *E. coli*. For example, the first arginine shown in red means only 7% of codon AGG is used in *E. coli* for arginine.

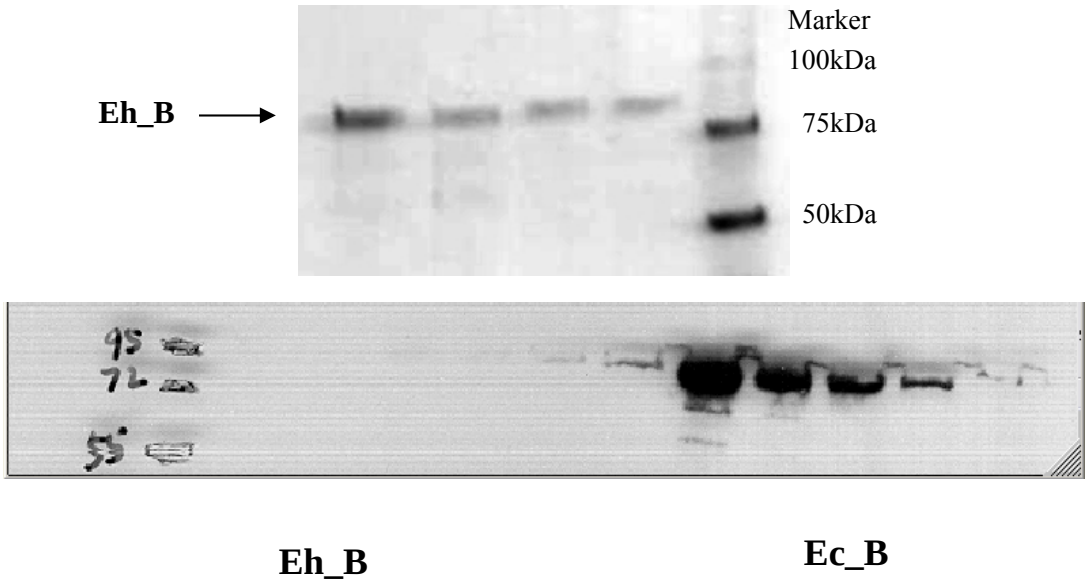


Fig. 4 Top panel: SDS-PAGE of purified Eh_B (7 μM, 3.5 μM, 1.75 μM, and 0.5 μM).

Bottom panel: Western blot was performed using polyclonal anti-Ec_B antibody. Same amount of Eh_B and Ec_B (7 μM, 3.5 μM, 1.75 μM, and 0.5 μM) were applied to SDS-PAGE and Eh_B could not be identified by Ec_B antibody.

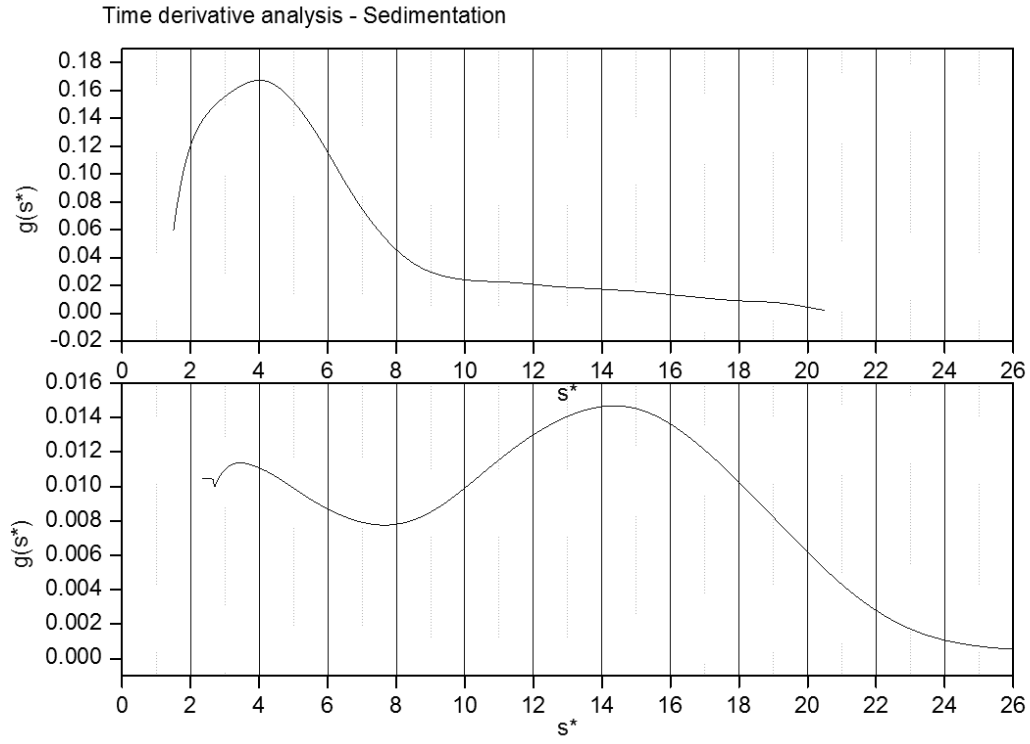


Fig. 5 Sedimentation velocity analysis of Eh_B. Ultracentrifugation was performed at 48,000 rpm and 20 °C for the 0.3-mg/ml protein samples in 50 mM Hepes/KOH pH 7.5, 0.2 M KCl, 20 mM $MgCl_2$, 1 mM EDTA, 2 mM β-mercaptoethanol. The apparent sedimentation coefficient distributions g(s*) are shown. Top: without nucleotide; bottom: with ATPγS.

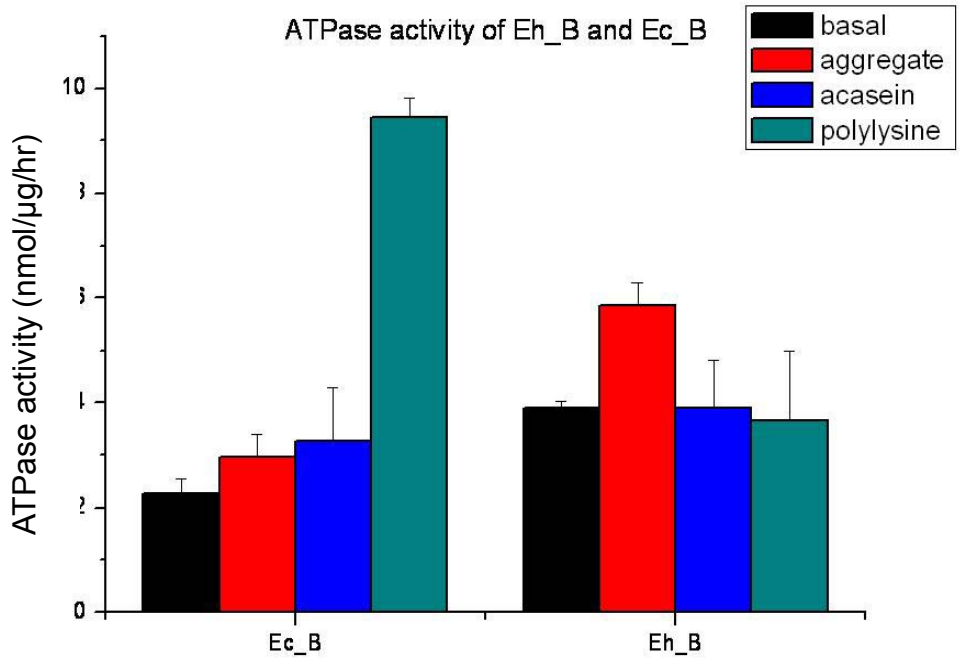
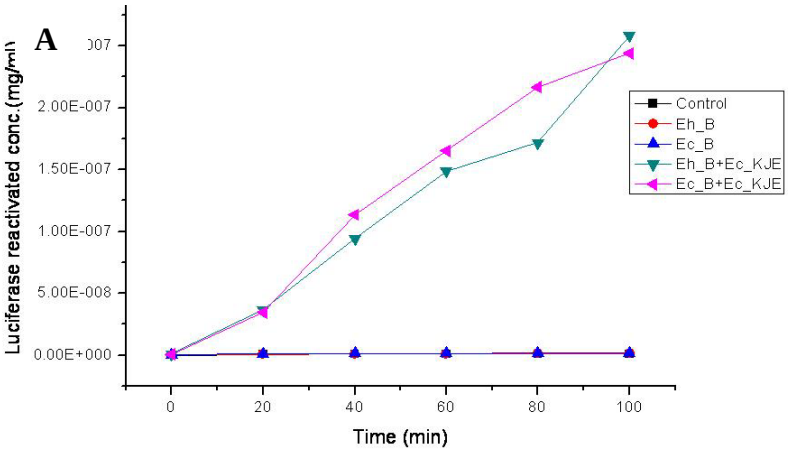
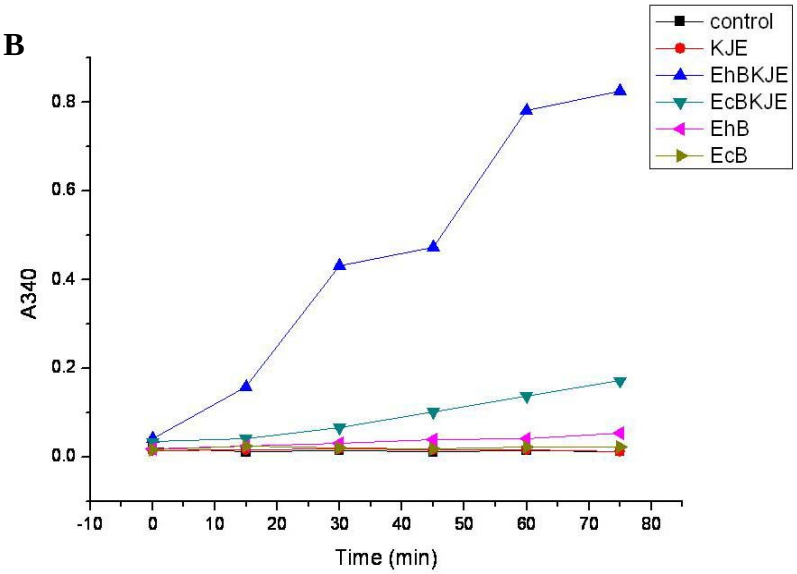


Fig. 6 ATPase activity was measured for Eh_B and Ec_B without other polypeptides (basal, black), with ~2 µM aggregated G6PDH (red), with 0.1 mg/ml α-casein (blue), or with 0.04 mg/ml poly-lysine (green).


A
Luciferase reactivated conc.(mg/ml)
2.00E-007
1.50E-007
1.00E-007
5.00E-008
0.00E+000
0
20
40
60
80
100
Time (min)
Control
Eh_B
Ec_B
Eh_B+Ec_KJE
Ec_B+Ec_KJE


B
A340
0.8
0.6
0.4
0.2
0.0
-10
0
10
20
30
40
50
60
70
80
Time (min)
control
KJE
EhBKJE
EcBKJE
EhB
EcB

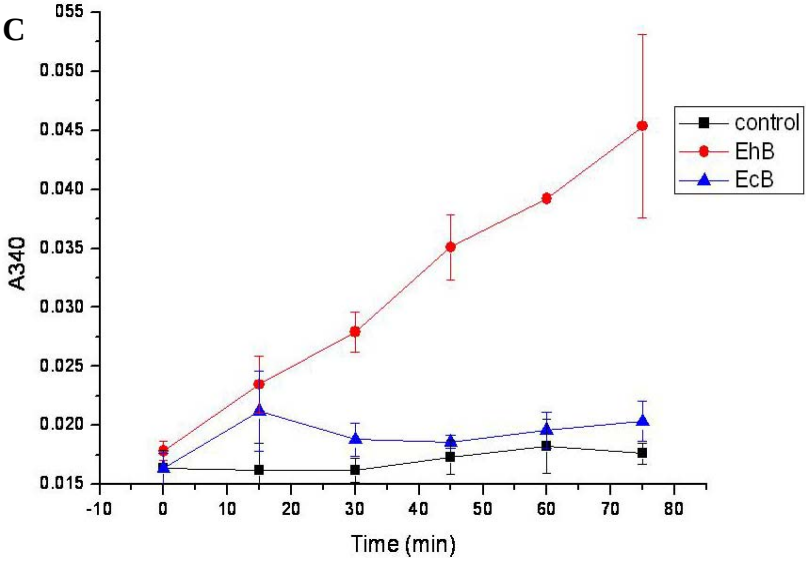

C
A340
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
-10
0
10
20
30
40
50
60
70
80
Time (min)
control
EhB
EcB

Fig. 7 Eh_B efficiently reactivates aggregated glucose-6-phosphate dehydrogenase (panel A) and luciferase (panel B) in cooperation with *E. coli* DnaK/DnaJ/GrpE.

G6PDH was unfolded with 10 M urea at 47 ℃ and then allowed to refold by dilution. Aggregation occurred during the refolding process and ClpB was added together with the co-chaperones DnaK, DnaJ and GrpE to disaggregate G6PDH. The recovery of G6PDH enzyme activity was measured at the indicated times. Luciferase was heat denatured at 45 ℃. Then ClpB and co-chaperones were added. The recovery of luciferase was monitored by luminometer.

Panel C shows the expanded scale of the reactivation of G6PDH by Eh_B without co-chaperones in panel B.

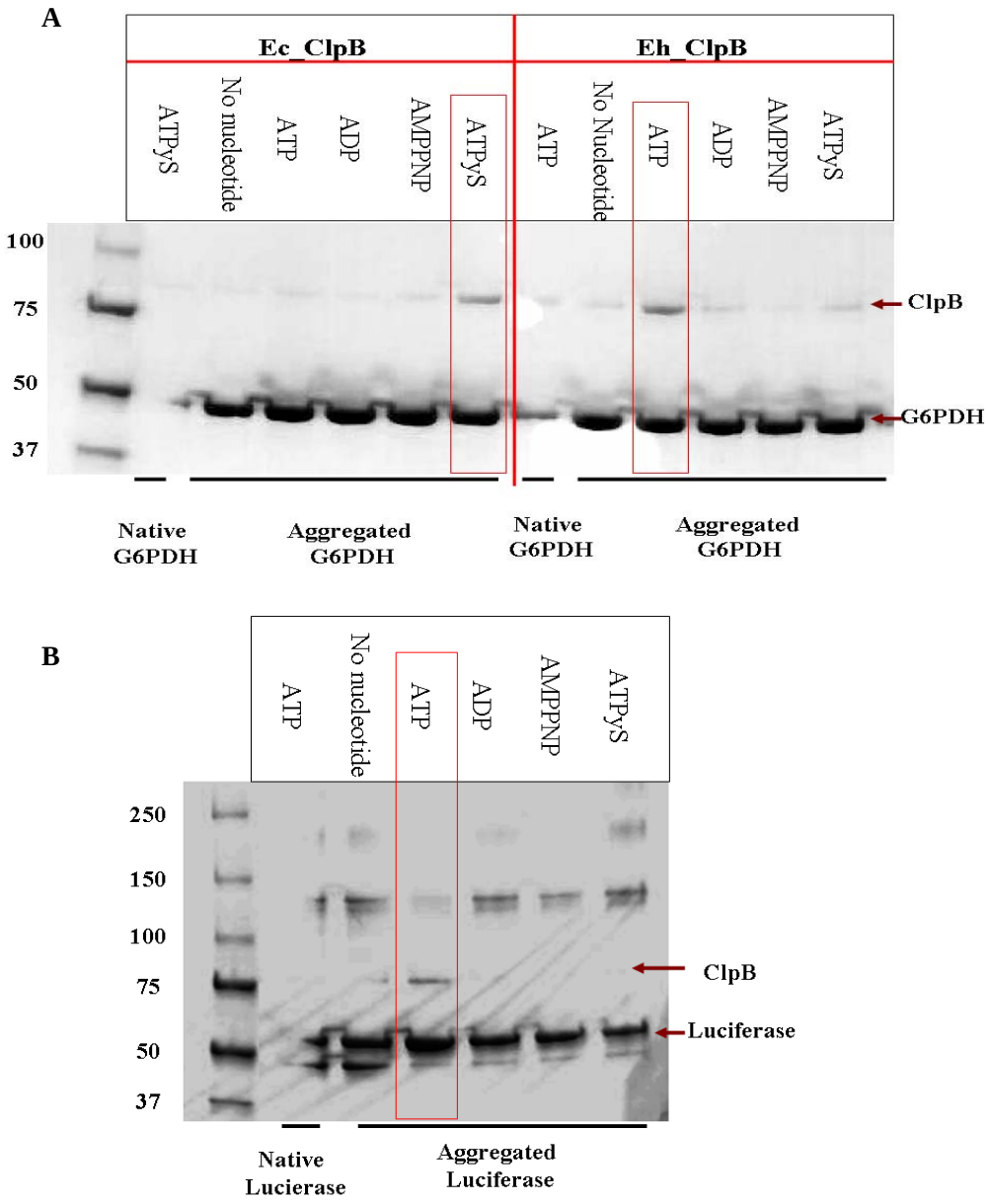


Fig. 8 1.5 μM Eh_ClpB/Ec_ClpB were incubated with 3 μM native/aggregated G6PDH (panel A left) or 2 μM native/aggregated luciferase (panel B) in the presence of different nucleotides. The mixtures were passed through a 0.1 μm filter and eluted with SDS. The elution fractions were analyzed by SDS-PAGE. Ec_ClpB bound to aggregates in the presence of ATPγS while Eh_ClpB bound to aggregates in the presence of ATP.

Chapter 4

Design and production of novel tools for studying the mechanism of substrate translocation by ClpB

4.1 Introduction

The detailed mechanism of ATP driven disaggregation of various substrates by ClpB is not clear. The most accepted model is substrate translocation through the central channel of the ClpB hexamer (Mogk, Dougan et al. 2004). In this model, the narrow channel of ClpB only allows unstructured peptides to go through, thus by using the energy of ATP hydrolysis, the aggregated substrate is unfolded while being threaded through the channel. Many AAA+ family members function by translocating the substrate through the central channel. For example, ATP-dependent protease ClpAP, ClpXP and HSiUV unfold the substrate by threading through the regulatory AAA+ channel and allow the unfolded peptide to enter the chamber of the attached protease for degradation (Sauer, Bolon et al. 2004). In the case of ClpB, the substrate is released from the channel and is allowed to refold into its native structure spontaneously or with the help of other chaperones.

The substrate pool of ClpB is large; how ClpB distinguishes between naturally folded proteins and aggregates remains unclear. Unlike its relatives ClpAP and ClpXP peptidase, ClpB does not use a tag peptide to help recognize its substrate. Peptide screening showed ClpB has a binding preference to aromatic and positively charged residues (Schlieker, Weibezahn et al. 2004). It is possible that the surface of aggregates is enriched in these types of amino acid residues. The substrate recognition of ClpB and the following translocation still need to be investigated.

To better understand substrate recognition and translocation we constructed the ClpB variant BAP. In BAP, we replaced the 27 amino acids in the C-terminal region of ClpB with the corresponding sequence of ClpA (Fig. 1). This makes BAP gain the capacity of binding ClpP, and converts BAP from a disaggregase to a peptidase. In

this Chapter, BAP was used to design a FRET experiment to study the recognition and following translocation of aggregated G6PDH.

4.2 Material and methods

Construction of BAP. A single C to A mutation was made immediately before the correspondent binding sequence of ClpB to ClpA to create a restriction enzyme cutting site for AvrII. Mutation was performed using site directed mutagenesis kit from Qiagen (Valencia, CA).

The *clpb* plasmid was digested by AvrII and EcoRI at 37℃ for 1hr. The product was analyzed on 0.5% agarose gel and the digested *clpb* band was extracted using gel extraction kit from Qiagen (Valencia, CA). Inserted sequences were synthesized by IDTDNA (Coralville, Iowa). Liagtion was performed using ligation kit from Invitrogen (Grand Island, NY).

Inserted sequence:

5'-GTACGGGAAACTGAGCGCAAATCCATTGGTCTTATCCACCAGGATAA CAGCACCGATGCGATGGAAGAGATCAAGAAGATC-3'

Proteins and Aggregates. Chaperones (ClpB, DnaK, DnaJ, GrpE) were produced or obtained as previously described (Barnett, Nagy et al. 2005). α-casein was obtained from Sigma (St. Louis, MO). Firefly luciferase was obtained from Promega (Madison, WI).

E. coli BL21(DE3) harboring plasmid pET25-G6PDH (obtained from Michael S. Cosgrove from Syracuse University) was used to purify G6PDH. Cells were grown at 37℃ with shaking to OD600~0.5 and then 0.4 mM IPTG was added. After 4 hours induction, cells were collected and sonicated. 0.004 g/g cell polyethyleneimine was added to the soluble fraction to precipitate DNA. After centrifugation, 50% saturation of $(NH_4)_2SO4$ was added to the supernatant and then the mixture was applied to centrifugation. The supernatant was dialyzed to remove extra salt and then was applied to ion-exchange column (Q-sepharose fast flow). Fractions containing

G6PDH were combined and concentrated using Amicon ultra centrifuge filter. Concentrated G6PDH was dialyzed to remove salt and then stored at -20℃.

E. coli BL21(DE3) harboring plasmid pET9a-ClpPS111A (obtained from Guillaume Thibault from University of Toronto) was used to purify ClpPS111A protein. Cells were grown at 37℃ with shaking to OD600~0.5 and then 0.4 mM IPTG was added. After 2 hours induction, cells were collected and sonicated. 0.004g/g cell polyethyleneimine was added to the soluble fraction to precipitate DNA. After centrifuge, the supernatant was applied to Ni-NTA column (Qiagen). Fractions containing ClpPS111A were combined and concentrated by precipitation using 60% saturation of (NH4)2SO4. Concentrated ClpPS111A was dialyzed to remove salt and then stored at -20℃.

Protein concentrations were determined spectrophotometrically. All mutations were introduced with QuikChange site-directed mutagenesis kit (Stratagene/Agilent Technologies, Santa Clara, CA).

To produce aggregates of G6PDH, the protein stock (324 μM) was diluted 2-fold with the unfolding buffer (10 M urea, 16% glycerol and 40 mM DTT) and was incubated at 47 °C for 5min. The mixture was then diluted 10-fold with the refolding buffer 1 (50 mM Tris/HCl pH 7.5, 20 mM $Mg(OAc)_2$, 30 mM KCl, 1 mM EDTA, and 1mM β-mercaptoethanol) and was incubated at 47 °C for 15 min and then on ice for 2 min. To produce aggregates of luciferase, 216 μM luciferase stock was diluted 300-fold with PBS containing 1 mg/ml BSA and then was incubated at 45 °C for 12 min. To produce aggregated GFP, 4.5 μM protein was heated for 10 min at 80 ˚C.

Gel-filtration chromatography. Gel-filtration experiments were performed at room temperature using a Superose 6 PC 3.2/30 column (Amersham Pharmacia Biotech) and a Shimadzu HPLC system. Aliquots (20 μl) of ClpB and/or ClpPS111A were chromatographed at 0.05 ml/min on a column equilibrated with 25 mM Hepes/KOH pH 7.4, 0.2 M KCl, 20 mM MgCl2, 0.1 mM EDTA, 1 mM DTT with 2 mM ATPγS. Collected 1-min fractions were analyzed by SDS–PAGE.

ClpB ATPase activity. The ClpB variants were incubated in assay buffer (100 mM Tris/HCl pH 8.0, 1 mM DTT, 1 mM EDTA, 10 mM $MgCl_2$, and 5 mM ATP) at 37 °C for 15 min without or with 0.1 mg/ml α-casein or 0.04 mg/ml poly-lysine. The concentration of ClpB was 0.05 mg/ml for the basal activity and in the presence of α-casein or 0.005 mg/ml in the presence of poly-lysine. The phosphate concentration generated by ClpB was measured as described before (Zolkiewski 1999).

Aggregate Reactivation Assays. Aggregated G6PDH or mutated G6PDH (16.2 μM) was diluted 10-fold with refolding buffer 1 containing 1.5 μM ClpB, 1 μM DnaK, 1 μM DnaJ, 0.5 μM GrpE and 6 mM ATP. The mixture was incubated at 30 °C and aliquots of the mixture were withdrawn to test the recovery of the G6PDH enzymatic activity. Aggregates diluted with refolding buffer without the chaperones were used as controls. To measure the G6PDH activity, aliquots from the refolding reaction were incubated in 50 mM Tris/HCl pH 7.8, 5 mM $MgCl_2$, 1.5 mM G6P and 1 mM $NADP^+$ for 10 min followed by the measurement of absorption at 340 nm.

ClpB-Aggregate Interaction Assay. Aggregated wild type or mutated G6PDH (16.2 μM) were diluted 10-fold with the refolding buffer 1 containing 1.5 μM ClpB and 6 mM ATPγS. The mixture was incubated at 30 °C with 600 rpm shaking for 20 min and then was applied to the filter device (Millipore Ultrafree-MC Centrifugal Filter Unit with pore size 0.1 μm). After 5 min incubation at room temperature, the filter device was centrifuged at 13,000 rpm for 4 min to get the flow-through fractions. The filter device was washed with the refolding buffer 1 containing ATPγS at 30 °C for 5 min and then re-centrifuged. Next, 1x SDS loading buffer was added to the filter device and the filter device was incubated at 50 °C for 5 min with shaking. Then, it was centrifuged to get the eluate fractions, which were applied to SDS-PAGE.

4.3 Results

4.3.1 FRET experiment design

To investigate the recognition and translocation of a substrate we designed a FRET experiment. In this experiment, BAP is incorporated with an inactive form of ClpP (S111A mutant) which can bind but cannot degrade substrates (substrate trap). A fluorescence donor label was introduced into the proteolytic chamber of ClpP, and a fluorescence acceptor label was introduced into different positions of the substrate. When translocation occurs, the donor and acceptor would be close enough to give FRET signal, allowing us to study which part of the aggregates enters the chamber first and how the substrate is translocated. (Fig. 2)

4.3.2 Construction and characterization of BAP/ClpP complex

4.3.2.1 Construction of BAP

We modified the construction of BAP described by another group (Weibezahn, Tessarz et al. 2004). By replacing the C-terminal sequence of ClpB with sequence responsible for binding ClpP from ClpA. First, two restriction enzyme sites were introduced into the proper position of *clpb* gene by mutagenesis. Then the ClpB C-terminal sequence was cut out by restriction enzyme digestion and the synthesized ClpA sequence was introduced into the plasmid. The binding site of BAP to ClpP is on the bottom of the hexamer cylinder, near the exit of the central channel (red in Fig. 1).

4.3.2.2 Biochemical properties of BAP

After purification of BAP protein its ATPase activity was tested. As shown in Fig. 3, BAP had similar basal ATPase activity as wt ClpB, suggesting its proper oligomerization since ATP hydrolysis occurs on the interface of two subunits in ClpB hexamer (Barnett, Zolkiewska et al. 2000). Also, its response to pseudo substrate alpha-casein was not affected. Thus, we conclude that the sequence substitution did not have much impact on the structure of BAP.

Proper substrate translocation requires the cooperation of all the domains. To

make sure that the changes in the C-terminus did not affect the motions and interactions between the domains, we also tested the chaperone activity of BAP. Fig. 4 shows the reactivation of G6PDH by BAP/KJE and ClpB/KJE. BAP had the same level of G6PDH reactivation as wt ClpB and the reactivation rate was also similar.

4.3.2.3 Construction of ClpP mutants

E. coli ClpP forms a tetradecamer (Fig. 5) and has ClpA binding sites at both side of the cylinder (Yu and Houry 2007). We obtained an inactive form of ClpP (S111A) plasmid from Dr Walid A. Houry (University of Toronto, Canada). The *E. coli* ClpP has no cys in the sequence thus we introduced 3 cys (F31C, L62C, and L139C) into inactive ClpPS111A at different positions of the preteolytic chamber. For example, mutant L139C in the middle of the channel is shown in green in bottom panel of Fig. 5.

4.3.2.4 Interaction between BAP and ClpP

We then tested the interaction between BAP and ClpP by gel filtration. As shown in Fig. 6, ClpB, BAP, ClpP alone and the mixture of BAP/ClpP, ClpB/ClpP were injected into the column and their elution times were recorded. The tetradecamer of ClpP has a molecular weight around 300 kDa, which was in correspondence with an elution peak at 33 min. BAP and ClpB had an earlier peak at around 29 min with a 570 kDa hexamer. The mixture of ClpB/ClpP also showed a 29 min peak with a shoulder, suggesting the mixture of molecules with two different sizes. In the mixture of BAP/ClpP, there was a shift of the peak to ~26 min, suggesting a large complex formed. The complex formed by BAP and ClpP is about 870 kDa, which is consistent with the component molecular weight.

In the normal aggregate reactivation by ClpB, substrate is translocated and released, which allows a spontaneous or chaperone assisted refolding. If ClpP binds to the exit of the ClpB central channel and captures the unfolded substrate with an inactive form, the translocation by ClpB will be stalled because the exit is blocked. To

test this hypothesis, we tested the reactivation of G6PDH by BAP in the presence of ClpP. As shown in Fig. 7, the reactivation was strongly inhibited by the presence of ClpP, which further confirmed BAP/ClpP interaction. We also tested the inhibition with different ratio of BAP/ClpP. In theory, the ratio of 2BAP:1ClpP should have the most inhibition since one ClpP molecule is able to bind two BAP molecules. However, the inhibition of different ratio was undistinguishable. It seemed that the more ClpP there was, the less substrate was reactivated. This makes sense since the interaction between ClpP and chaperones bound to it is dynamic (Martin, Baker et al. 2007); more ClpP in the solution will give BAP more chance to bind to it, thus inhibition of the reactivation was much stronger.

Overall, the constructed BAP interacted with ClpP and was ready for further investigation.

4.3.3 Substrate selection and labeling

The substrate in the FRET experiment should have the following properties: a known crystal structure so that the position of the fluorescence labeling will be accessible; easy to label; and has an activity assay to test recovery by ClpB reactivation. Considering all these requirements, we chose glucose-6-phosphate dehydrogenase (G6PDH) from *Leuconostoc mesenteroides* as our substrate. The crystal structure of this G6PDH was solved in 1994, and the enzyme is a dimer, each subunit consisting of two domains (Rowland, Basak et al. 1994). There is no cys in the sequence of this G6PDH, which makes it easier to introduce cys into the position we wanted to label. And the reactivation of urea unfolded G6PDH aggregates is commonly used in chaperone assays.

Where should we put the fluorescence acceptor label? In the designed FRET experiment, the label has to be on the surface of the aggregate to be accessible by the donor label in ClpP. However the nature of aggregated protein is poorly understood. There are studies identifying amino acid residues that promote protein aggregation such as residues responsible for maintaining the net charge of the whole protein

(Dobson 2001; Chiti, Calamai et al. 2002). To further predict the aggregation propensity of an amino acid residue in a protein, a statistical mechanic algorithm called Tango was developed (Fernandez-Escamilla, Rousseau et al. 2004). Tango calculates the propensity of an amino acid residue to be in cross-beta aggregation region based on the assumptions that in the aggregated region, beta-strand is the main structure and regions involved in aggregation are fully buried.

Fig. 8 shows the tango beta aggregation index of G6PDH. We can see that there are four highly aggregation-prone regions. The two termini of G6PDH seemed highly prone to form aggregates. The aggregation prone regions were labeled in the structure of G6PDH dimer. Most of these regions are partially on the surface of the protein, which makes them easy to access during the unfolding of the protein and formation of aggregates.

According to the principle of Tango, the aggregation-prone region should be fully buried in aggregated protein. We assumed that the least aggregation-prone region will be most probably to stay on the surface of aggregates. Thus, we designed several cys mutants of G6PDH spreading the whole sequence, in both aggregation-prone regions and regions that are highly unlikely to aggregate. We also made sure that these mutated Cys were on the surface of G6PDH for fluorescence dye access and not on the interface of the dimer so that enzyme activity would not be affected. Pannel C of Fig. 8 shows the distribution of all G6PDH mutants.

4.3.4 Characterization of labeled substrates

Two of the G6PDH mutants were purified in *E. coli*: one with a Cys in the aggregation prone region (S3C) and the other in the region that was predicted to have low propensity of aggregation (A392C).

We compared the enzyme activity of recombinant wt G6PDH, G6PDH purchased from Sigma and the two mutants. Both mutants lowered the enzymatic activity and S3C to a larger extent (Fig. 10). The difference was not significant and as long as the mutant was active for recovery examination (Fig. 4 lower panel).

Both mutants were labeled with Fluorescein-5-Maleimide. The spectra are shown in Fig. 9. As we can see, both labeling reactions were successful but S3C had a better labeling efficiency (~80%) compared to A392C (only ~25%). Both F5M S3C and F5M A392C were as active as the unlabeled form.

We then tested how aggregation would influence the fluorescence signal. Hypothetically, F5M S3C should be fully buried in aggregates and thus should lose the fluorescence signal after aggregation, while F5M A392C should stay on the surface and the fluorescence signal should not be much affected. As we can see in Fig. 11, the fluorescence intensity of both mutants significantly dropped after aggregation. However, F5M S3C had a larger decrease (~80%) compared to F5M A392C (~50%). This suggested that F5M S3C may be buried more in the aggregates. For F5M A392C, the acute structural changes during aggregation could affect the fluorescein, thus the decrease in the fluorescence intensity was totally reasonable.

We also tested how the binding of ClpB altered the fluorescence signal. Both WT clpB/ATPγS and Trap ClpB/ATP were used. The same filter assay described in chapter 2 was used. As shown in Fig. 12, both aggregates bound to ClpB and Trap ClpB. However, the binding of ClpB did not have any influence on fluorescence intensity of the mutants, indicating there were not many conformational changes in the aggregates upon ClpB binding or the changes were subtle and could not be revealed by fluorescence signal. One interesting phenomenon we discovered was that labeling of S3C seemed to abolish aggregation (Fig. 12). This was not seen in the F5M A392C mutants either because it had to be specific labeling to avoid aggregation or simply because the A392C mutant was not labeled as much as S3C. We also saw a band of ~130kDa in the aggregated F5M A392C. According to the size, this should be two dimers of A392C. Mass spectrum analysis discovered both A392C and ClpB in the band. However we suspect that the ClpB was only traces left in the gel and the band should be some form of stable complex of A392C. Further investigation is needed to better understand the properties of these fluorescence labeled aggregates.

We also tested the fluorescence signal change during reactivation of F5M A392C.

We did not test the F5M S3C mutant because there were much less aggregates after urea unfolding and refolding. We can see in Fig. 13 that the labeled mutant could be reactivated similar to wild type protein, with the fluorescence signal increasing slowly over time. The increase was rather small considering the fact that ClpB/KJE can only reactivate a very small portion of the whole aggregates pool.

All together, we labeled two G6PDH mutants and tested the properties of their aggregates. The aggregation step needs more investigation but A392C appeared to be ready for use in a FRET experiment.

4.4 Discussion

The mechanism of aggregate reactivation by ClpB/KJE machinery is still unclear, although there are different models proposed: A "crowbar" model in which aggregates are broke into smaller pieces by the M domain of ClpB or by deoligomerization of ClpB (Glover and Lindquist 1998; Lee, Hisayoshi et al. 2003); substrate translocation model in which ClpB recognizes aggregates and threads single polypeptides through its central channel. These models may not be mutually exclusive; however the construction of BAP (Weibezahn, Tessarz et al. 2004) showed the applicability of the substrate translocation model.

In this Chapter, we designed an experiment that will give us detailed insights into substrate recognition and translocation by ClpB. By fluorescence labeling the aggregated substrate and the chamber of BAP/ClpP complex, we could use FRET to analyze which part of the aggregate is recognized by ClpB first and how aggregates are translocated through the ClpB central channel.

The helix-loop-helix with the IGF/L binding motif in ClpA is responsible for interaction with ClpP (Kim, Levchenko et al. 2001; Ortega, Lee et al. 2004). To construct BAP, the corresponding helix-loop-helix in ClpB was replaced by the sequence from ClpA. The constructed BAP gained the ability to bind ClpP and had similar ATPase activity and disaggregation activity as wild type ClpB. In order to capture the translocated substrate in the channel so that there would be enough time to

analyze the interaction, an inactive form of ClpP which binds to substrates but does not degrade or release them (ClpPS111A) was used. In the BAP/ClpPS111A complex, the aggregates were first translocated through the central channel of BAP and then directly entered the proteolytic chamber of ClpP. Three cys mutants of inactive ClpP were constructed in order to put the donor fluorescein at different positions in the chamber.

The acceptor fluorescein on the aggregates has to be accessible to the donor fluorescein, that is to say, it has to be on the surface of the aggregates. The undefined nature of aggregates makes it difficult to decide where the fluorescence label should be on the native protein, because during aggregation the protein undergoes such significant conformational changes that how the position of the labeled amino acid residue will be hard to determine.

Despite the seemingly random nature of protein aggregation, there are clues of aggregation prone regions (APRs) in a given protein. These APRs have unique features in their charges, hydrophobicity and secondary structure (Agrawal, Kumar et al. 2011). Many computational algorithms have been developed to discover the possible APRs in protein sequences such as AGGRESCAN (http://bioinf.uab.es/aggrescan/), AMYLPRED (http://biophysics.biol.uoa.gr/AMYLPRED), TANGO (http://tango.crg.es/), and Zyggregator (http://www-vendruscolo.ch.cam.ac.uk/) etc. We used Tango to predict the possible APRs in G6PDH. We replaced the amino acid residues in the APRs and outside the APRs with cys for further fluorescence labeling. In Tango prediction the APRs are supposed to be buried inside in the aggregate. Fluorescent signal changes after aggregation confirmed this assumption. In Fig. 11, the fluorescent signal decrease in S3C (within APRs) was much greater than that in A392C (outside APRs), indicating in S3C more fluorescein were buried. After the labeled proteins were reactivated by ClpB/KJE, the fluorescent signal recovered slowly (Fig. 13); this could result from the reactivated G6PDH, or during reactivation the aggregate was unfolded and the buried fluorescein was exposed.

To better understand the features of labeled substrates, we tested the binding of ClpB to aggregated fluorescently labeled G6PDH. The binding did not have effects on the fluorescent signal, indicating no significant conformational change of aggregate upon ClpB binding. Interestingly, we found that labeling of G6PDH S3C seemed to inhibit aggregation. It is possible that the fluorescein on the APR disrupted the formation of aggregates since this was not seen in the labeled non-APR mutant G6PDH A392C. However, the different labeling efficiency could also explain this, since much less G6PDH A392C was labeled. It is less possible that a small label (molecular weight ~427) would have disrupted aggregation at any given position in the protein sequence. We favor the hypothesis that only labeling in the APRs will inhibit aggregate formation. Further investigation is needed to better understand the properties of fluorescenctly labeled aggregates.

In conclusion, we successfully constructed a BAP/ClpP complex which could trap substrates after their translocation; we also constructed labeled substrates according the prediction of Tango. The materials for the FRET experiment are ready, although further investigation of substrate labeling needs to be done.

References

Agrawal, N. J., S. Kumar, et al. (2011). "Aggregation in protein-based biotherapeutics: computational studies and tools to identify aggregation-prone regions." J Pharm Sci **100**(12): 5081-5095.

Barnett, M. E., A. Zolkiewska, et al. (2000). "Structure and activity of ClpB from Escherichia coli. Role of the amino-and -carboxyl-terminal domains." J Biol Chem **275**(48): 37565-37571.

Chiti, F., M. Calamai, et al. (2002). "Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases." Proc Natl Acad Sci U S A **99 Suppl 4**: 16419-16426.

Dobson, C. M. (2001). "The structural basis of protein folding and its links with human disease." Philos Trans R Soc Lond B Biol Sci **356**(1406): 133-145.

Fernandez-Escamilla, A. M., F. Rousseau, et al. (2004). "Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins." Nat Biotechnol **22**(10): 1302-1306.

Glover, J. R. and S. Lindquist (1998). "Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins." Cell **94**(1): 73-82.

Kim, Y. I., I. Levchenko, et al. (2001). "Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase." Nat Struct Biol **8**(3): 230-233.

Lee, S., M. Hisayoshi, et al. (2003). "Crystallization and preliminary X-ray crystallographic analysis of the Hsp100 chaperone ClpB from Thermus thermophilus." Acta Crystallogr D Biol Crystallogr **59**(Pt 12): 2334-2336.

Martin, A., T. A. Baker, et al. (2007). "Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease." Mol Cell **27**(1): 41-52.

Mogk, A., D. Dougan, et al. (2004). "Broad yet high substrate specificity: the challenge of AAA+ proteins." J Struct Biol **146**(1-2): 90-98.

Ortega, J., H. S. Lee, et al. (2004). "ClpA and ClpX ATPases bind simultaneously to opposite ends of ClpP peptidase to form active hybrid complexes." J Struct Biol **146**(1-2): 217-226.

Rowland, P., A. K. Basak, et al. (1994). "The three-dimensional structure of glucose 6-phosphate dehydrogenase from Leuconostoc mesenteroides refined at 2.0 A resolution." Structure **2**(11): 1073-1087.

Sauer, R. T., D. N. Bolon, et al. (2004). "Sculpting the proteome with AAA(+) proteases and disassembly machines." Cell **119**(1): 9-18.

Schlieker, C., J. Weibezahn, et al. (2004). "Substrate recognition by the AAA+ chaperone ClpB." Nat Struct Mol Biol **11**(7): 607-615.

Weibezahn, J., P. Tessarz, et al. (2004). "Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB." Cell **119**(5): 653-665.

Yu, A. Y. and W. A. Houry (2007). "ClpP: a distinctive family of cylindrical energy-dependent serine proteases." FEBS Lett **581**(19): 3749-3757.

```
ClpA.          IGAPPGYVGFDQGGLLTDAVIKHPHAVLLLDEIEKAHPDVFNILLQVMDNGTLTDNNGRK 593
ClpB.          VGAPPGYVGYEEGGYLTEAVRRRPYSVILLDEVEKAHPDVFNILLQVLDDGRLTDGQGRT 706
               :********:::** **:** ::*::*:****:**************:*:* ***.:**.

ClpA.          ADFRNVVLVMTTNAGVRETERKSIGLIHQDNSTDAMEEIKKITPEFRNRLDNIIWFDHL 653
ClpB.          VDFRNTVVIMTSNLGSDLIQERFGELDYAHMKELVLGVVSHNFRPEFINRIDEVVVFHPL 766
               .****.*::**:* *    :.:   * : . .  .:  :.: * *** **:*::: *. *
```

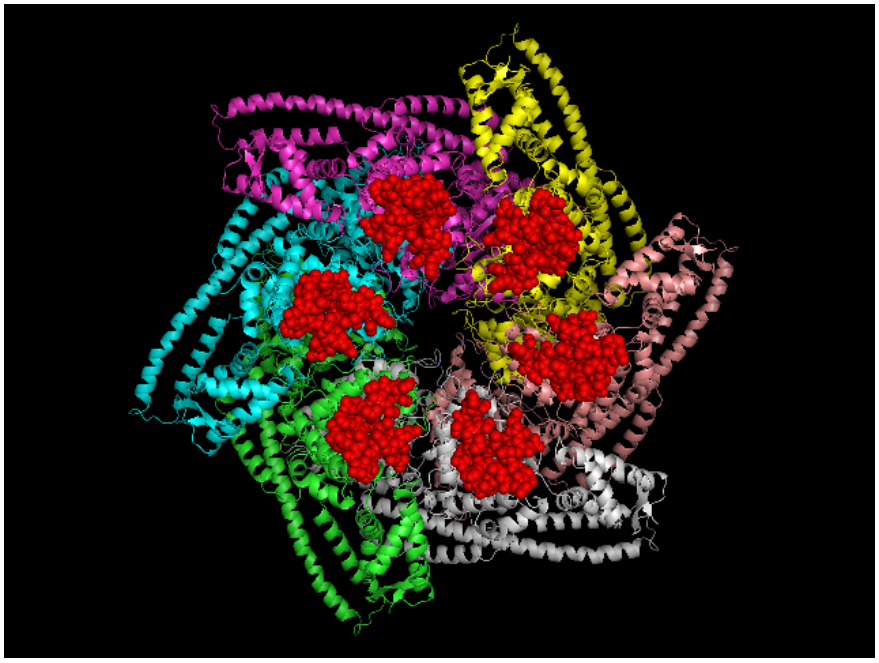


Fig. 1 Construction of BAP.

Top panel: sequence alignment of ClpA and ClpB in the ClpP binding region. The red square shows the 27 amino acids that had been exchanged.

Bottom panel: bottom view of ClpB hexamer. The exchanged amino acids are shown in red.

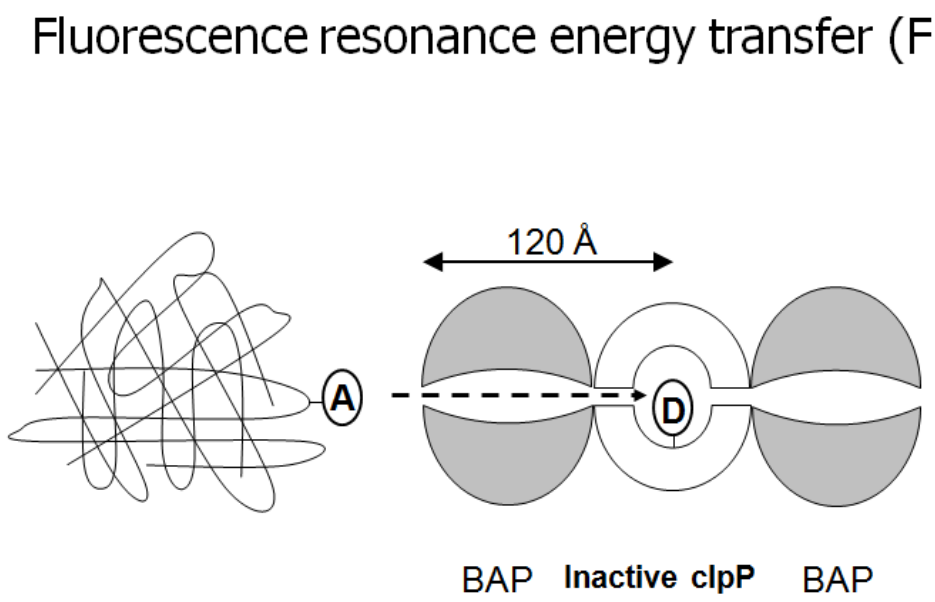


Fig. 2 FRET experiment design.

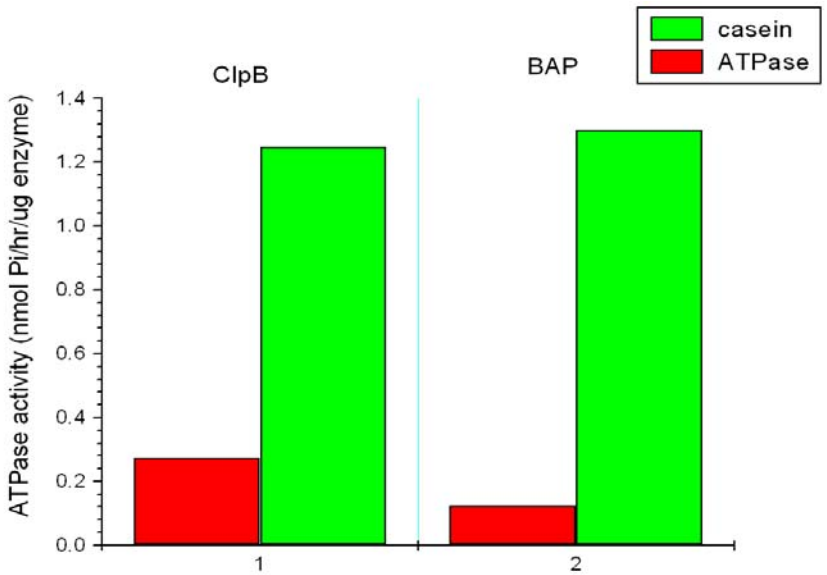


Fig. 3 ATPase activity of BAP. BAP had similar basal ATPase activity comparing to wild type ClpB. In the presence of 0.4 mg/ml pseudo substrate alpha-casein, ATPase activity of BAP was also activated to the similar level of wild type.

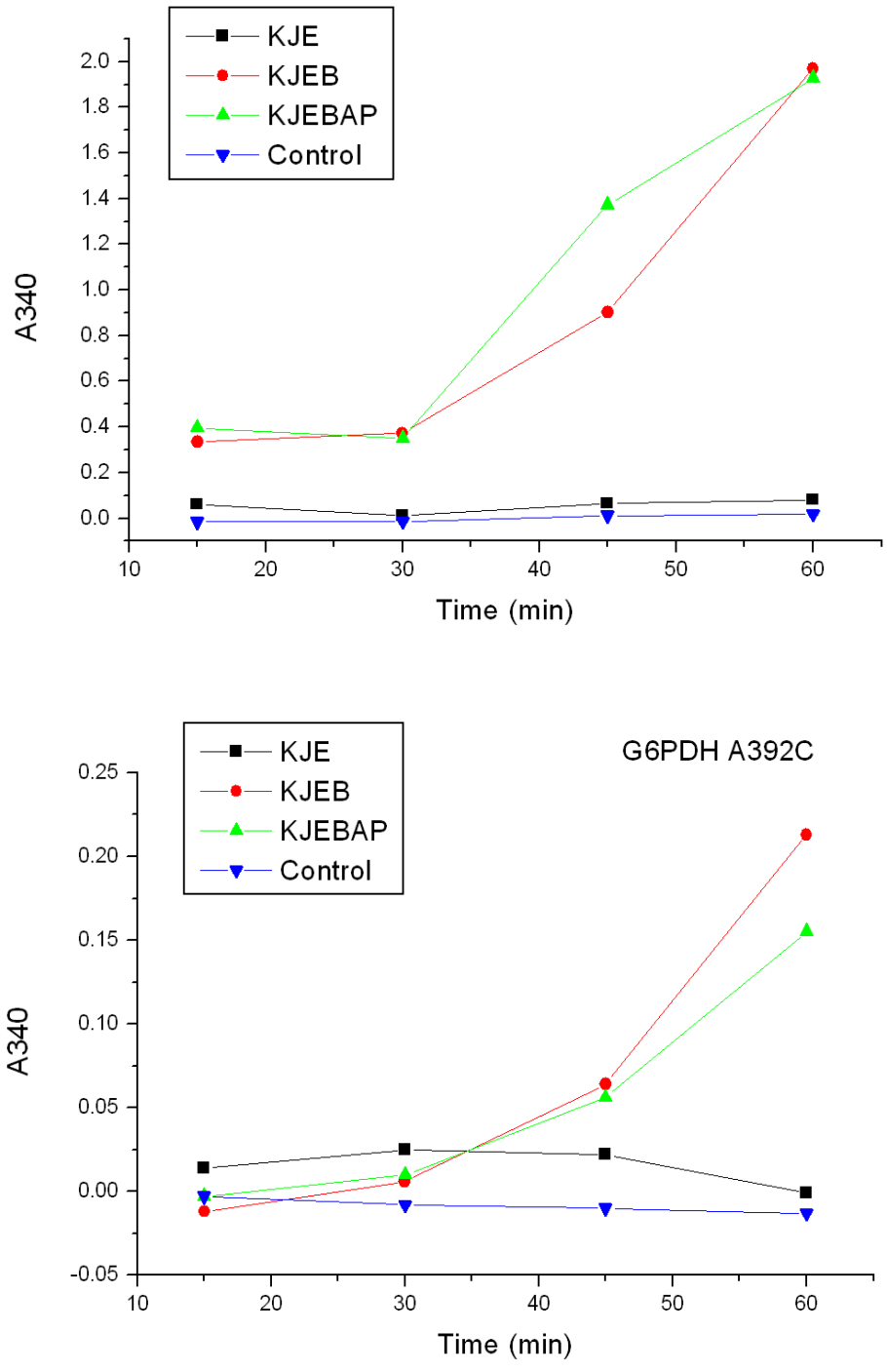


Fig. 4 Chaperone activity of BAP. BAP reactivated aggregated G6PDH and G6PDH mutant at similar rate and yield comparing to wild type.

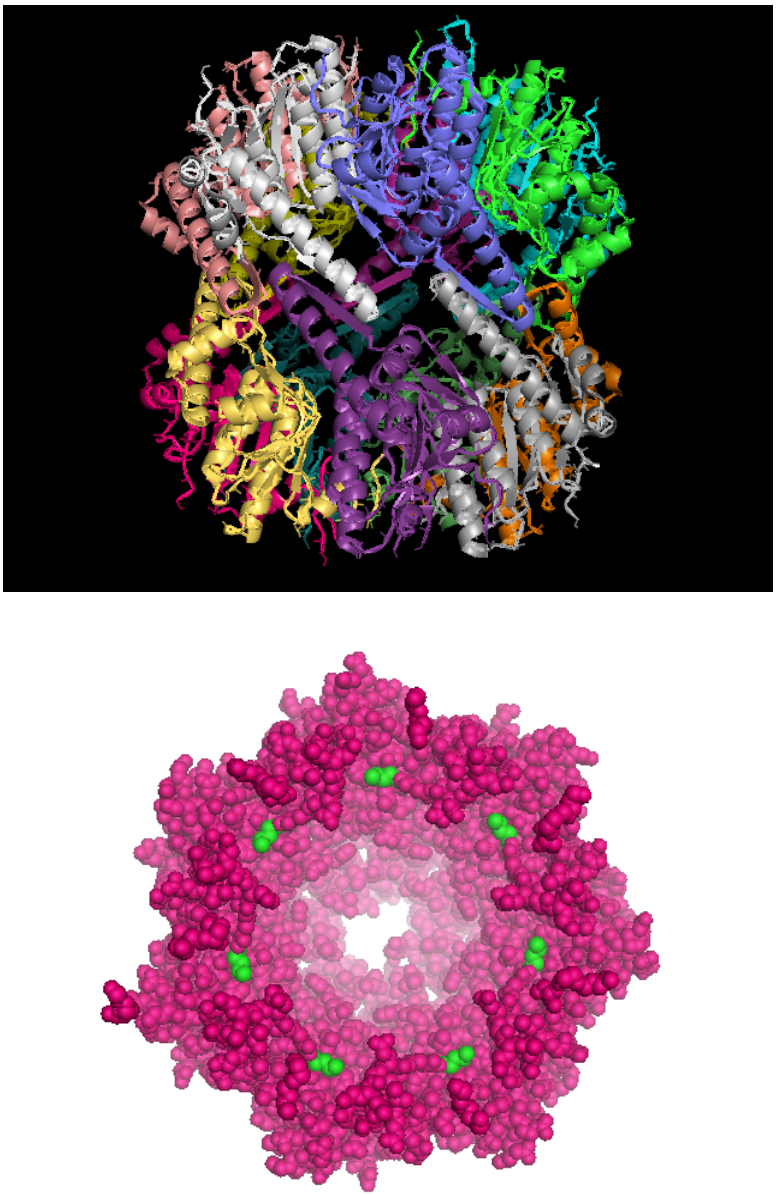


Fig. 5 Structure of ClpP tetradecamer.

The bottom panel shows the cut view of the preteolytic chamber of ClpP tetradecamer and L139C mutation is shown in green.

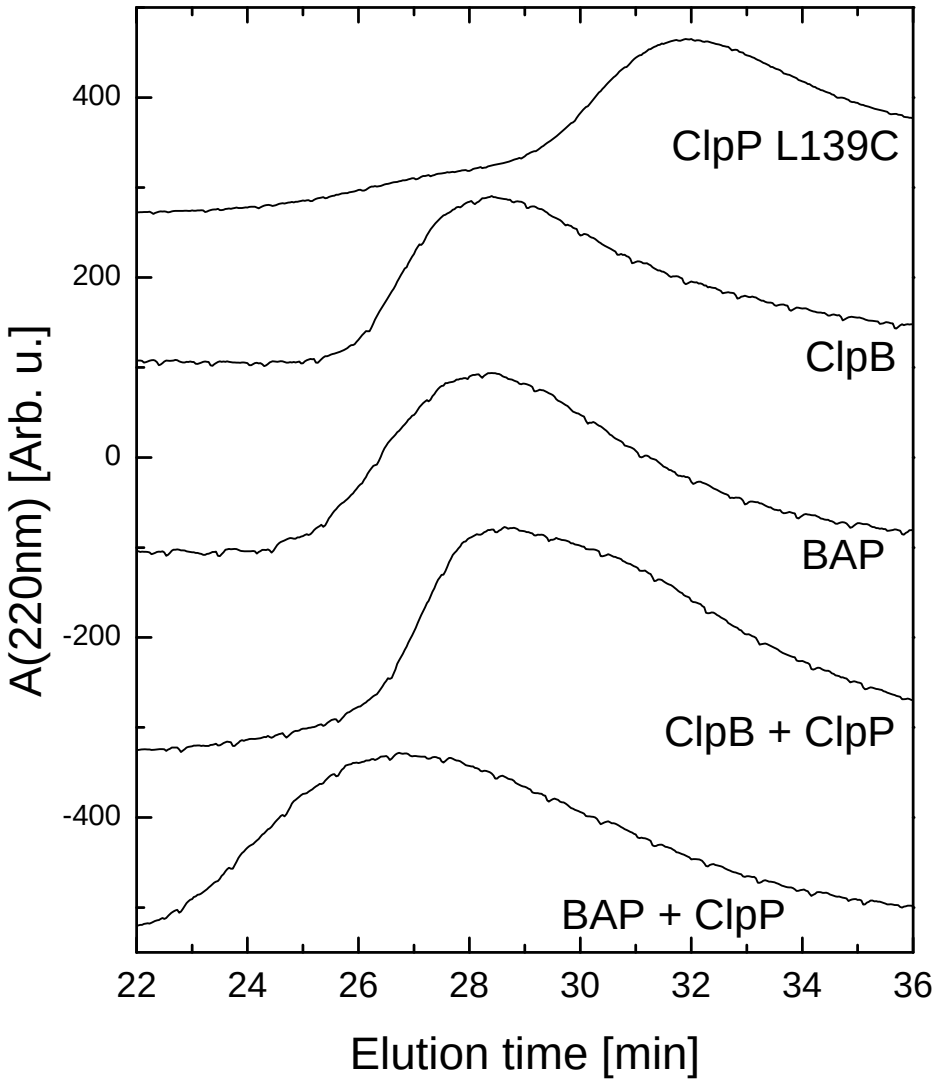


Fig.6 Gel filtration chromatography of ClpP/BAP mixture in the presence of ATPγS. Sizes of protein oligomers: ClpB and BAP hexamer ~570 kDa; ClpPS111A tetradecamer: ~300 kDa; BAP/ClpPS111A complex: ~870 kDa. The shift of BAP/ClpP peak confirmed the interaction between BAP and ClpP.

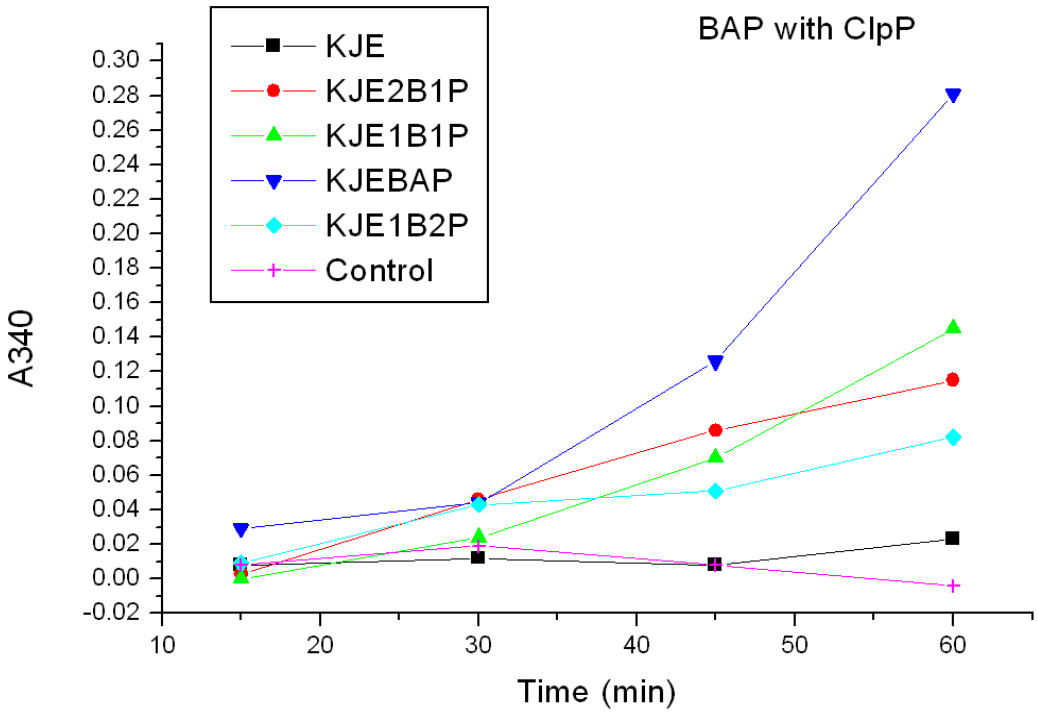


Fig.7 Reactivation of G6PDH aggregate in the presence of ClpPS111A. BAP and ClpPS111A were mixed at different ratios (2:1; 1:1; 1:2) and the reactivation of G6PDH aggregate was tested. The decrease of reactivation yield by ClpPS111A further confirmed the interaction between BAP and ClpP.

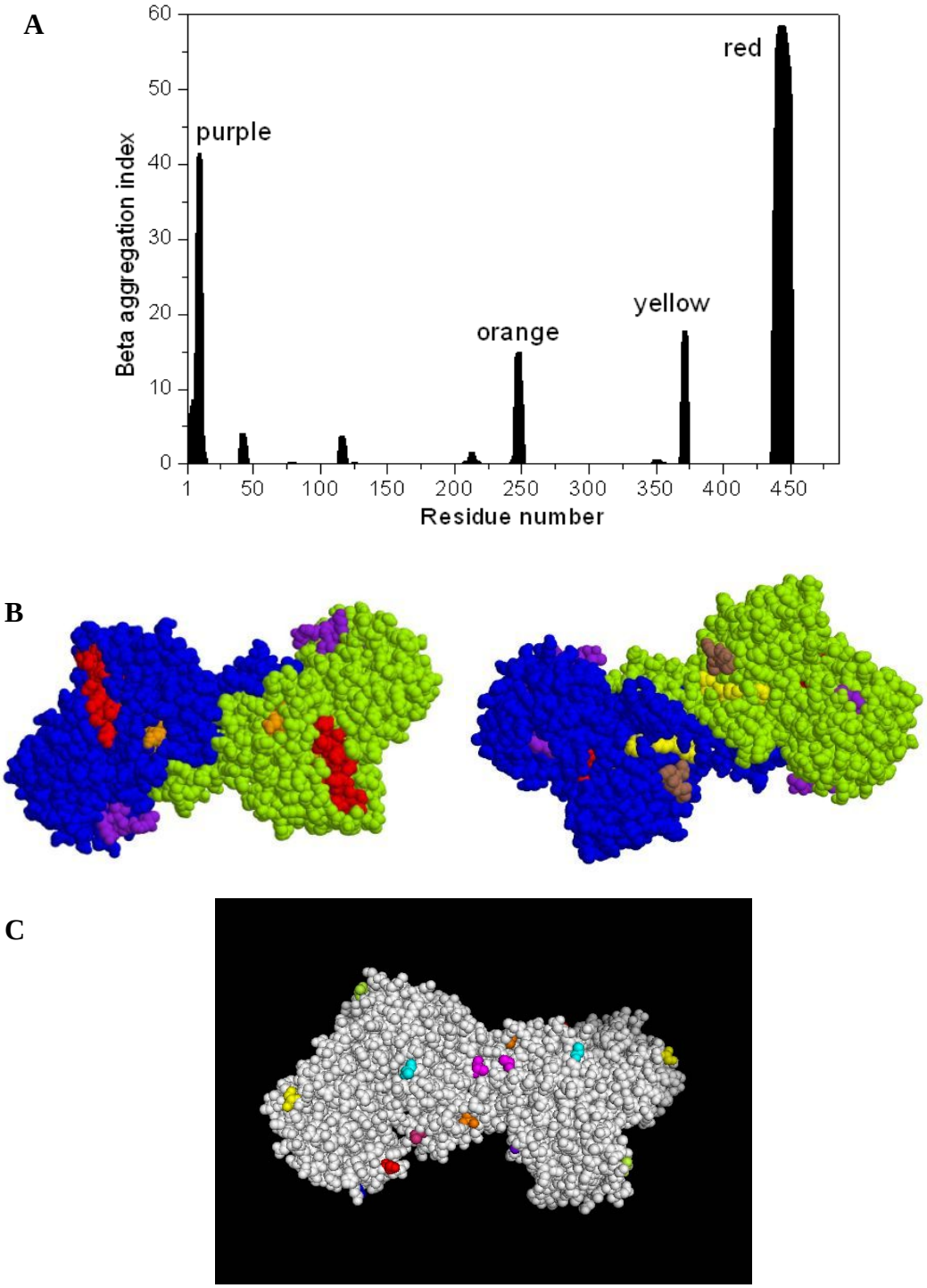


Fig. 8 Tango prediction of G6PDH. (A), Tango beta aggregation index of G6PDH. (B), aggregation prone area in G6PDH dimer. Color is consistent with the index graph. The two monomer of G6PDH are shown as blue and green. (C), mutations of G6PDH. S3C, A39C, A71C, A264C, A382C, A453C, and A481C. Some are buried inside.

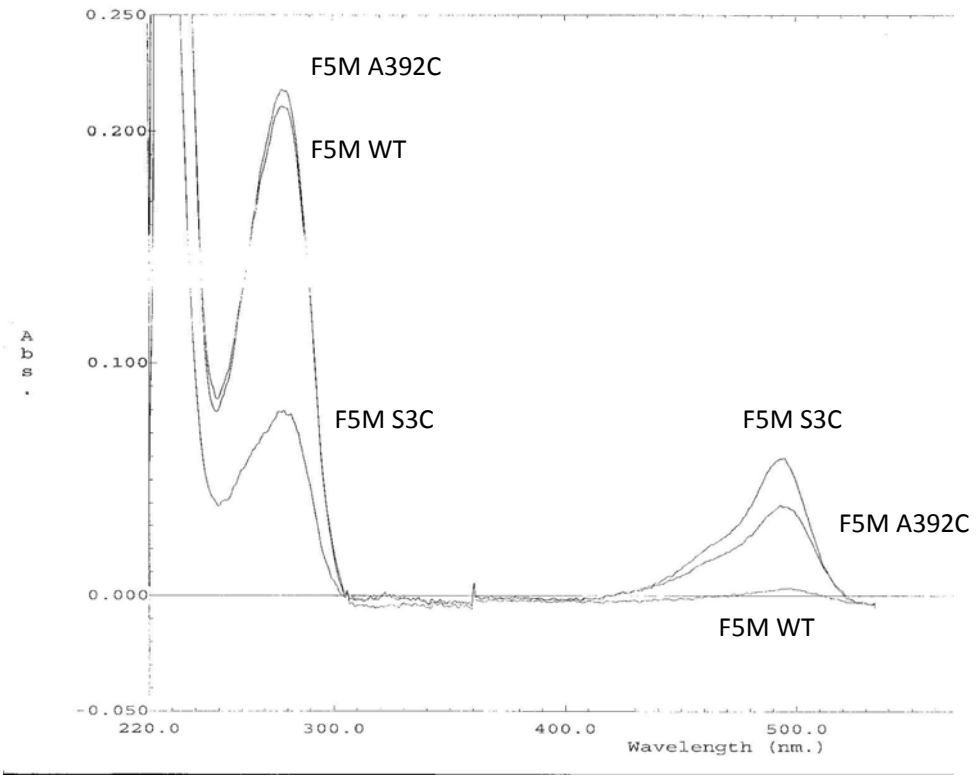


Fig. 9 Spectra of F5M labeled G6PDH WT, S3C and A392C. Absorbance of Fluorescein-5-maleimide is at 493nm.

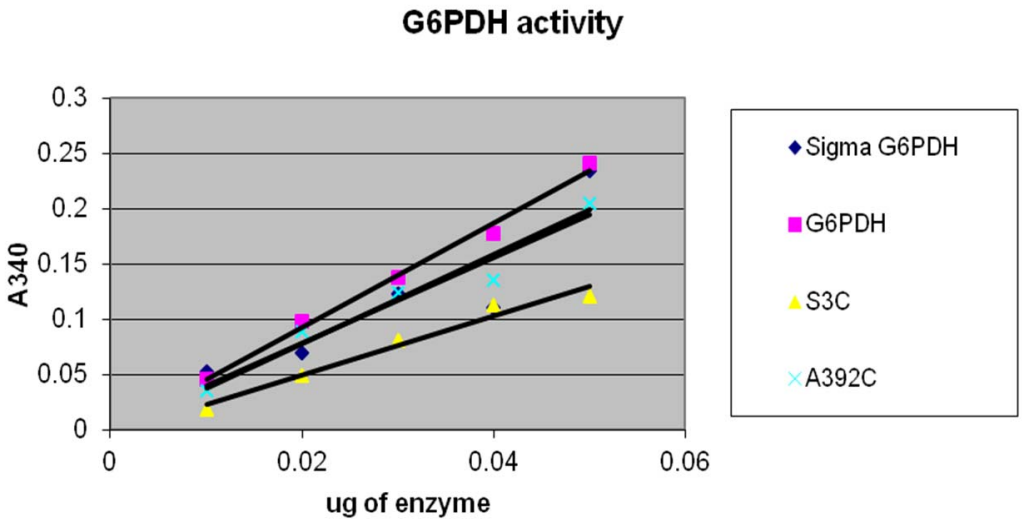
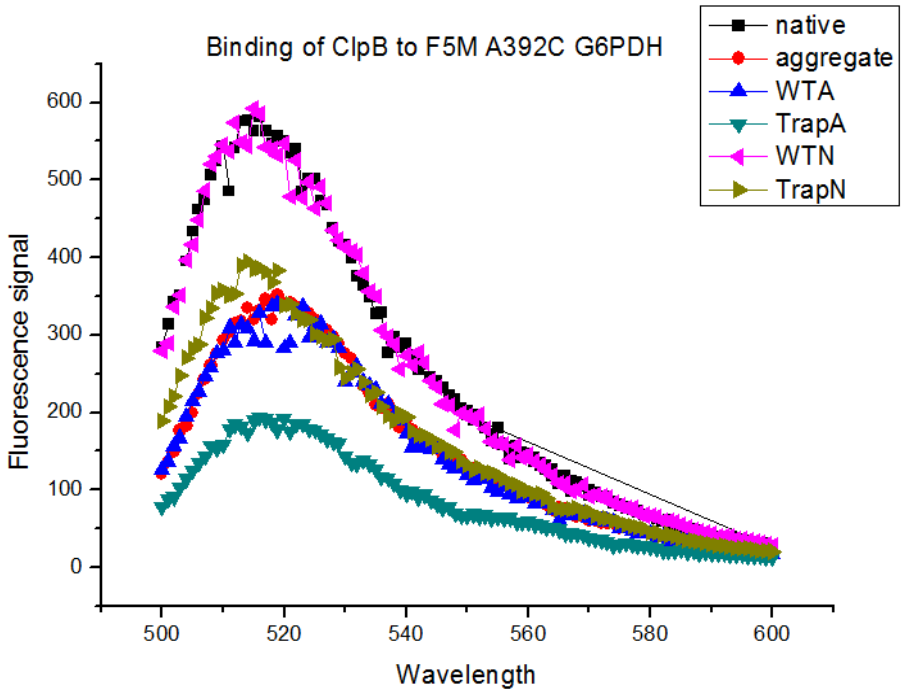


Fig. 10 Standard curve of G6PDH activity of the mutants.

Activity of purified wt G6PDH and S3C, A392C mutants were compared with G6PDH purchased from Sigma.



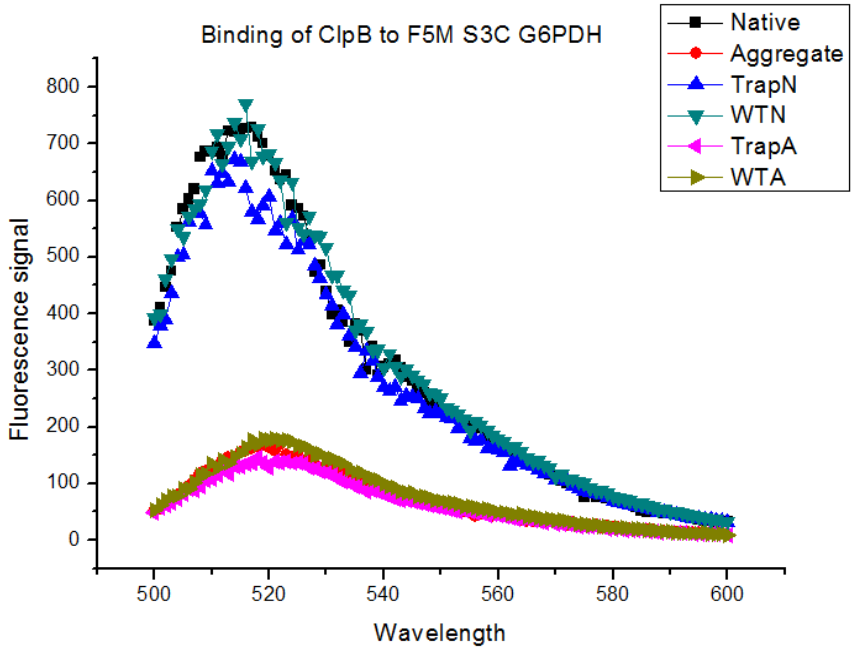


Fig. 11 Aggregation of both G6PDH mutants inhibited fluorescence.

Conditions were the same as in filter assay described in Chapter 2.

WTA: WT ClpB+Aggregated G6PDH+ATPγS;

WTN: WT ClpB+Native G6PDH+ATPγS;

TrapA: Trap ClpB+Aggregated G6PDH+ATP;

TrapN: Trap ClpB+Native G6PDH+ATP

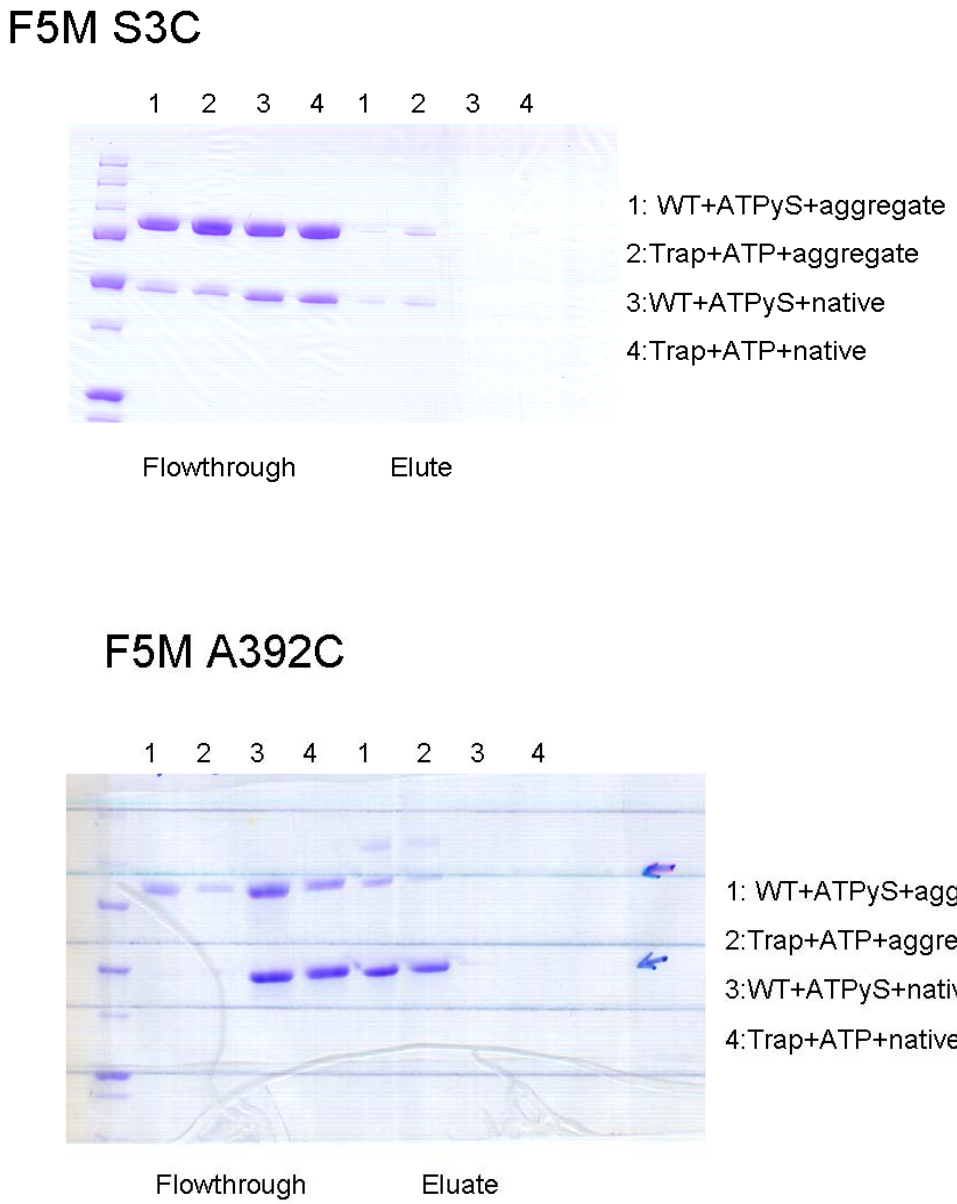
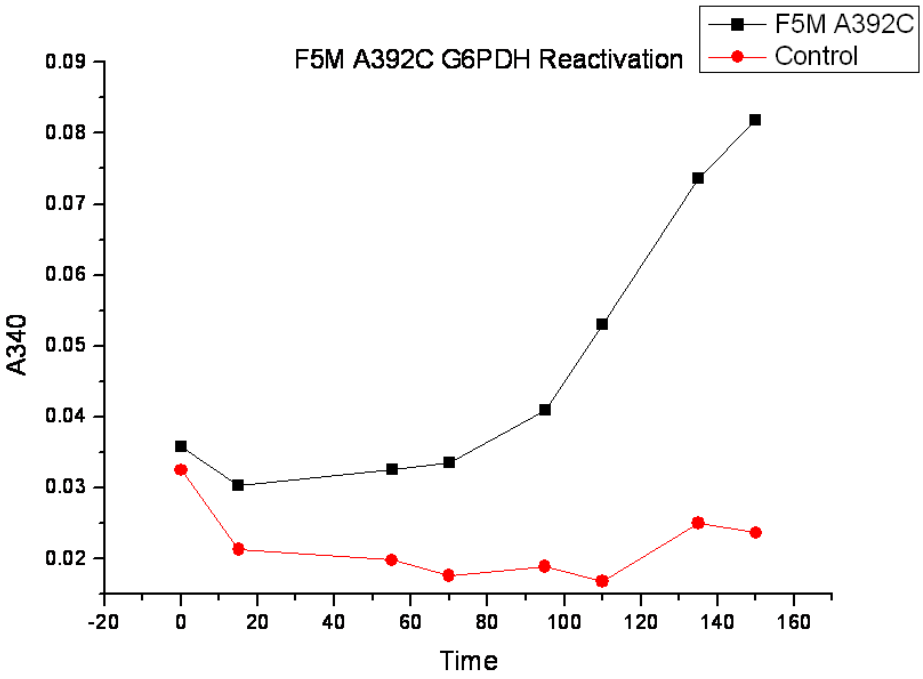


Fig. 12 Binding of wild type ClpB and Trap ClpB to labeled G6PDH aggregates. The same filter assay was used as described in Chapter 2. In the case of G6PDH A392C aggregate (bottom panel), the binding was similar to wild type G6PDH aggregate; while in the case of G6PDH S3C, the labeling seemed to prevent aggregation (top panel) since only a small amount of G6PDH S3C was left on the filter.



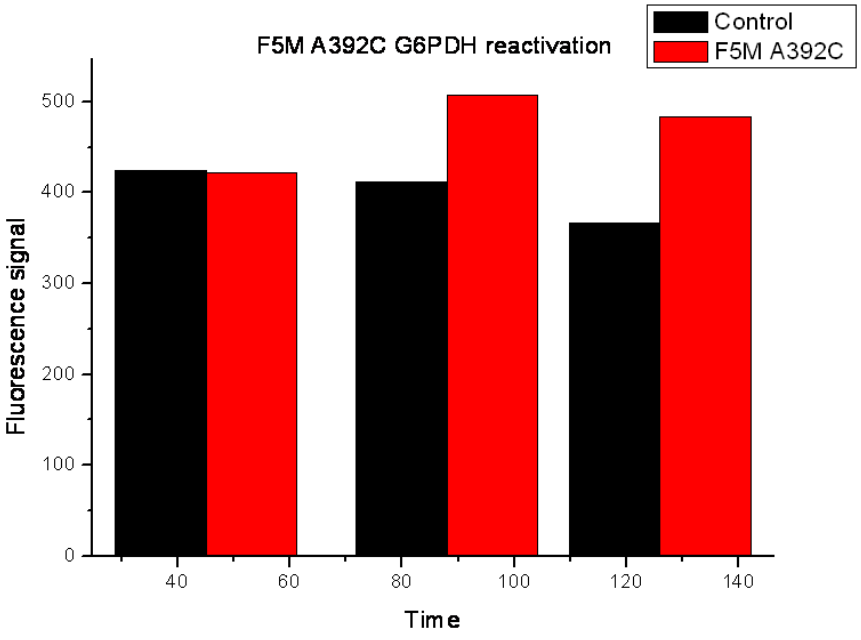


Fig. 13 Reactivation of labeled G6PDH A392C. The reactivation assay was performed as the reactivation of wild type G6PDH. Both enzyme activity and fluorescence were monitored over time.