Bacterial AAA+ Disaggregase ClpB: Mechanism and inhibition

by

Chathurange Bharathee Ranaweera

B.Sc., University of East London, 2009
M.Sc., University of East London, 2012

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Abstract

Molecular chaperones are essential proteins that bind to unfolded and partially folded polypeptide chains to prevent their aggregation and precipitation. Many molecular chaperones were first described as heat shock proteins (Hsp) because their expression is increased at elevated temperatures. Bacterial ClpB is a molecular chaperone, which is involved in reactivation of aggregated proteins. ClpB is the only known bacterial chaperone capable of suppressing and reversing protein aggregation in cooperation with other heat-shock proteins (DnaK, DnaJ, GrpE). In the presence of ATP, six ClpB monomers associate and form the hexameric functional unit of ClpB. Our understanding of the mechanism of substrate recognition by ClpB is still limited and is under investigation. In order to study substrate recognition by *E. coli* ClpB and substrate binding affinity, we used model peptide substrates in fluorescence anisotropy experiments. We hypothesized that the peptides containing predicted aggregation-prone sequences would preferentially bind to ClpB. Our results demonstrated that substrate recognition mechanism of ClpB may build upon global surface properties of aggregated proteins rather than on local sequence motifs. Next, we explored a potential of ClpB as a target for the development of novel antimicrobials. We investigated ClpB activity in the presence of several previously described inhibitor candidates and identified N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ) as a promising lead compound for future structural optimization aimed at selective targeting of ClpB and/or DnaK. We also studied protein aggregation patterns in two obligate intracellular pathogenic bacteria, *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. We performed the first biochemical characterization of ClpB and DnaK from *Anaplasma phagocytophilum* and investigated substrate recognition mechanisms of ClpB and DnaK from *Anaplasma phagocytophilum* with similar conclusions to those obtained with *E. coli* chaperones. Finally, we

demonstrated that DBeQ displays a potent activity against both *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* in infected cell cultures. Thus, DBeQ can be used as a lead compound for further development of novel treatments for Human Granulocytic Anaplasmosis (HGA) and Human Granulocytic Ehrlichiosis (HGE).

Bacterial AAA+ Disaggregase ClpB: Mechanism and inhibition

by

Chathurange Bharathee Ranaweera

B.Sc., University of East London, 2009
M.Sc., University of East London, 2012

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Approved by:

Major Professor
Michal Zolkiewski

Copyright

Abstract

Molecular chaperones are essential proteins that bind to unfolded and partially folded polypeptide chains to prevent their aggregation and precipitation. Many molecular chaperones were first described as heat shock proteins (Hsp) because their expression is increased at elevated temperatures. Bacterial ClpB is a molecular chaperone, which is involved in reactivation of aggregated proteins. ClpB is the only known bacterial chaperone capable of suppressing and reversing protein aggregation in cooperation with other heat-shock proteins (DnaK, DnaJ, GrpE). In the presence of ATP, six ClpB monomers associate and form the hexameric functional unit of ClpB. Our understanding of the mechanism of substrate recognition by ClpB is still limited and is under investigation. In order to study substrate recognition by *E. coli* ClpB and substrate binding affinity, we used model peptide substrates in fluorescence anisotropy experiments. We hypothesized that the peptides containing predicted aggregation-prone sequences would preferentially bind to ClpB. Our results demonstrated that substrate recognition mechanism of ClpB may build upon global surface properties of aggregated proteins rather than on local sequence motifs. Next, we explored a potential of ClpB as a target for the development of novel antimicrobials. We investigated ClpB activity in the presence of several previously described inhibitor candidates and identified N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ) as a promising lead compound for future structural optimization aimed at selective targeting of ClpB and/or DnaK. We also studied protein aggregation patterns in two obligate intracellular pathogenic bacteria, *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. We performed the first biochemical characterization of ClpB and DnaK from *Anaplasma phagocytophilum* and investigated substrate recognition mechanisms of ClpB and DnaK from *Anaplasma phagocytophilum* with similar conclusions to those obtained with *E. coli* chaperones. Finally, we

demonstrated that DBeQ displays a potent activity against both *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* in infected cell cultures. Thus, DBeQ can be used as a lead compound for further development of novel treatments for Human Granulocytic Anaplasmosis (HGA) and Human Granulocytic Ehrlichiosis (HGE).

Table of Contents

List of Figures

List of Tables

Acknowledgements

First of all, I want to convey my astounding and heartfelt gratitude to my major professor, Dr Michal Zolkiewski for his incredible continuous support and guidance throughout my PhD studies at Kansas State University. I am eternally grateful to him for accepting me as a graduate student in his research group and allowing me to develop my skills freely and independently. Having arrived here in Manhattan Kansas, leaving a young family behind My advisor made sure to give me enough time to get settled down and start my studies smoothly. He provided me a safe, secure and happy research environment which allowed me to grow as a research scientist and as a good human being. It has been an absolute privilege to have you as my major professor.

Next, I want to thank Dr Roman Ganta for allowing me to work in his lab, sharing his knowledge and expertise throughout my PhD journey. The environment at MZ and RG labs encouraged and allowed me to test science without fear of failure and pursue something I am truly passionate about. I would like to extend my sincere thanks to my other committee members, Dr Brian Geisbrecht and Dr Om Prakash for taking time to serve on my committee and providing me with invaluable advices from my preliminary examination to the very last committee meeting. At the same time, I would like to thank Dr Masaaki Tamura for serving as the chair for my doctoral defense.

My sincere thanks go to all current and previous lab members of Michal Zolkiewski lab group and Roman Ganta lab group. It is an absolute privilege to meet you and work along with you all. Countless lunch outs and lab meetings are wonderful memories that I will carry for the rest of my life. These two labs are truly a safe haven for me. I am glad that I ended up at Manhattan Kansas and had the chance to work in these two great labs with these wonderful individuals. In no particular order I want to mention and thank Dr Przemyslaw Glaza, Dr Sunitha Shiva, Dr Lynn

Schrag, Dr Indhujah Thevarajan, Swetha Madesh and Zachary Spaulding for everything you have done for me throughout my studies. I will never forget what you all have done for me.

I want to express my sincerest gratitude to the staff and faculty of the entire Biochemistry and Molecular Biophysics Department at Kansas State University for providing the facilities, education and support to pursue my studies. Most importantly I want to thank all my teachers, colleagues and great friends who stood by me and supported me from my education at Rahula College Matara Sri Lanka, University of East London, Bielefeld University of Applied Sciences, General Sir John Kotelawala Defence University Sri Lanka and Kansas state University.

I would not be here if it were not for the free education and healthcare system in Sri Lanka supported by all hard-working generous Sri Lankans. I want to say thank you to all these unknown individuals and hope that I can repay part of the debt one day. I also want to say thank you to all hard-working generous taxpayers in America for funding my studies and allowing me to explore and enjoy new avenues in science. You all made me realised and showed me that science could break any barrier, connect and unify all of us for a common goal.

Finally, I want to express my overwhelming and sincerest appreciation to my beautiful family, my wife Anomali Vidanagamage , my son Lisas Ranaweera ,my mother Pushpa Jayawardena, my father Siri Ranaweera, my late grandparents Nelly Wickramasinghe Jayasekara and Don Samuel Abeysinghe Jayawardena , my parents-in-law Ananda Vidanagamage and Udithaa Gajadeera, Aunty Guni and Ranjanee, aunty Kanthie Fernando, aunty Amala Mitchelle, aunty Indrani Jayawardena, Uncle Jeevana Fernando ,Daya Jayawardena and uncle Tissera, my siblings Udayanga Ranaweera, Sathiranga Ranaweera, Sandasiya Ranaweera, Uditha Fernando, Dinusha Fernando and Sandali Tissera. You all stood by me, looked after me and family and gave

your unconditional love and support allowing me to pursue my studies abroad. Thank you all, I will forever be in your debt.

Dedication

To My Beautiful Family.

For Everything you all have done for me,

Thank you all,

I will forever be in your debt.

Chapter 1 - Introduction.

1.1 Evolution and biological functions of molecular chaperones.

Proteins are the most versatile and structurally complex macromolecules found in any biological living system. These macromolecules perform many essential biological functions and are involved in most biological reactions that occur in a living organism. Therefore, it is very important for all organisms to maintain a proper balance in protein homeostasis, such as protein synthesis, protein folding, protein refolding, and protein degradation. One of the most important functions of this protein quality control and a maintenance of proteome homeostasis (also known as proteostasis) is to prevent or minimize protein aggregation and maintain proteins in a soluble active state. In a living cell, protein quality control and proteostasis is supported by a large number of proteins, most prominently by molecular chaperones and their regulators. These proteins assist de novo protein folding and refolding under stress, but also include the ubiquitin proteasome system (UPS) and the autophagy system which are responsible for timely removal of misfolded/aggregated proteins (Hartl, Bracher et al. 2011).

Protein aggregation is defined as non-physiological association of partially folded or misfolded polypeptides. In order to prevent protein aggregation and promote protein folding, cells have adopted a special set of proteins known as molecular chaperones. A molecular chaperone can be defined as any protein which helps, promotes, and stabilizes other protein molecules to reach their biologically active native conformation without being present in the final active structure of the client protein (Hartl, Bracher et al. 2011, Zolkiewski, Zhang et al. 2012).

In vitro, a majority of small proteins can refold on its own without any energy input as denaturants are removed from the solution. This signifies the fact that the amino acid sequence of these small proteins contains all necessary information for protein folding and reaching its final

active conformation (Dobson, Šali et al. 1998). However, it has been observed that, despite the fact that small proteins can fold into the native active structure relatively fast, large multi domain proteins tend to take longer times to fold properly and most often these large proteins fail to fold properly *in vitro*. The protein folding process becomes even more challenging in a cellular environment which is crowded with other protein molecules and where proteostasis needs to be maintained properly at all times (Hartl, Bracher et al. 2011).

Research over the past few decades had firmly established a fundamental fact that the majority of proteins in a cell require assistance of molecular chaperones to reach their biologically active native structure. Apart from the involvement in protein folding, molecular chaperones are important in prevention of formation of protein aggregates, re-solubilization of protein aggregates, and are involved protein degradation as well. It is important to note here that protein aggregation and misfolding tend to occur more frequently during cellular stress events and it is not surprising that all heat shock proteins (Hsp) contain an inherent chaperone activity (Bukau, Weissman et al. 2006). At the same time, when considering the evolution of life, it is possible that early cells required a way to fold proteins relatively quickly and efficiently while keeping all proteins soluble as cells become more crowded with various molecules. Hence, it is fair to assume that chaperones may have evolved early to perform these specific requirements in a crowded cellular environment (Hartl, Bracher et al. 2011).

In a cellular milieu, as unstructured polypeptides emerge from ribosomes, there a is a chance that some polypeptides may end up misfolded while a majority are correctly folded into the active native state (Figure 1.1). At the same time, when a cell is under stress, certain proteins could lose its native structure, becoming misfolded. However, as mentioned earlier, cells contain molecular chaperones to help in assisting polypeptides/proteins to fold into their native state or

guide misfolded polypeptides/proteins to fold back into the active native structure. For this purpose, cells produce molecular chaperones such as GroEL (Hsp 60), DnaK (Hsp 70), DnaJ (Hsp 40) and GrpE (nucleotide exchange factor). It is important to note that protein aggregation is an irreversible process and once protein aggregates are formed, they may accumulate in the cytoplasm and form inclusion bodies. Hence, rapid removal of protein aggregates is essential for the cell survival as inefficient aggregate removal could build up large clumps of proteins inside the cells, leading towards protein toxicity or even cell death.

Interestingly, certain cells have evolved a certain type of molecular chaperone which has the unique ability to break down protein aggregates and convert them into unstructured polypeptides (a reversal of protein aggregation). These molecular chaperones are known as Caseinolytic Peptidase B (ClpB) in bacteria, Hsp104 in yeast and Hsp101 in plants. In literature, these proteins are generally termed as Hsp100 family since their monomers have a molecular weight around 100 kDa (kiloDaltons). Once the aggregates are broken into unstructured polypeptides, these polypeptides either can be folded back into active proteins (with the help from other molecular chaperones) or can be degraded by the protease machinery in the cell (Doyle, Hoskins et al. 2012).

It is worth mentioning here that, while Hsp100 family proteins are found in bacteria, yeast and plants, they are absent in animals and in humans. However, DnaK /Hsp70 family, which can break down small aggregates on its own, is found in all domains of life (Karlin, Mrázek et al. 2005, Zolkiewski, Zhang et al. 2012, Mattoo, Sharma et al. 2013). This indicates that different cells have evolved their chaperones in different ways to maintain the equilibrium between folded and misfolded proteins while ensuring that at least one type of chaperone is present at all the time in all domains of life.

1.2 Protein folding, misfolding, and the associated diseases.

Proteins are required to achieve their native active state in order to perform their intended cellular functions. In 1973, Christian B. Anifinsen demonstrated that the primary amino acid sequence of a protein contains all necessary information for a protein to be folded correctly (Anfinsen 1973). At the same time, it has been shown that small proteins may fold relatively quickly (microseconds) without any assistance, while larger multi-domain proteins take a longer time (minutes to hours) to reach their native active state (Hartl, Bracher et al. 2011). Hence it is obvious that these larger multi domain proteins require extra assistance to fold quickly to perform their intended biological functions efficiently. The required assistance for correct protein folding is provided by molecular chaperones. As a protein folds, local regions within the polypeptide/s tend to have inherent secondary structure preference (alpha helices, beta strands, etc) and these local regions could initiate the folding process. As secondary structures are being built within a polypeptide/s, these secondary structures could start interacting with each other and start to stabilize the final shape of the protein. These steps in protein folding are proposed by the nucleation condensation model (Gianni, Guydosh et al. 2003). Therefore, protein folding is a highly cooperative process where native state is achieved via progressive stabilization of secondary structure intermediates, rather than by a random search. As the protein folds, hydrophobic side chains of amino acids are buried inside the core of the protein leaving hydrophilic sides chains exposed to water (Camilloni, Bonetti et al. 2016).

As mentioned earlier, as a protein folds by generating secondary strictures, there is a chance for the formation of non-native conformations as well. If these non-native states have exposed hydrophobic regions, then they could associate with each other to form protein aggregates. It is

worth mentioning here that that incorrect protein folding is likely to occur as proteins are folded in a crowded cellular environment. These incorrectly folded polypeptides are known as misfolded proteins. Misfolded proteins either need to be refolded correctly by chaperones or degraded by proteases/proteasomes.

Once misfolded proteins accumulate inside cells, they lead to a development and progression of many diseases with devastating impact on human health. Interestingly, misfolded proteins are responsible for the majority of diseases that are not caused by an infectious agent. Generally, pathologies associated with misfolded proteins are categorized into two main groups: the loss of function and the toxic gain of function diseases. Sickle cell anemia and cystic fibrosis are categorized as loss of function diseases while Alzheimer disease (AD) and Parkinson disease (PD) are categorized as toxic gain of function diseases (Valastyan and Lindquist 2014, Hartl 2017). In fact, sickle cell anemia is the first described inherited disease with a known molecular mechanism (a missense mutation, Glutamine 6 to Valine in the amino acid sequence of hemoglobin beta chain). This single point mutation in the beta hemoglobin chain causes the protein misfolding in homozygous individuals. However, it is important to note that the relationship between genetic changes, protein misfolding and disease onset is not well understood and is still under investigation.

Cystic fibrosis is an inherited genetic disease caused by mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene. The most common variant is caused by deletion of Phe508 in the CFTR amino acid sequence. This mutation produces a misfolded version of CFTR protein. Misfolded CFTR leads to defects in transmembrane ion transport, which produces a thick, sticky mucus that can block lung airways. Generally, this misfolded CFTR is degraded at a rapid rate by autophagy and endoplasmic-reticulum-associated protein degradation

(ERAD) (Valastyan and Lindquist 2014). It has been shown that interactions with chaperones are essential for the maturation of wildtype CFTR protein. Disrupting interactions of co-chaperones such as Hsp90 and Aha1 (Activator of Hsp90 ATPase activity 1) allows the mutant CFTR escape the degradation pathways and allows the mutant protein to partially function as the wildtype CFTR. This is one example where partial inhibition of certain chaperones could be beneficial for patients when treating for disease such as cystic fibrosis (Valastyan and Lindquist 2014).

In another group of diseases (Alzheimer's disease (AD), Parkinson's disease (PD)), partially folded or misfolded proteins form aggregates which eventually become toxic to the cells. Both AD and PD are neurodegenerative diseases in which protein aggregation could be caused by heritable mutations. In both cases, disease onset and manifest are age dependent and are facilitated by a decline of proteostasis network as a result of aging (Hartl 2017). AD disease is characterized by the presence of protein aggregates within the surrounding of brain neurons. These protein aggregates are known as amyloid plaques and are formed from amyloid beta peptides (Aβ). Accumulated Aβ plaques are believed to induce toxic effects which eventually destroy brain neurons leading to the loss of memory and communication abilities (Ashraf, Greig et al. 2014).

Parkinson disease affects mid brain cells which are responsible for the production of a neurotransmitter dopamine. The exact cause for this disease is not fully understood yet. PD patients often manifest intracellular protein aggregates known as Lewy Bodies (LBs), in the cells found in the midbrain regions. The major component of LBs is a protein called alpha synuclein. Two mutations (A53T and A30P) in alpha synuclein gene are responsible for major conformational changes in the protein structure which drives the mutated protein towards self-aggregating states. These self-aggregated proteins form irreversible aggregates in the cells which become toxic and cause cells death within the midbrain region. It is believed that these missense mutations alter the

alpha synuclein degradation pathway, allowing for accumulation of protein aggregates which eventually become toxic to the cells (Forloni, Terreni et al. 2002).

Apart from the diseases discussed above, which are associated with protein misfolding and aggregate formation, there are many other examples of similar pathogenicities and more are getting discovered on a regular basis. However, treatments, early diagnosis and other therapeutic interventions remain at their infancy. Therefore, it is essential to understand underlying mechanisms involved in diseases associated with protein misfolding and aggregation and their pathogenesis to open new avenues for drug discovery and better treatment options for the patients.

1.3 ClpB participates in a multi-chaperone network.

ClpB is a bacterial chaperone which belongs to Heat Shock Protein 100 family (Hsp100). Hsp100 proteins belong to the AAA+ (ATPase associated with various cellular activities) family and hence they require energy from ATP hydrolysis to perform their cellular functions. In a cellular environment, proper balance of protein homeostasis is essential for the maintenance of optimal cellular functions and ClpB plays a vital role in the bacterial protein homeostasis by reactivation of aggregated proteins. It has been a well-established fact that the ClpB-mediated aggregate reactivation and breakdown mechanism is linked with two other major molecular chaperones (DnaK and DnaJ) and a nucleotide exchange factor (GrpE) which are believed to be working in cycles (Zolkiewski, Zhang et al. 2012, Ranaweera, Glaza et al. 2018). *E. coli* DnaK (Hsp70) is one of the most abundant and constitutively expressed protein found in the cytosol. DnaK acts as the major central hub for *E. coli* chaperone network (Calloni, Chen et al. 2012). *E. coli* DnaK plays a vital role in protein folding and refolding while preventing formation of protein aggregates. At the same time, it is also believed that DnaK is essential for the activation of ClpB and driving ClpB towards cellular aggregates (Acebrón, Martín et al. 2009, Schramm, Heinrich et al. 2017).

Experimental data (Miot, Reidy et al. 2011) showed that the DnaK nucleotide binding domain interacts with the middle domain of ClpB and this interaction induces ClpB activation (figure 1.2).

DnaJ (Hsp40) also plays a vital role in the ClpB-mediated aggregate reactivation and breakdown cycle by helping DnaK to bind to substrate or by delivering the substrate to DnaK. It has been shown that DnaJ binds to DnaK and accelerates its ATP hydrolysis by delivering the protein substrate to DnaK (Kampinga and Craig 2010). The nucleotide exchange factor GrpE binds to DnaK and removes adenosine 5′-diphosphate (ADP) from the DnaK nucleotide binding domain while replacing it with ATP. It has been well documented that substrate release mechanism of DnaK depends on the DnaK ATP cycle and hence the GrpE nucleotide exchange activity plays a vital role in delivering the substrate to ClpB (Harrison 2003) .

1.4 Structure of Bacterial ClpB.

Bacterial ClpB belongs to the AAA+ (ATPase Associated with various cellular Activities) superfamily and is found in both prokaryotic and eukaryotic cells. ClpB is a molecular chaperone which utilizes the energy from ATP hydrolysis to disassemble large complexes of aggregated proteins. The biologically active unit of ClpB is a hexamer (≈570 kDa) which is formed in presence of nucleotides (ATP or ADP). Six ClpB monomers (≈95 kDa, figure 1.3) are assembled to form the biologically active hexamer of ClpB in a cellular milieu (Zolkiewski, Zhang et al. 2012). However, it is worth mentioning here that a heptameric assembly of ClpB had been reported under non-physiological conditions as well (Kim, Cheong et al. 2000, Akoev, Gogol et al. 2004). Therefore, it is believed that binding of ATP induces a conformational change in self associating monomers which drives the ClpB assembly from a heptameric state towards a biologically active hexameric unit. When six monomers are assembled together, ClpB forms a two-tiered ring-shaped hexamer with a central open channel (figure 1.4).

Each *E. coli* ClpB monomer contains 857 amino acids (AA) and the domain organization of a monomer is shown in figure 1.4. The N-terminal domain is made of the first 147 amino acids (AA) followed by a linker (from 147 AA to 163 AA) connecting the N terminal domain and first ATP binding domain (from 164 AA to 413 AA). The middle domain (from 414 AA to 525 AA) is located between the first and the second ATP binding domains (from 526 AA to 769 AA). The middle domain is a unique feature of ClpB and is not found in any other AAA+ proteins. The C-terminal domain is made of the last 87 amino acids (from 770 AA to 857 AA) giving each monomer a molecular weight around 95 kDa.

Mobile and globular shaped N terminal domain is made of alpha helices and contributes to a detection and binding to aggregates. It has been shown that N-terminal domain is not essential for disaggregation of certain substrates. However, it becomes essential for optimal bacterial growth in the absence of a functional DnaK (Chow, Barnett et al. 2005). By designing elegant Nuclear Magnetic Resonance (NMR) experiments, Rosenzweig and colleagues demonstrated that there is a substrate binding patch in the N-terminal domain, and they showed that this substrate binding patch is made of aliphatic hydrophobic amino acids (Rosenzweig, Farber et al. 2015). The NMR results demonstrated that the N-terminal domain detects and interacts with hydrophobic patches in client proteins and that substrate interaction with the N-terminal domain elevates the ATPase activity of ClpB. Most importantly, Rosenzweig and colleagues established that the N-terminal domain plays a regulatory role in the ClpB-substrate interaction and subsequent translocation steps. They suggested that the N-terminal domain is responsible for blocking the channel entrance in a ClpB hexamer, thus it is responsible for a correct recognition of client proteins by preventing the engagement of properly folded proteins by ClpB. *In vitro* data with an N-terminally truncated ClpB demonstrated that the truncated versions can still form active hexamers with a basal ATPase

activity not affected by the truncation. However those data showed that the truncation had a significant effect (decreased activity) on casein-activated ATPase activity (Barnett, Zolkiewska et al. 2000). The unstructured linker is located between the N-terminal domain and the first nucleotide binding domain (NBD1) and is responsible for dynamic movements of the N-terminal domain.

Bacterial ClpB contains two nucleotide binding domains (NBD1 & NBD2) which are believed to have evolved independently from each other. These nucleotide binding domains generally have a similar structure and contain a number of conserved sequence motifs, including Walker A, Walker B, sensor 1, sensor 2 , and Arginine finger (Zolkiewski, Zhang et al. 2012). Walker A and Walker B motifs are important sites for ATP binding and hydrolysis. Walker A directly interacts with the phosphate groups of ATP while Walker B is essential for coordination of Mg^{2+} ion and water which are essential for ATP hydrolysis (figure 1.5). Lysine is a key conserved residue in Walker A motif which is essential for ATP binding and mutations of this residue eliminate ATP binding and inhibit the AAA+ protein activities. The conserved lysine is located in the following Walker A consensus sequence GXXXXGK[T/S] where X represents any amino acid (Hanson and Whiteheart 2005). Walker B motifs also form contacts with bound nucleotides and contain two key conserved residues in its structure: aspartate and glutamate. The aspartate residue is responsible for proper coordination of Mg^{2+} ion. The glutamate residue is responsible for activation of water molecules which is required for catalysis of the ATP hydrolysis reaction. Mutations of the Walker B glutamate blocks bound nucleotide hydrolysis but does not affect a substrate or ATP binding to the AAA+ proteins. Both the conserved aspartate and glutamate are found in the following consensus sequence (hhhhDE, where h represents a hydrophobic amino acid) in Walker B motif. Since mutations of the Walker B glutamate preserves

ATP and substrate binding to ClpB but not the ATP hydrolysis, these mutations have been used to generate a substrate trapping version of ClpB for research purposes. These substrate trap versions of ClpB are unable to hydrolyze bound ATP & release bound substrates. These substrate trap versions of ClpB have become efficient biological tools to study general functions of AAA+ proteins and in designing small-molecules inhibitors against specific AAA+ proteins (Hanson and Whiteheart 2005).

Besides the motifs described above, AAA+ proteins have a consensus region which is located next to C-terminus of Walker B motifs. This region is known as the Second Region of Homology (SRH). At the N-terminus of SRH there is a small motif known as Sensor 1 and at the C-terminus of SRH, there is another motif known as Arginine finger. Polar residues found in Sensor 1 motif are important in ATP hydrolysis as these polar residues are located in between Walker A and Walker B regions (Figure 1.5). Mutation of the Sensor 1 polar residues are known to impair ATP hydrolysis reactions in NBDs. The Arginine fingers are important for the formation of nucleotide binding pocket in the active ClpB hexamer (nucleotide binding pocket is located between the interface of adjacent monomers). In all known structures of AAA+ proteins, the nucleotide binding pocket is formed between the interface of adjacent monomers and thus bacterial ClpB hexamer has twelve (12) nucleotide binding pockets (six in each NBD). The Arginines found in arginine finger motifs are essential for ATP binding, ATP hydrolysis and ATP dependent oligomerization of AAA+ proteins. In addition, Sensor 2 motif with a key conserved arginine is located at C-terminal end of nucleotide binding domain. Based on the structural studies, the arginine in Sensor 2 is in direct contact with the gamma phosphate of bound ATP. Therefore Sensor 2 is essential for ATP binding and hydrolysis mechanism associated with AAA+ proteins (Hanson and Whiteheart 2005).

Once ClpB monomers have assembled to form a biologically active hexamer, the central pore/channel is formed and lined by the residues from each NBD. These residues are arranged in a way where central channel is surrounded by loops from each monomer (two loops for each monomer, one loop per each NBD). There are three conserved key residues found in these pore loops which are essential for substrate binding and substrate translocation in ClpB. Generally, these pore loops have conserved aromatic-hydrophobic/hydrophilic-glycine residues in the pore lining regions where the aromatic residue is essential for substrate binding while the glycine residue contributes to the pore loop flexibility and substrate translocation in ClpB. In all Hsp 100 proteins, the central pore loop aromatic residue is tyrosine (Y). This conserved tyrosine is crucial for substrate binding. Mutation of this conserved tyrosine residue eliminates substrate binding ability of all Hsp100 proteins. In bacterial ClpB, these pore loop residues are Tyr251 (first NBD) and Tyr653 (second NBD) (figure 1.6 and 1.7). There are twelve (12) pore loops found in hexameric ClpB (two loops per monomer, figure 1.6 and 1.7). At any given point, when substrate is attached to ClpB, it has been shown that ten pore loops are in direct contact with bound substrate. Based on structural studies it has revealed that these pore loops are arranged as a spiral staircase around bound substrate (Rizo, Lin et al. 2019, Gates and Martin 2020).

Middle domain (MD) is a unique feature of bacterial ClpB which is not found in any other bacterial chaperone. The middle domain of ClpB has a coiled-coil structural arrangement and is wrapped around the outer edge of hexameric ClpB (figure 1.3 and 1.4). Previous work on the ClpB middle domain has revealed that MD plays a critical role in *in vitro* stabilization of oligomer assembly (Kedzierska, Akoev et al. 2003). Later Bernd Bukau and colleagues (Carroni, Kummer et al. 2014) showed that MD regulates ClpB ATPase activity and interacts with DnaK during activation of ClpB. By using elegant electron microscopic and cryo-electron microscopic data they

showed that MD adopts different positional arrangements in an active ClpB hexamer. They demonstrated that MD would have either horizontal configuration forming head to tail contacts among adjacent MD units in a hexamer or MD will have a slightly tilted configuration breaking head to tail contacts (figure 1.8). A horizontal arrangement of MD with head to tail contacts prevents interaction of ClpB with DnaK, thus repressing ClpB activity. A slight tilt of MD breaks the head to tail contacts of adjacent MDs and allows DnaK to bind to ClpB. Binding of DnaK to ClpB signals activation of ClpB, rapid increase in ATPase activity and aggregate disaggregation in a finely tuned manner. According to (Barnett, Zolkiewska et al. 2000), the carboxy terminal domain of ClpB (from 770AA to 857AA) plays a vital role in formation of ClpB oligomers. Their data revealed that deletion of N-terminal or C-terminal domain from ClpB has two opposite effects on ClpB. N-terminal domain deletion did not have any effect on hexamer formation of ClpB or its basal ATPase activity while C-terminal deletion had a detrimental effect on ClpB oligomer formation, its basal ATPase activity, and on the aggregate reactivation process.

1.5 Specific cellular functions of ClpB.

Bacterial ClpB and its yeast ortholog Hsp104 are required for thermotolerance and for cellular survival under stress conditions (Doyle, Genest et al. 2013). Yeast electron micrographic studies had revealed that Hsp104 was responsible for dissolving large aggregates in yeast cells as these cells were exposed to higher temperatures. In fact Hsp104 mutant strains were unable to dissolve these aggregates and were unable to survive for loner times under the same conditions (Parsell, Kowal et al. 1994). Similar to yeast, in bacteria, it has been demonstrated that ClpB is essential for induced thermotolerance and survival under stressed conditions. It has been shown in vivo , that in the absence of ClpB, bacterial cells tend to form large amounts of protein aggregates under stress conditions (Mogk, Tomoyasu et al. 1999).

It is important to note that although bacteria and yeast cells have their own protein degradation machineries, the removal of misfolded proteins is not sufficient for the cells to survive under severe stress conditions. This observation led to an interesting hypothesis that the recovery of aggregated proteins in cells could be vital for cell survival also under normal conditions. In fact, recently it has been shown that Hsp104 is essential for protection of thermo sensitive proteins from degradation under stressed conditions (Cabrera, Boronat et al. 2020). By using elegant in vivo experiments these authors showed these thermo sensitive proteins got localized into specific Protein Aggregate Centers (PAC) at the periphery of the cells, as a defensive mechanism that protects against protein degradation under stressed conditions. At the same time, the authors showed that these thermo sensitive proteins become resolubilized and recovered from PAC centers by Hsp104 during recovery to non-stressed conditions. Therefore, it is fair to assume that ClpB could perform a similar function in bacterial cells to that of Hsp104, i.e., protecting thermo sensitive proteins under stressed and non-stressed conditions.

In fact, it has been shown that the PAC centers are formed in bacterial cells under heat stress conditions, and they disappear as the cells are moved to normal growing conditions (Winkler, Seybert et al. 2010). Interestingly, from a separate study, it has been shown that ClpB chaperones have the ability to break down amyloid fibers when ClpB is co-expressed with DnaK, DnaJ and GrpE (Reidy, Miot et al. 2012). These results indicate the nature of the inherent power of ClpB to induce aggregate breakdown and protein resolubilization in a cellular environment. Recently Sander J. Tans and colleagues showed that ClpB can extract polypeptide loops from aggregates by generating a tremendous force while translocating both arms of the polypeptide loop through the central channel of ClpB (Avellaneda, Franke et al. 2020). They also showed that when ClpB encounters any disturbance in translocation of both arms of a polypeptide loops at once, the

chaperone can switch to translocation of a single arm of the loop. This arm switching mechanism helps ClpB (figure 1.10) to breakdown aggregates at a tremendous speed while generating a sufficient force for aggregate breaking. However, it is important to note here that even though these chaperones (ClpB in bacteria and Hsp104 in yeast) are essential for survival of these cells under stressed conditions, they are not essential for the growth and survival of the cells under normal conditions.

In summary, ClpB provides thermotolerance to bacterial cells, supports their survival under stressed conditions, mediates a breakdown of protein aggregates, and maintains cellular proteins in a soluble state where these proteins can be refolded with the help of other co-chaperones found in the cell.

1.6 Proposed mechanism of the ClpB mediated protein disaggregation.

The active physiological unit of bacterial ClpB is a hexamer (~ 575 kDa). In the presence of nucleotides (ATP or ADP) six ClpB monomers (each ~ 95 kDa) assemble in to an active hexametric ClpB unit with a six-fold symmetry (Figure 1.4 & (Zolkiewski, Zhang et al. 2012). As the monomers self-associate, they form a narrow channel at the center of the hexamer. Extracted polypeptides from aggregates are translocated through this central channel as ATP is hydrolyzed by ClpB (figure 1.9).

After encountering an aggregate, ClpB has the ability to pull an exposed end (N or C terminus) of the polypeptide or an exposed polypeptide loop from the aggregate through its central channel with the use of energy from ATP hydrolysis. It has been shown that DnaK is required for ClpB to detect protein aggregates and to direct ClpB towards aggregates within a cellular environment (Winkler, Seybert et al. 2010). Recent experimental data demonstrated that when extracting a polypeptide loop, ClpB has the ability to pull the peptide through its central channel

by both arms of the loop or it will switch to a single arm pulling if encounters resistance during translocation (figure 1.10). The ClpB pulling mechanism (loop extrusion) is believed to be rapid and relatively powerful, as it can generate forces over 50 pN at speeds over 500 residues per second when suitable conditions prevail (Avellaneda, Franke et al. 2020) .

As the extracted polypeptide exits the ClpB central channel, the substrate can start refolding on its own or it can be delivered to other co-chaperones (Dnak, DnaJ & GrpE or GroEL/ GroES) for a subsequent refolding process (Hartl, Bracher et al. 2011). However, if a peptide extraction from an aggregate becomes impossible, a ClpB hexamer will dissociate into its monomer units and will be reassembled again at the same location on the aggregate or at a different location on the aggregate, to engage a new exposed loop or a terminus. This structural flexibility of ClpB allows multiple ClpB units to work simultaneously on the same aggregate, making the aggregate breakdown a very rapid and an efficient process.

1.7 Therapeutic use of ClpB as a novel antimicrobial target.

Over the past few decades, with the advancement of scientific discoveries, many diseases either haven successfully treated, cured or eradicated even if once they were considered untreatable. Many novel antibiotics were discovered in the last century to target disease causing pathogens and have become an essential fighting tool against infectious diseases. However, during the past few decades, many bacterial pathogens developed resistance against the antibiotics due to their overuse and misuse. This phenomenon has led to a rapid proliferation of antibiotic-resistant bacterial strains, which has emerged as one of the most pressing threats to human health. According to latest data from Center for Disease Control and Prevention (CDC), there were more than 2.8 million antibiotic resistance cases each year and more than thirty-five thousand (35000) people die as result of these infections (https://www.cdc.gov/drugresistance/biggest-threats.html,

accessed on 01/31/2021). Hence there is an urgent requirement for the discovery of novel antibiotic targets and antibiotic molecules to tackle these problems.

In the post genomic era, with advanced instrumentation and novel intelligent detection methods, scientists managed to sequence a few hundreds of bacterial genomes and identify novel genes and gene products as possible antimicrobial targets. In order to be considered as a possible antimicrobial target, the target must be essential for the survival of a microorganism, ideally no homologous targets should be present in humans, the target must be druggable *in vitro* and in vivo with small molecules (Alksne and Dunman 2008). During last few years, few research groups have explored and managed to demonstrate that host pathogen stress response mechanism could become a novel antimicrobial target for drug development process (Martin, Underhaug et al. 2013, Zhang, Kedzierska-Mieszkowska et al. 2013, Torrente, Castellano et al. 2014, AhYoung, Koehl et al. 2015, Glaza, Ranaweera et al. 2020). At the same time, bacterial heat shock response has emerged as a very specific and unique target paving a new avenue for novel antibiotic discovery. When an infectious organism invades a host, the pathogen experiences sudden changes in pH, oxidative stress and heat shock. The survival and replication of the pathogen largely depend on the pathogen's molecular response against these different kind of stresses (Poole 2012).

Even though there are data to support a notion that pathogen stress response is a suitable novel target for antimicrobial development, so far, no successful inhibitors have been developed to target the heat shock response mechanism. Perhaps underlying reasons could be that heat shock proteins are presented in all domains of life and there is a high conservation of the amino acids sequences (figure 1.11) in most heat shock proteins. This sequence conservation may also indicate the essential role that these heat shock proteins perform in protein homeostasis in all domains of life (Zhang, Kedzierska-Mieszkowska et al. 2013). It is important to note that despite high heat

shock protein sequence conservation found in all domains of life, bacterial heat shock protein 100 (Hsp100) family can be further studied as a suitable antimicrobial target. As mentioned earlier bacterial ClpB belongs to the Hsp100 family and is essential for survival under stressful conditions. ClpB is an ideal antimicrobial target as it is not found in humans and animals. As ClpB orthologs are found in lower eukaryotes (fungi and protozoa), targeting these heat shock proteins opens new avenues to develop novel antifungal and antiprotozoal drugs as well (Zhang, Kedzierska-Mieszkowska et al. 2013, Glaza, Ranaweera et al. 2020). As shown in figure 1.11, ClpB is present in all ESKAPE pathogens (*Pseudomonas aeruginosa*, *Enterococcus faecium, Klebsiella pneumoniae, Acinetobacter baumannii, Staphylococcus aureus* and *Enterobacter* spp.) and malaria causing parasites. Hence it is fair to assume that inhibition of ClpB would suppress the infectivity and survival of broad range of pathogens and opens a unique avenue to develop wide variety of novel antimicrobials.

There are existing *in vivo and in vitro* data which show that ClpB is druggable by small molecules such as 5-(2-Chlorobenzyl)-2-hydroxy-3-nitrobenzaldehyde (C3 compound, structure is shown in appendix A supplementary figure A.1), Guanidine hydrochloride (GuHCl) and N^2,N^4 -dibenzylquinazoline-2,4-diamine (DBeQ, structure is shown in appendix A supplementary figure A.1) (Martin, Underhaug et al. 2013, Zeymer, Werbeck et al. 2013, Kuczynska-Wisnik, Cheng et al. 2017, Glaza, Ranaweera et al. 2020). Moreover, *in vivo* experiments with these compounds indicated that pathogen viability had been reduced significantly as well.

It is important to mention that when screening for small compounds against ClpB, one must make sure not to select compounds which affect ClpB basal ATPase activity as these compounds could affect other ATP-binding proteins. Therefore, an ideal compound/s would be a small molecule which could interfere with ClpB substrate recognition/interaction and its subsequent

allosteric response to substrate binding. Discovery of specific inhibitors against ClpB will help us to expand our existing knowledge on host-pathogen stress response mechanism and identify novel proteins which are important for survival and virulence. Identification of binding site/s in ClpB, and detailed studies on ligand binding may even help us to design selective novel compounds against other AAA+ proteins which are important in human health and well-being.

1.8 Goals and accomplishments of current research.

Recent Cryo Electron Microscopy (Cryo EM) data and structures have revealed new insights into the ClpB substrate attachment and translocation process (Rizo, Lin et al. 2019). However, the substrate recognition mechanism of ClpB remains poorly understood. How does ClpB distinguish an aggregate from a properly folded protein or even partially folded protein in a cellular environment? Are there any exposed motifs or patterns found on these aggregated proteins which are recognized by ClpB or by the co-chaperone DnaK? In order to address these questions, we have designed and used substrate mimicking peptides to investigate *in vitro* substrate recognition patterns of bacterial ClpB and DnaK (Chapter 2).

While ClpB and DnaK are found in many pathogenic bacteria, a complete biochemical characterization is still missing for ClpB and DnaK from obligate intracellular pathogens. To address this knowledge gap, we have cloned ClpB and DnaK from obligate intracellular pathogen, *Anaplasma phagocytophilum* and performed their full biochemical characterization (chapter 4). At the same time, we attempted to identify novel inhibitors of bacterial ClpB and the host-pathogen stress response mechanism. We have screened small inhibitor molecules against bacterial ClpB and identified N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ , structure is shown in figure 1.12) as a potential lead compound for our further studies (chapter 3). We have used DBeQ against *E. coli* and against obligate intracellular pathogens *Anaplasma phagocytophilum* and *Ehrlichia*

chaffeensis (chapter 5). We also tested DBeQ's potency against other bacterial AAA+ proteins, such as ClpA, ClpP, ClpX and studied DBeQ effects on the bacterial protein degradation machineries (appendix material) as well.

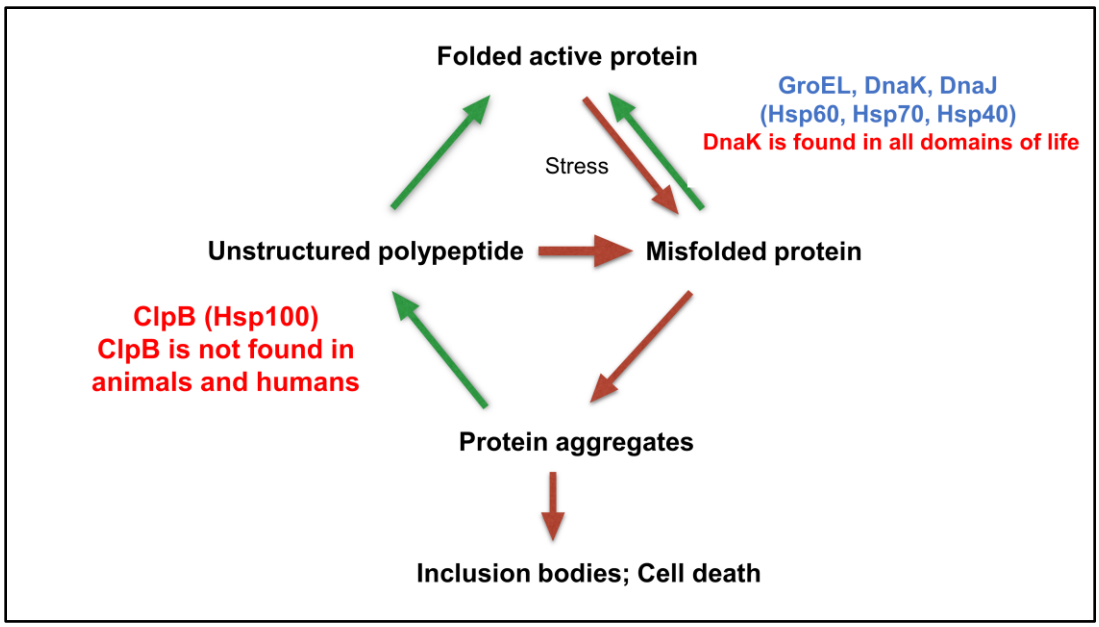


Figure 1.1. Model of the cellular protein folding, misfolding, and aggregation.

During de novo protein folding, misfolded proteins may get accumulated inside a cell, especially upon stressed conditions. Native proteins could lose its native structure and end up as misfolded proteins. These misfolded proteins have a higher tendency to associate with each other to form cellular aggregates. Accumulation of protein aggregates can lead to cell death. Chaperones such as GroEL, DnaK and DnaJ promote protein folding/refolding and prevent aggregation. Hsp100 family proteins have the ability to reverse protein aggregation process.

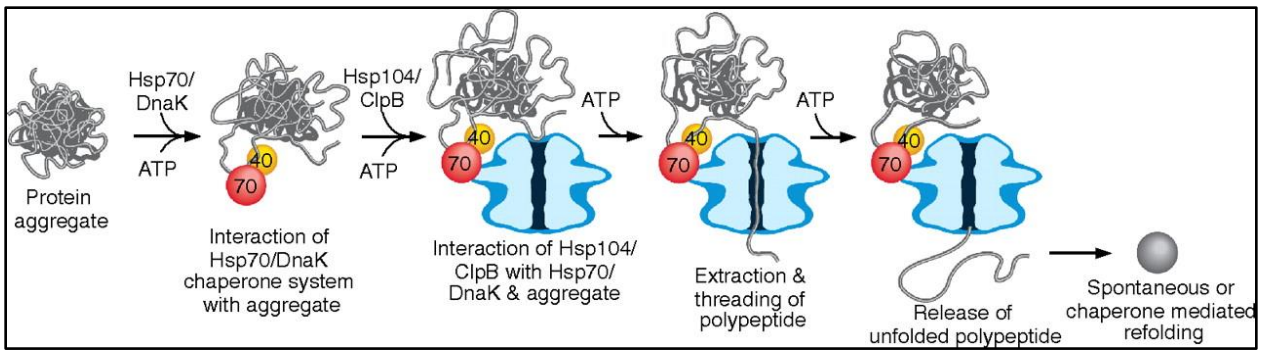


Figure 1.2. ClpB interaction with other co-chaperones.

An active unit of ClpB is a hexamer arranged as a cylinder with a central pore in the middle (blue color, Hsp104 is the ClpB homolog found in yeast). When aggregates are found in the cell, first DnaJ/Hsp40 would bind to aggregates and recruit DnaK/Hsp70. Binding of DnaJ and DnaK may loosen and remodel the aggregates to some extent. ClpB interacts with DnaK via its middle domain which results in ClpB activation and substrate delivery to ClpB. The aggregated polypeptides become unfolded during translocation through the central channel of ClpB. After translocation, the unfolded polypeptides are released from ClpB and can refold on their own or will be refolded with the assistance of other cellular chaperones. Reprinted with permission. Original figure is taken from (Miot, Reidy et al. 2011). PNAS April 26, 2011 108 (17) 6915-6920.

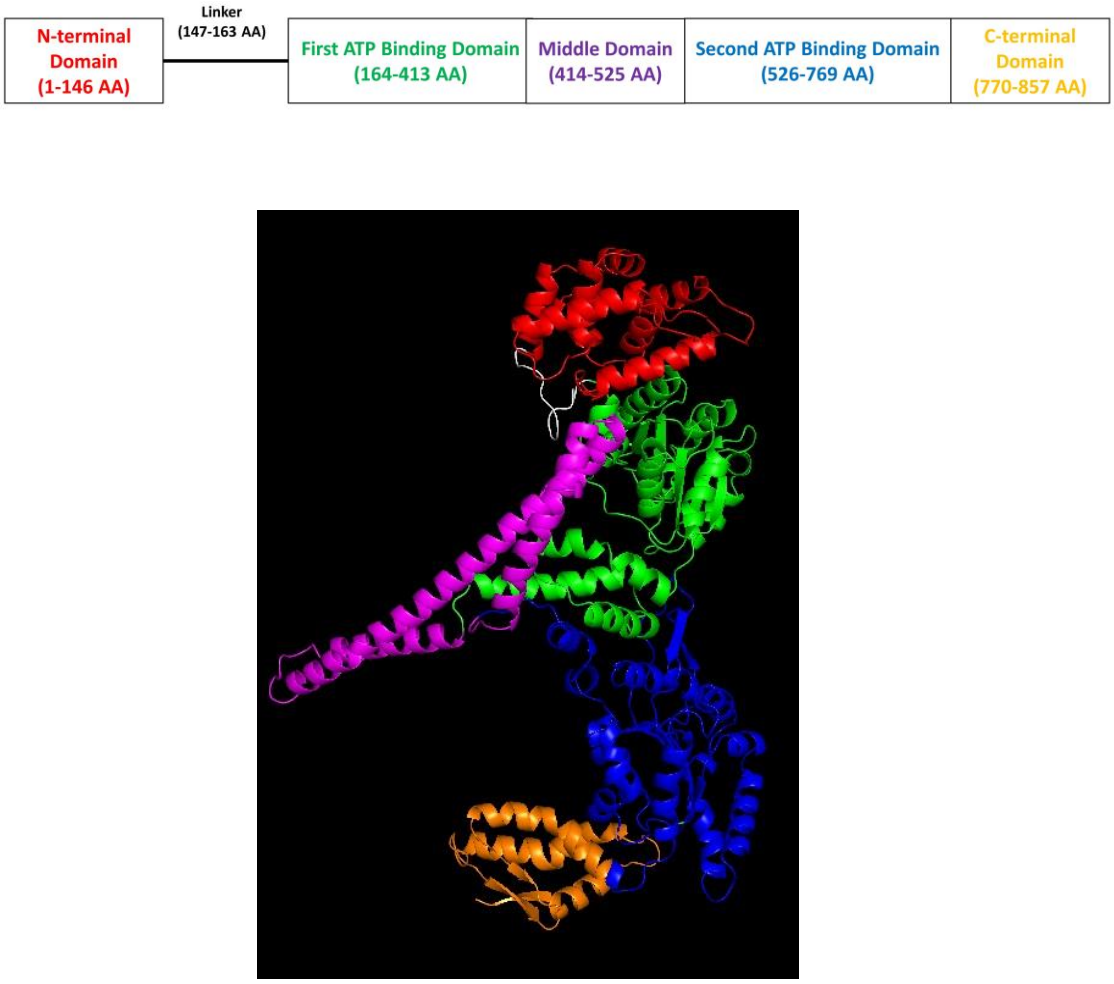


Figure 1.3. Domain organization of ClpB monomer.

E. coli ClpB monomer consists of 857 amino acids. The N terminal domain (red) is made of first 147 amino acids (AA) followed by a linker (white) connecting N terminal domain and first ATP binding domain (green). Middle domain (magenta) located between the first and second ATP binding domain (blue) and C terminal domain (orange) is made last 87 amino acids. ClpB monomer was built using PyMOL software, from published Structure of ClpB N-terminal domain (Li and Sha 2003) and Thermus thermophilus ClpB crystal structure (PDB Id: IQVR)(Lee, Sowa et al. 2003).

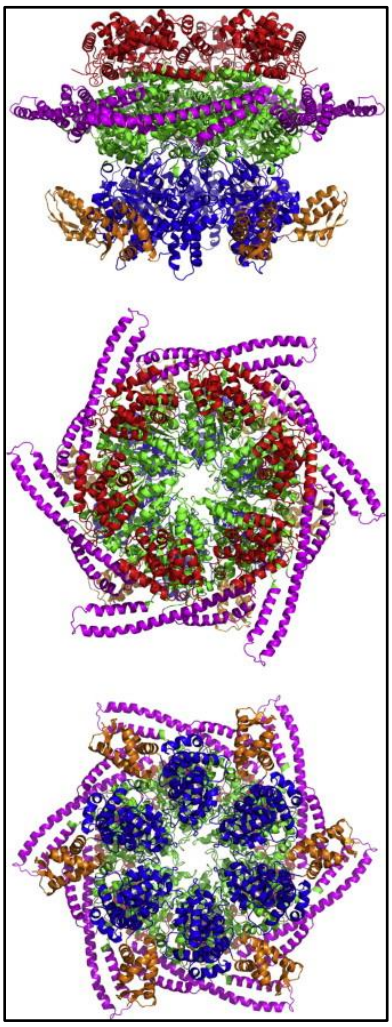


Figure 1.4. Structural model of biologically active ClpB from *E. coli*.

This structure was obtained by assembling six monomers into a ring-shaped assembly. The top, side and bottom views of the assembly are shown here. Individual domains of a monomer are colored as follows, N terminal domain -red, first ATP binding domain- Green, middle domain-magenta, second ATP binding domain-blue and C terminal domain-orange. Reprinted with permission. Original figure is taken from (Zolkiewski, Zhang et al. 2012).

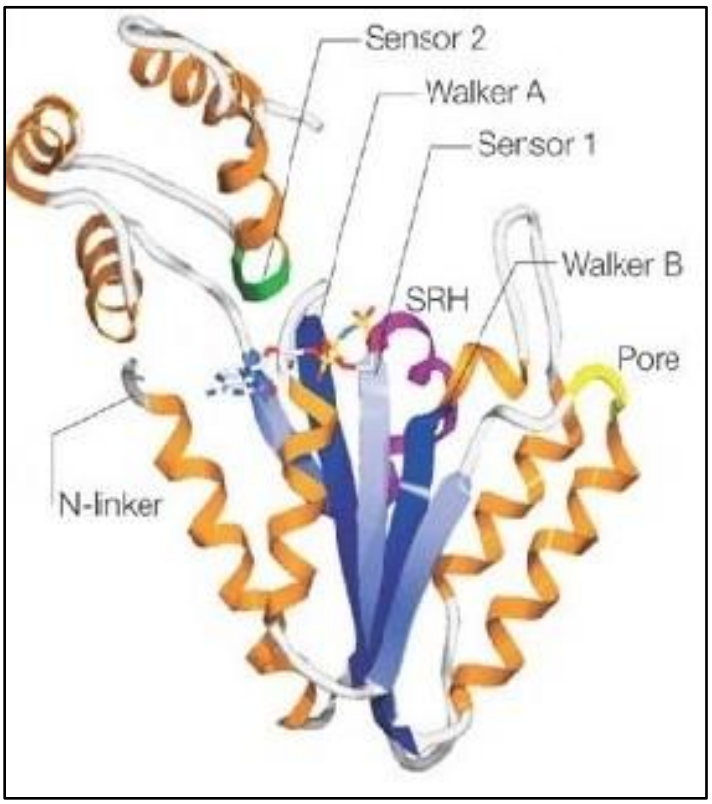


Figure 1.5. Structure of an AAA+ domain.

The crystal structure of second AAA+ domain of N-ethylmaleimide-sensitive factor (PDB Id:1D2N) is presented here. Bound nucleotide analogue (AMP-PNP) is shown in stick representation. Mg2+ ion (not shown here) is required for proper coordination of AMP-PNP molecule. SRH stands for Second Region of Homology. Picture is adapted from (Hanson and Whiteheart 2005). Reprinted with permission.

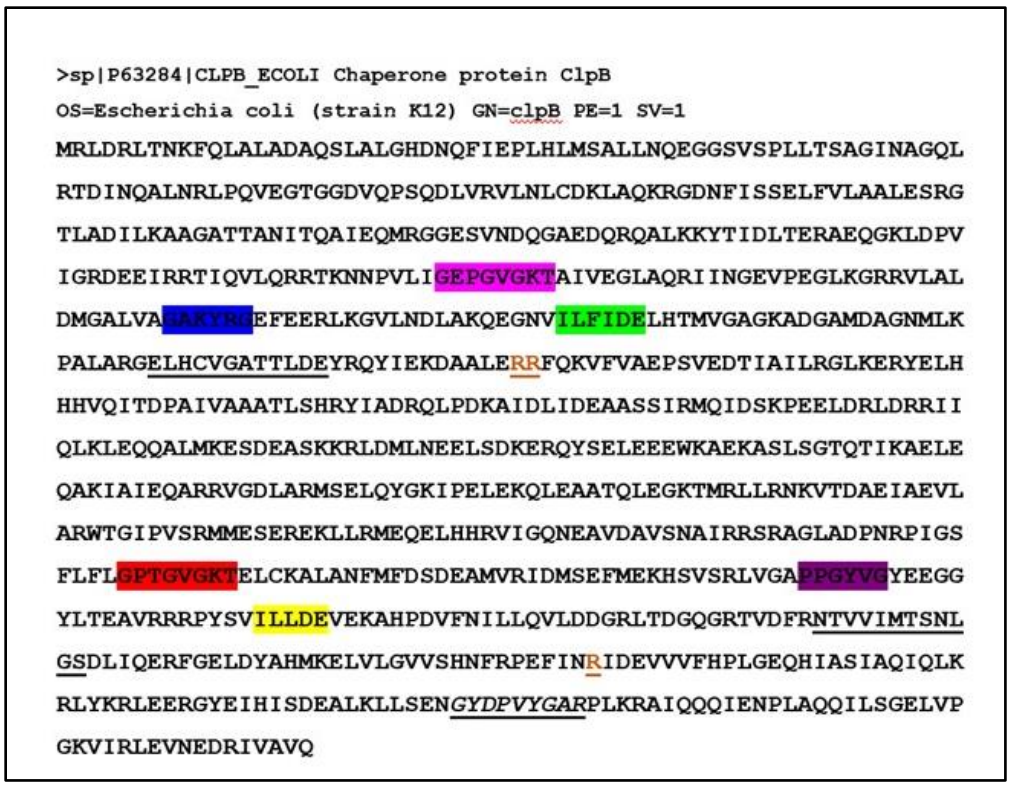


Figure 1.6. Amino acid sequence of ClpB from *Escherichia coli*.

E. coli ClpB sequence from UniProtKB - P63284 (CLPB_ECOLI) from *Escherichia coli* (strain K12). Residues which are part of Walker A of first NBD are highlighted in pink color box, Walker B of first NBD are highlighted in green color box, Walker A of second NBD are highlighted in red color box and Walker B of second NBD are highlighted in yellow color box. Pore loops residues of first NBD are highlighted in blue colored box and pore loops residues of second NBD are highlighted in magenta color box. Residues of Sensor 1 are underlined, and Sensor 2 residues are shown in italics and underlined. Arginine finger residue are shown in brown color and underlined.

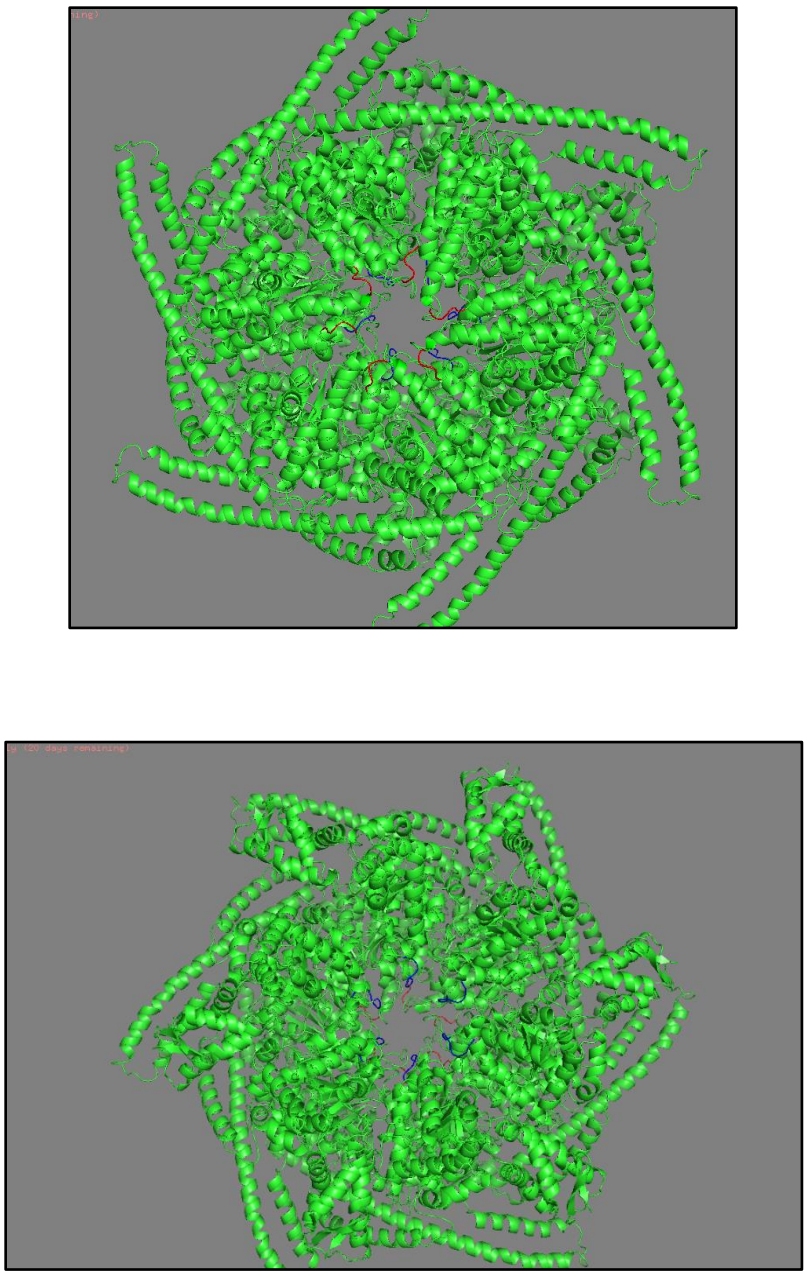


Figure 1.7. Pore loops arrangement in hexameric ClpB.

There are twelve (12) pore loops found in hexameric ClpB (two loops per monomer). The top panel shows six (6) loops in red color in the first NBD (view from top of the hexamer) and the bottom panel shows six (6) loops in blue color in the second NBD (view from bottom of the hexamer).

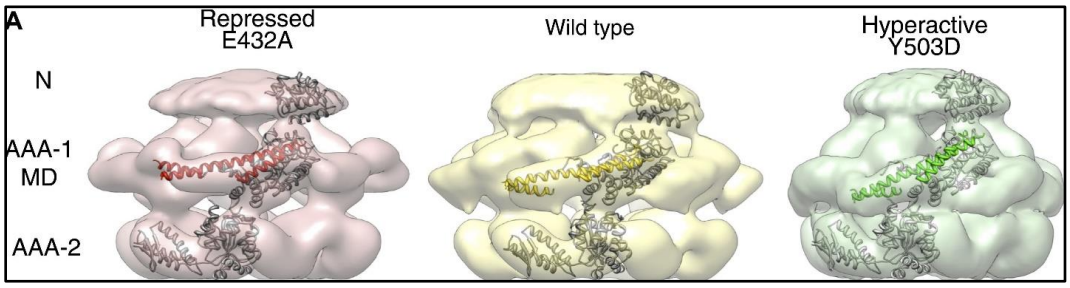


Figure 1.8. Middle domain arrangement in ClpB.

Repressed state of ClpB (left and middle), middle domain is in a horizontal arrangement shown in Red and Yellow color. Hyperactive state of ClpB (right) is shown in green color where the middle domain is slightly tilted. In this figure, N stands for N-terminal domain, AAA-1 stands for first NBD, AAA-2 stands for second NBD and MD stands for Middle Domain. The point mutations, E432A and Y503D were created in the middle domain of ClpB to stabilize the repressed and active conformation, respectively. Reprinted with permission. Original figure is taken from (Carroni, Kummer et al. 2014). eLife 2014;3:e02481 DOI: 10.7554/eLife.02481

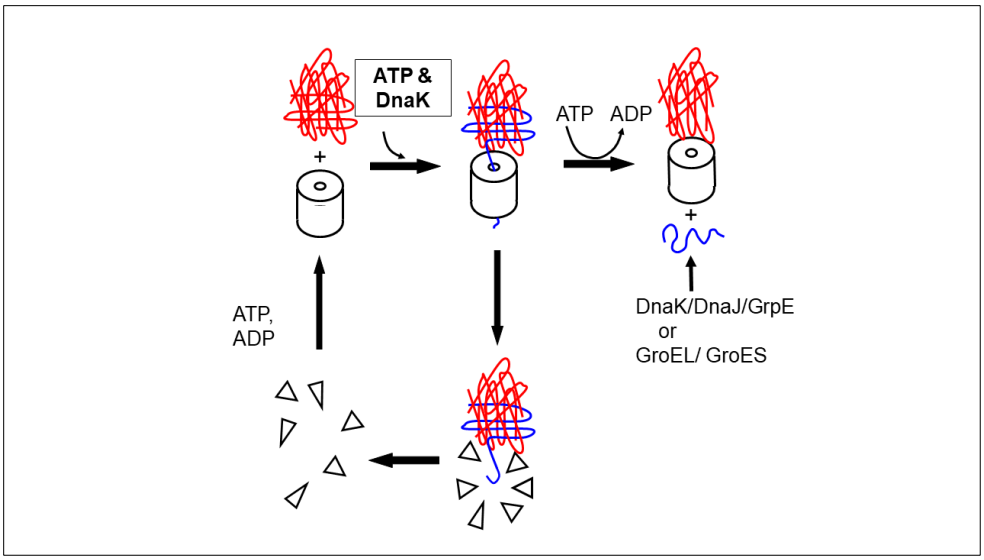


Figure 1.9. Proposed mechanism of protein disaggregation by ClpB.

Aggregated protein is shown in red and the hexameric ClpB unit is shown as a cylinder with an opening on top. Six ClpB monomers are assembled into a biologically active hexameric ClpB in the presence of nucleotides. DnaK activates ClpB and guides ClpB towards aggregates. When ClpB encounters an aggregate, ClpB is committed to pull a polypeptide loop or an exposed end of a polypeptide (blue) through it central channel and energy from ATP hydrolysis is required for this process. Extracted unfolded polypeptides (blue) are refolded either by Dnak/DnaJ/GrpE or GroEL/GroES chaperones. When ClpB is unable to extract a polypeptide completely from an aggregate, the hexamer dissociates into monomers which starts a new ClpB disaggregation cycle. Reprinted with permission. Original figure is taken from (Zolkiewski, Zhang et al. 2012) and presented here with a minor modification.

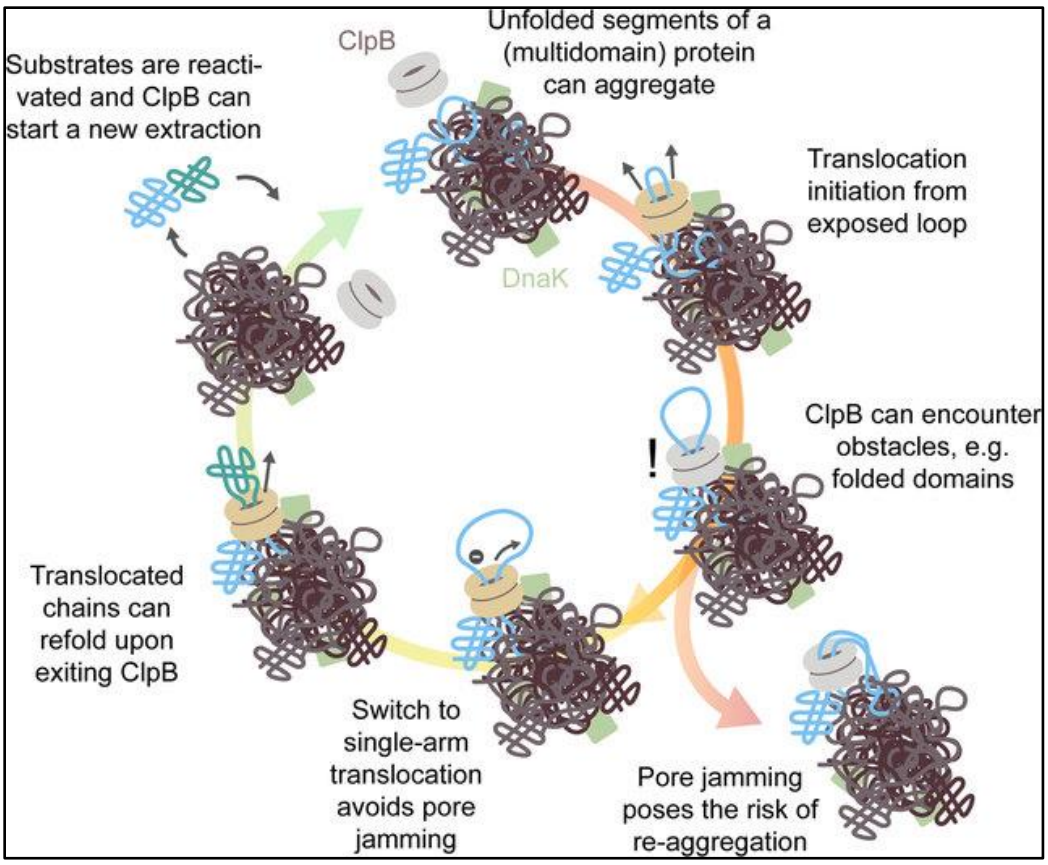
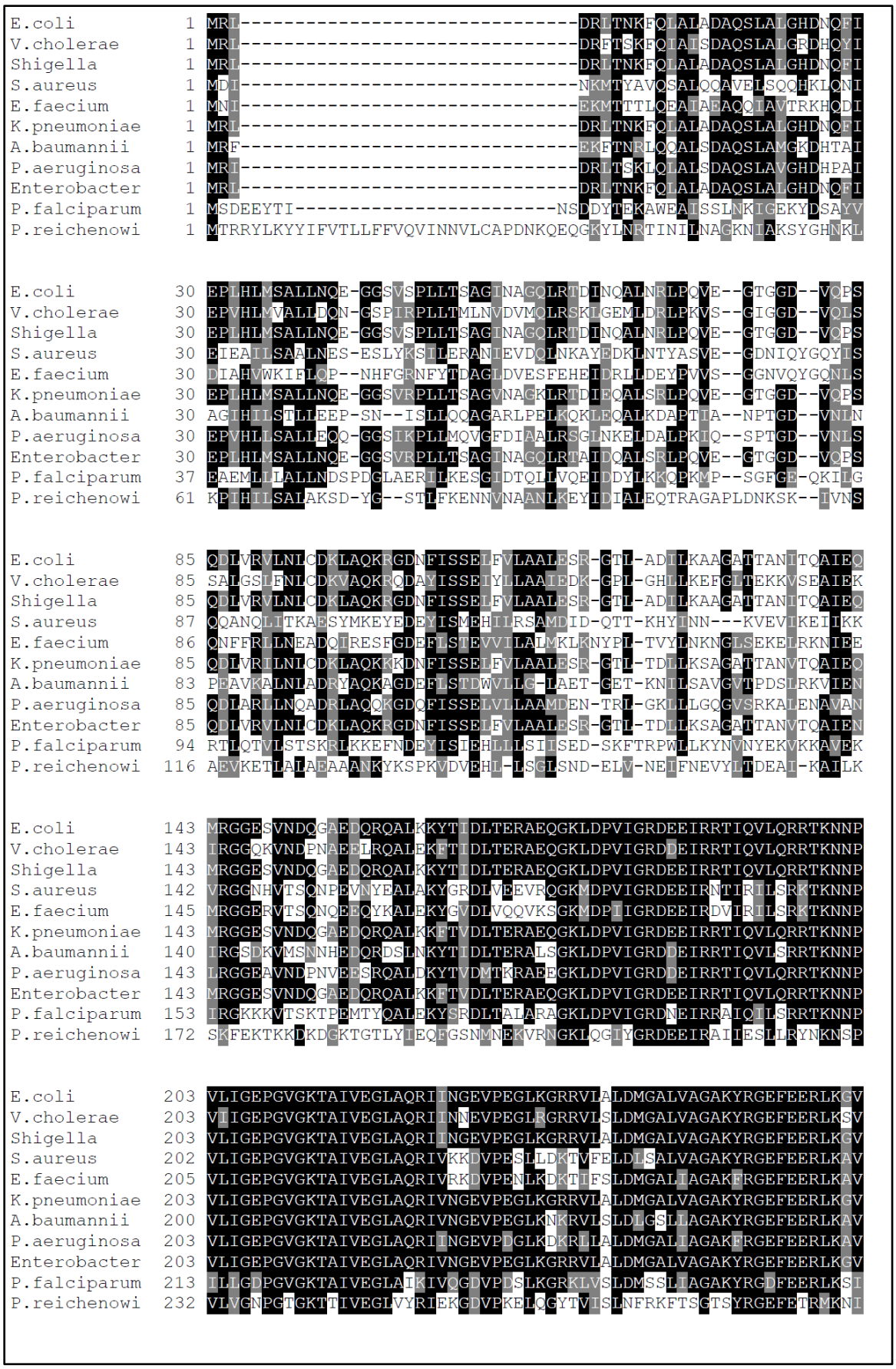
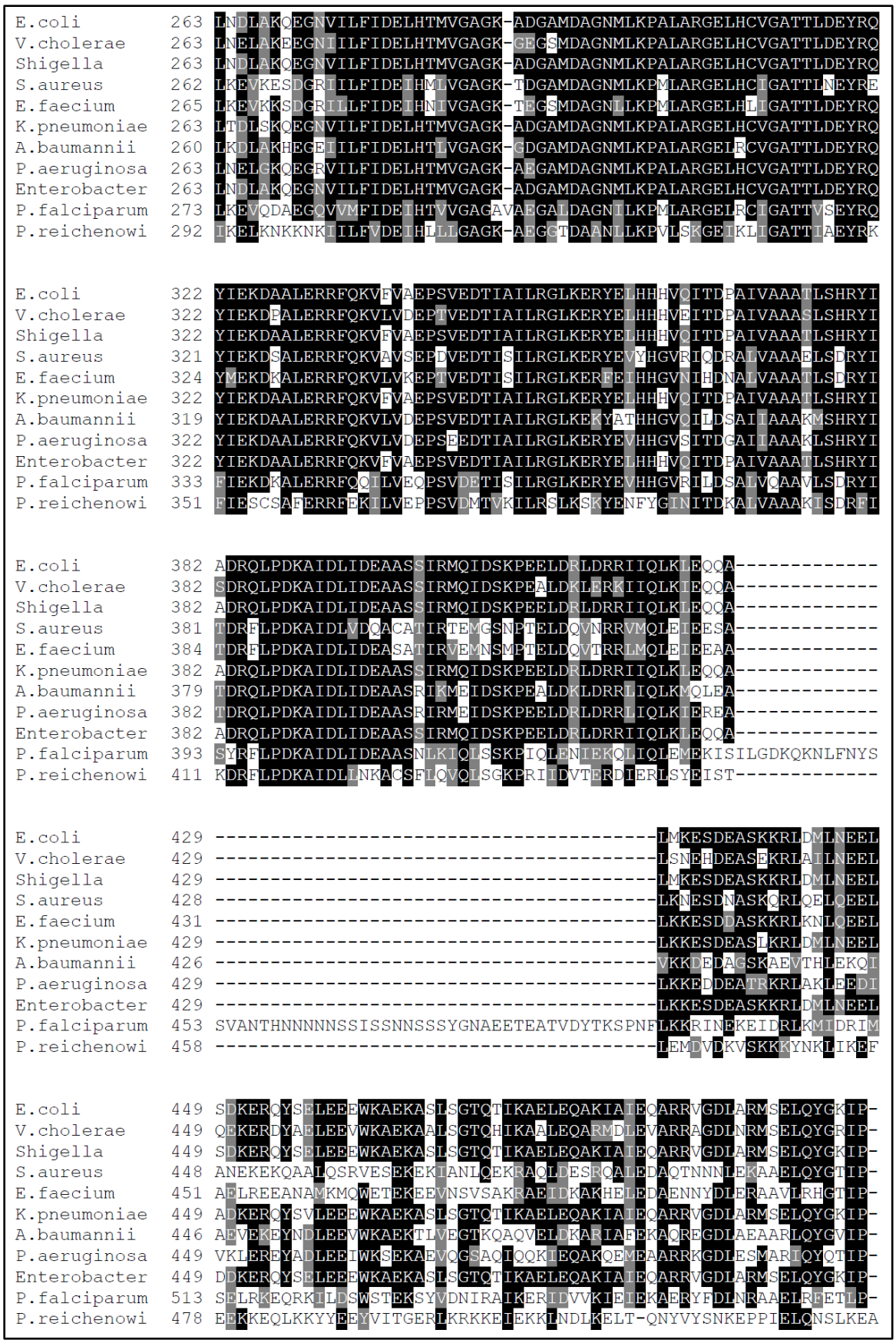
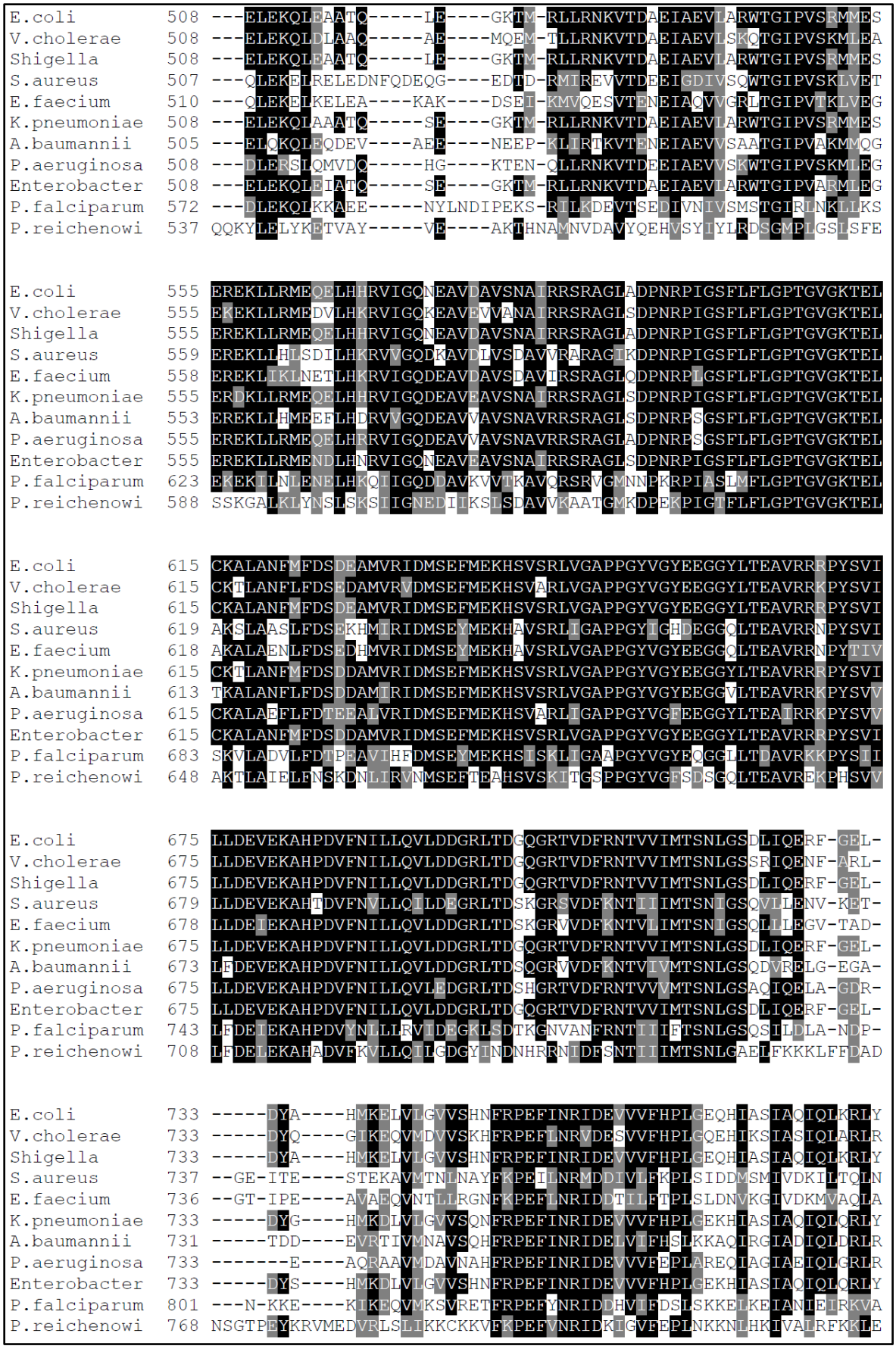


Figure 1.10. Proposed loop extrusion mechanism of ClpB.

Following a single ClpB molecule and its substrate in real time, Sander J.Tans and colleagues (Avellaneda, Franke et al. 2020) demonstrated that ClpB has the ability to pull exposed loops from the aggregates. When pulling these loops, ClpB can either pull both arms of the loops at the same time or shift to a single arm pulling if one arm is obstructed. Reprinted with permission. Original figure is taken from (Avellaneda, Franke et al. 2020).


E.coli 1 MRL----------DRLTNKFQLALADAQSLALGHDNQFI
V.cholerae 1 MRL----------DRFTSKFQIALSDAQSLALGRDHQYI
Shigella 1 MRL----------DRLTNKFQLALADAQSLALGHDNQFI
S.aureus 1 MDI----------NKMTYAVQSALQQAVELSQQHKLQNI
E.faecium 1 MNI----------EKMTTTLQEAIAEAQQIAVTRKHQDI
K.pneumoniae 1 MRL----------DRLTNKFQLALADAQSLALGHDNQFI
A.baumannii 1 MRF----------EKFTNRLQQALSDAQSLAMGKDHTAI
P.aeruginosa 1 MRI----------DRLTSKLQLALSDAQSLAVGHDHPAI
Enterobacter 1 MRL----------DRLTNKFQLALADAQSLALGHDNQFI
P.falciparum 1 MSDEEYTI----------NSDDYTEKAWEAISSLNKIGEKYDSAYV
P.reichenowi 1 MTRRYLKYYIFVTLLFFVQVINNVLCAPDNKQEQGKYLNRTINILNAGKNIAKSYGHNKL
E.coli 30 EPLHLMSALLNQE-GGSVSPLLTSAGINAGQLRTDINQALNRLPQVE--GTGGD--VQPS
V.cholerae 30 EPVHLMVALLDQN-GSPIRPLLTMLNVDVMQLRSKLGEMLDRLPKVS--GIGGD--VQLS
Shigella 30 EPLHLMSALLNQE-GGSVSPLLTSAGINAGQLRTDINQALNRLPQVE--GTGGD--VQPS
S.aureus 30 EIEAILSAALNES-ESLYKSILERANIEVDQLNKAYEDKLNTYASVE--GDNIQYGQYIS
E.faecium 30 DIAHVWKIFLQP--NHFGRNFYTDAGLDVESFEHEIDRLLDEYPVVS--GGNVQYGQNLS
K.pneumoniae 30 EPLHLMSALLNQE-GGSVRPLLTSAGVNAGKLRTDIEQALSRLPQVE--GTGGD--VQPS
A.baumannii 30 AGIHILSTLLEEP-SN--ISLLQQAGARLPELKQKLEQALKDAPTIA--NPTGD--VNIN
P.aeruginosa 30 EPVHLLSALLEQQ-GGSIKPLLMQVGFDIAALRSGINKELDALPKIQ--SPTGD--VNLS
Enterobacter 30 EPLHLMSALLNQE-GGSVRPLLTSAGINAGQLRTAIDQALSRLPQVE--GTGGD--VQPS
P.falciparum 37 EAEMLLLALLNDSPDGLAERILKESGIDTQLLVQEIDDYLKKQPKMP--SGFGE-QKILG
P.reichenowi 61 KPIHILSALAKSD-YG--STLFKENNVNAANLKEYIDIALEQTRAGAPLDNKSK--IVNS
E.coli 85 QDLVRVLNLCDKLAQKRGDNFISSELFVLAALESR-GTL-ADILKAAGATTANITQAIEQ
V.cholerae 85 SALGSLFNLCDKVAQKRQDAYISSEIYLLAAIEDK-GPL-GHLLKEFGLTEKKVSEAIEK
Shigella 85 QDLVRVLNLCDKLAQKRGDNFISSELFVLAALESR-GTL-ADILKAAGATTANITQAIEQ
S.aureus 87 QQANQLITKAESYMKEYEDEYISMEHILRSAMDID-QTT-KHYINN---KVEVIKEIIKK
E.faecium 86 QNFFRLLNEADQIRESFGDEFLSTEVVILALMKLKNYPL-TVYLNKNGLSEKELRKNIEE
K.pneumoniae 85 QDLVRILNLCDKLAQKKKDNFISSELFVLAALESR-GTL-TDLLKSAGATTANVTQAIEQ
A.baumannii 83 PEAVKALNLADRYAQKAGDEFLSTDWVLLG-LAET-GET-KNILSAVGVTPDSLRKVIEN
P.aeruginosa 85 QDLARLLNQADRLAQQKGDQFISSELVLLAAMDEN-TRI-GKLLLGQGVSRKALENAVAN
Enterobacter 85 QDLVRVLNLCDKLAQKRGDNFISSELFVLAALESR-GTL-TDLLKSAGATTANVTQAIEN
P.falciparum 94 RTLQTVLSTSKRLKKEFNDEYISIEHLLLSIISED-SKFTRPWLLKYNVNYEKVKKAVEK
P.reichenowi 116 AEVKETLALAEAAANKYKSPKVDVEHL-LSGLSND-ELV-NEIFNEVYLTDEAI-KAILK
E.coli 143 MRGGESVNDQGAEDQRQALKKYTIDLTERAEQGKLDPVIGRDEEIRRTIQVLQRRTKNNP
V.cholerae 143 IRGGQKVNDPNAEELRQALEKFTIDLTERAEQGKLDPVIGRDDEIRRTIQVLQRRTKNNP
Shigella 143 MRGGESVNDQGAEDQRQALKKYTIDLTERAEQGKLDPVIGRDEEIRRTIQVLQRRTKNNP
S.aureus 142 VRGGNHVTSQNPEVNYEALAKYGRDLVEEVRQGKMDPVIGRDEEIRNTIRILSRKTKNNP
E.faecium 145 MRGGERVTSQNQEEQYKALEKYGVDLVQQVKSGKMDPIIGRDEEIRDVIRILSRKTKNNP
K.pneumoniae 143 MRGGESVNDQGAEDQRQALKKFTVDLTERAEQGKLDPVIGRDEEIRRTIQVLQRRTKNNP
A.baumannii 140 IRGSDKVMSNNHEDQRDSLNKYTIDLTERALSGKLDPVIGRDDEIRRTIQVLSRRTKNNP
P.aeruginosa 143 LRGGEAVNDPNVEESRQALDKYTVDMTKRAEEGKLDPVIGRDDEIRRTIQVLQRRTKNNP
Enterobacter 143 MRGGESVNDQGAEDQRQALKKFTVDLTERAEQGKLDPVIGRDEEIRRTIQVLQRRTKNNP
P.falciparum 153 IRGKKKVTSKTPEMTYQALEKYSRDLTALARAGKLDPVIGRDNEIRRAIQILSRRTKNNP
P.reichenowi 172 SKFEKTKKDKDGKTGTLYIEQFGSNMNEKVRNGKLQGIYGRDEEIRAIIESLLRYNKNSP
E.coli 203 VLIGEPGVGKTAIVEGLAQRIINGEVPEGLKGRRVLALDMGALVAGAKYRGEFEERLKGV
V.cholerae 203 VIIGEPGVGKTAIVEGLAQRIINNEVPEGLRGRRVLSLDMGALVAGAKYRGEFEERLKSV
Shigella 203 VLIGEPGVGKTAIVEGLAQRIINGEVPEGLKGRRVLALDMGALVAGAKYRGEFEERLKGV
S.aureus 202 VLIGEPGVGKTAIVEGLAQRIVKKDVPESLLDKTVFELDLSALVAGAKYRGEFEERLKAV
E.faecium 205 VLIGEPGVGKTAIVEGLAQRIVRKDVPENLKDKTIFSLDMGALIAGAKFRGEFEERLKAV
K.pneumoniae 203 VLIGEPGVGKTAIVEGLAQRIVNGEVPEGLKGRRVLALDMGALVAGAKYRGEFEERLKGV
A.baumannii 200 VLIGEPGVGKTAIVEGLAQRIVNGEVPEGLKNKRVLSLDLGSLIAGAKYRGEFEERLKAV
P.aeruginosa 203 VLIGEPGVGKTAIVEGLAQRIINGEVPDGLKDKRLIALDMGALIAGAKFRGEFEERLKAV
Enterobacter 203 VLIGEPGVGKTAIVEGLAQRIVNGEVPEGLKGRRVLALDMGALVAGAKYRGEFEERLKGV
P.falciparum 213 ILLGDPGVGKTAIVEGLAIKIVQGDVPDSLKGRKLVSLDMSSLIAGAKYRGDFEERLKSI
P.reichenowi 232 VLVGNPGTGKTTIVEGLVYRIEKGDVPKELQGYTVISLNFRKFTSGTSYRGEFETRMKNI


E.coli 263 LNDLAKQEGNVILFIDELHTMVGAGK-ADGAMDAGNMLKPALARGELHCVGATTLDEYRQ
V.cholerae 263 LNELAKEEGNIILFIDELHTMVGAGK-GEGSMDAGNMLKPALARGELHCVGATTLDEYRQ
Shigella 263 LNDLAKQEGNVILFIDELHTMVGAGK-ADGAMDAGNMLKPALARGELHCVGATTLDEYRQ
S.aureus 262 LKEVKESDGRIILFIDEIHMLVGAGK-TDGAMDAGNMLKPMLARGELHCIGATTLNEYRE
E.faecium 265 LKEVKKSDGRILLFIDEIHNIVGAGK-TEGSMDAGNLLKPMLARGELHLIGATTLDEYRQ
K.pneumoniae 263 LTDLSKQEGNVILFIDELHTMVGAGK-ADGAMDAGNMLKPALARGELHCVGATTLDEYRQ
A.baumannii 260 LKDLAKHEGEIILFIDELHTLVGAGK-GDGAMDAGNMLKPALARGELRCVGATTLDEYRQ
P.aeruginosa 263 LNELGKQEGRVILFIDELHTMVGAGK-AEGAMDAGNMLKPALARGELHCVGATTLDEYRQ
Enterobacter 263 LNDLAKQEGNVILFIDELHTMVGAGK-ADGAMDAGNMLKPALARGELHCVGATTLDEYRQ
P.falciparum 273 LKEVQDAEGQVVMFIDEIHTVVGAGAVAEGAIDAGNILKPMLARGELRCIGATTVSEYRQ
P.reichenowi 292 IKELKNKKNKIILFVDEIHLLLGAGK-AEGGTDAANILKPVLSKGEIKLIGATTIAEYRK
E.coli 322 YIEKDAALERRFQKVFVAEPSVEDTIAILRGLKERYELHHHVQITDPAIVAAATLSHRYI
V.cholerae 322 YIEKDPALERRFQKVLVDEPTVEDTIAILRGLKERYELHHHVEITDPAIVAAASLSHRYI
Shigella 322 YIEKDAALERRFQKVFVAEPSVEDTIAILRGLKERYELHHHVQITDPAIVAAATLSHRYI
S.aureus 321 YIEKDSALERRFQKVAVSEPDVEDTISILRGLKERYEVYHGVRIQDRALVAAAELSDRYI
E.faecium 324 YMEKDKALERRFQKVLVKEPTVEDTISILRGLKERFEIHHGVNIHDNALVAAATLSDRYI
K.pneumoniae 322 YIEKDAALERRFQKVFVAEPSVEDTIAILRGLKERYELHHHVQITDPAIVAAATLSHRYI
A.baumannii 319 YIEKDAALERRFQKVLVDEPSVEDTIAILRGLKEKYATHHGVQILDSAIIAAAKMSHRYI
P.aeruginosa 322 YIEKDAALERRFQKVLVDEPSEEDTIAILRGLKERYEVHHGVSITDGAIIAAAKLSHRYI
Enterobacter 322 YIEKDAALERRFQKVFVAEPSVEDTIAILRGLKERYELHHHVQITDPAIVAAATLSHRYI
P.falciparum 333 FIEKDKALERRFQQILVEQPSVDETISILRGLKERYEVHHGVRILDSALVQAAVLSDRYI
P.reichenowi 351 FIESCSAEERRFEKILVEPPSVDMTVKILRSLKSKYENFYGINITDKALVAAAKISDRFI
E.coli 382 ADRQLPDKAIDLIDEAASSIRMQIDSKPEELDRLDRRIIQLKLEQQA----------
V.cholerae 382 SDRQLPDKAIDLIDEAASSIRMQIDSKPEALDKLERKIIQLKIEQQA----------
Shigella 382 ADRQLPDKAIDLIDEAASSIRMQIDSKPEELDRLDRRIIQLKLEQQA----------
S.aureus 381 TDRFLPDKAIDLVDQACATIRTEMGSNPTELDQVNRRVMQLEIEESA----------
E.faecium 384 TDRFLPDKAIDLIDEASATIRVEMNSMPTELDQVTRRLMQLEIEEAA----------
K.pneumoniae 382 ADRQLPDKAIDLIDEAASSIRMQIDSKPEELDRLDRRIIQLKLEQQA----------
A.baumannii 379 TDRQLPDKAIDLIDEAASRIKMEIDSKPEALDKLDRRLIQLKMQLEA----------
P.aeruginosa 382 TDRQLPDKAIDLIDEAASRIRMEIDSKPEELDRLDRRLIQLKIEREA----------
Enterobacter 382 ADRQLPDKAIDLIDEAASSIRMQIDSKPEELDRLDRRIIQLKLEQQA----------
P.falciparum 393 SYRFLPDKAIDLIDEAASNLKIQLSSKPIQLENIEKQLIQLEMEKISILGDKQKNLFNYS
P.reichenowi 411 KDRFLPDKAIDLLNKACSFLQVQLSGKPRIIDVTERDIERLSYEIST----------
E.coli 429 ----------LMKESDEASKKRLDMLNEEL
V.cholerae 429 ----------LSNEHDEASEKRLAILNEEL
Shigella 429 ----------LMKESDEASKKRLDMLNEEL
S.aureus 428 ----------LKNESDNASKQRLQELQEEL
E.faecium 431 ----------LKKESDDASKKRLKNLQEEL
K.pneumoniae 429 ----------LKKESDEASLKRLDMLNEEL
A.baumannii 426 ----------VKKDEDAGSKAEVTHLEKQI
P.aeruginosa 429 ----------LKKEDDEATRKRLAKLEEDI
Enterobacter 429 ----------LKKESDEASKKRLDMLNEEL
P.falciparum 453 SVANTHNNNNNSSISSNNSSSYGNAEETEATVDYTKSPNFLKKRINEKEIDRLKMIDRIM
P.reichenowi 458 ----------LEMDVDKVSKKKYNKLIKEF
E.coli 449 SDKERQYSELEEEWKAEKASLSGTQTIKAELEQAKIAIEQARRVGDLARMSELQYGKIP-
V.cholerae 449 QDKERDYAILEEVWKAEKAALSGTQHIKAALEQARMDLEVARRAGDLNRMSELQYGRIP-
Shigella 449 SDKERQYSELEEEWKAEKASLSGTQTIKAELEQAKIAIEQARRVGDLARMSELQYGKIP-
S.aureus 448 ANEKEKQAALQSRVESEKEKIANLQEKRAQLLESFQALEDAQTNNNLEKAAELQYGTIP-
E.faecium 451 AELREEANAMKMQWETEKEEVNSVSAKRAEIDKAKHELEDAENNYDLERAAVLRHGTIP-
K.pneumoniae 449 ADKERQYSVLEEEWKAEKASLSGTQTIKAELEQAKIAIEQARRVGDLARMSELQYGKIP-
A.baumannii 446 AEVEKEYNDLEEVWKAEKTLVEGTKQAQVELLKARIAFEKAQREGDLAEAARLQYGVIP-
P.aeruginosa 449 VKLEREYADLEEIWKSEKAEVQGSAQIQQKIEQAKQEMEAARRKGDLESMARIQYQTIP-
Enterobacter 449 DDKERQYSELEEEWKAEKASLSGTQTIKAELEQAKIAIEQARRVGDLARMSELQYGKIP-
P.falciparum 513 SELRKEQRKILDSWSTEKSYVDNIRAIKERIDVVKIEIEKAERYFDLNRAAELRFETLP-
P.reichenowi 478 EEKKEQLKKYYEEYVITGERLKRKKEIEKKINDLKELT-QNYVYSNKEPPIELQNSLKEA


E.coli 508 ---ELEKQLEAATQ-----LE-----GKTM-RLLRNKVTDAEIAEVLARWTGIPVSRMMES
V.cholerae 508 ---ELEKQLDLAAQ-----AE-----MQEM-TLLRNKVTDAEIAEVLSKQTGIPVSKMLEA
Shigella 508 ---ELEKQLEAATQ-----LE-----GKTM-RLLRNKVTDAEIAEVLARWTGIPVSRMMES
S.aureus 507 ---QLEKELRELEDNFQDEQG----EDTD-RMIREVVTDEEIGDIVSQWTGIPVSKLVET
E.faecium 510 ---QLEKELKELEA----KAK----DSEI-KMVQESVTENEIAQVVGRLTGIPVTKLVEG
K.pneumoniae 508 ---ELEKQLAAATQ-----SE-----GKTM-RLLRNKVTDAEIAEVLARWTGIPVSRMMES
A.baumannii 505 ---ELQKQLEQDEV----AEE-----NEEP-KLIRTKVTENEIAEVVSAATGIPVAKMMQG
P.aeruginosa 508 ---DLERSLQMVDQ-----HG-----KTEN-QLLRNKVTDEEIAEVVSKWTGIPVSKMLEG
Enterobacter 508 ---ELEKQLEIATQ-----SE-----GKTM-RLLRNKVTDAEIAEVLARWTGIPVARMLEG
P.falciparum 572 ---DLEKQLKKAEE-----NYLNDIPEKS-RILKDEVTSEDIVNIVSMSTGIRLNKLLKS
P.reichenowi 537 QQKYLELYKETVAY-----VE-----AKTHNAMNVDAVYQEHVSYIYLRDSGMPLGSLSFE
E.coli 555 EREKLLRMEQELHHRVIGQNEAVDAVSNAIRRSRAGLADPNRPIGSFLFLGPTGVGKTEL
V.cholerae 555 EKEKLLRMEDVLHKRVIGQKEAVEVVANAIRRSRAGLSDPNRPIGSFLFLGPTGVGKTEL
Shigella 555 EREKLLRMEQELHHRVIGQNEAVDAVSNAIRRSRAGLADPNRPIGSFLFLGPTGVGKTEL
S.aureus 559 EREKLLHLSDILHKRVVGQDKAVDLVSDAVVRARAGIKDPNRPIGSFLFLGPTGVGKTEL
E.faecium 558 EREKLIKLNETLHKRVIGQDEAVDAVSDAVIRSRAGLQDPNRPLGSFLFLGPTGVGKTEL
K.pneumoniae 555 ERDKLLRMEQELHHRVIGQDEAVEAVSNAIRRSRAGLSDPNRPIGSFLFLGPTGVGKTEL
A.baumannii 553 EREKLLHMEEFLHDRVVGQDEAVVAVSNAVRRSRAGLSDPNRPSGSFLFLGPTGVGKTEL
P.aeruginosa 555 EREKLLRMEQELHRRVIGQDEAVVAVSNAVRRSRAGLADPNRPSGSFLFLGPTGVGKTEL
Enterobacter 555 EREKLLRMENDLHNRVIGQNEAVEAVSNAIRRSRAGLSDPNRPIGSFLFLGPTGVGKTEL
P.falciparum 623 EKEKILNLENELHKQIIGQDDAVKVVTKAVQRSRVGMNNPKRPIASLMFLGPTGVGKTEL
P.reichenowi 588 SSKGALKLYNSLSKSIIGNEDIIKSLSDAVVKAATGMKDPEKPIGTFLFLGPTGVGKTEL
E.coli 615 CKALANFMFDSDEAMVRIDMSEFMEKHSVSRLVGAPPGYVGYEEGGYLTEAVRRRPYSVI
V.cholerae 615 CKTLANFLFDSEDAMVRVDMSEFMEKHSVARLVGAPPGYVGYEEGGYLTEAVRRKPYSVI
Shigella 615 CKALANFMFDSDEAMVRIDMSEFMEKHSVSRLVGAPPGYVGYEEGGYLTEAVRRRPYSVI
S.aureus 619 AKSLAASLFDSEKHMIRIDMSEYMEKHAVSRLIGAPPGYIGHDEGGQLTEAVRRNPYSVI
E.faecium 618 AKALAENLFDSEDHMVRIDMSEYMEKHAVSRLVGAPPGYVGYEEGGQLTEAVRRNPYTIV
K.pneumoniae 615 CKTLANFMFDSDDAMVRIDMSEFMEKHSVSRLVGAPPGYVGYEEGGYLTEAVRRRPYSVI
A.baumannii 613 TKALANFLFDSDDAMIRIDMSEFMEKHSVSRLVGAPPGYVGYEEGGVLTEAVRRKPYSVV
P.aeruginosa 615 CKALAEFLFDTEEALVRIDMSEFMEKHSVARLIGAPPGYVGFEEGGYLTEAIRRKPYSVV
Enterobacter 615 CKALANFMFDSDDAMVRIDMSEFMEKHSVSRLVGAPPGYVGYEEGGYLTEAVRRRPYSVI
P.falciparum 683 SKVLADVLFDTPEAVIHFDMSEYMEKHSISKLIGAAPGYVGYEQGGLLTDAVRKKPYSII
P.reichenowi 648 AKTLAIELFNSKDNLIRVNMSEFTEAHSVSKITGSPPGYVGFSDSGQLTEAVREKPHSVV
E.coli 675 LLDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFRNTVVIMTSNLGSDLIQERF-GEL-
V.cholerae 675 LLDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFRNTVVIMTSNLGSSRIQENF-ARL-
Shigella 675 LLDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFRNTVVIMTSNLGSDLIQERF-GEL-
S.aureus 679 LLDEVEKAHTDVFNVLLQILDEGRLTDSKGRSVDFKNTIIIMTSNIGSQVLLENV-KET-
E.faecium 678 LLDEIEKAHPDVFNILLQVLDDGRLTDSKGRVVDFKNTVLIMTSNIGSQLLLEGV-TAD-
K.pneumoniae 675 LLDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFRNTVVIMTSNLGSDLIQERF-GEL-
A.baumannii 673 LEDEVEKAHPDVFNILLQVLDDGRLTDSQGRVVDFKNTVIVMTSNLGSQDVRELG-EGA-
P.aeruginosa 675 LLDEVEKAHPDVFNILLQVLEDGRLTDSHGRTVDFRNTVVVMTSNLGSAQIQELA-GDR-
Enterobacter 675 LLDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFRNTVVIMTSNLGSDLIQERF-GEL-
P.falciparum 743 LEDEIEKAHPDVYNLLLRVIDEGKLSDTKGNVANFRNTIIIFTSNLGSQSILDLA-NDP-
P.reichenowi 708 LEDELEKAHADVFKVLLQILGDGYINDNHRRNIDFSNTIIIMTSNLGAELFKKKLFFDAD
E.coli 733 -----DYA----HMKELVLGVVSHNFRPEFINRIDEVVVFHPLGEQHIASIAQIQLKRLY
V.cholerae 733 -----DYQ----GIKEQVMDVVSKHFRPEFLNRVDESVVFHPLGQEHIKSIASIQLARLR
Shigella 733 -----DYA----HMKELVLGVVSHNFRPEFINRIDEVVVFHPLGEQHIASIAQIQLKRLY
S.aureus 737 --GE-ITE----STEKAVMTNLNAYFKPEILNRMDDIVLFKPLSIDDMSMIVDKILTQLN
E.faecium 736 --GT-IPE----AVAEQVNTLLRGNFKPEFLNRIDDTILFTPLSLDNVKGIVDKMVAQLA
K.pneumoniae 733 -----DYG----HMKDLVLGVVSQNFRPEFINRIDEVVVFHPLGEKHIASIAQIQLQRLY
A.baumannii 731 -----TDD----EVRTIVMNAVSQHFRPEFINRIDELVIFHSLKKAQIRGIADIQLDRLR
P.aeruginosa 733 -------E----AQRAAVMDAVNAHFRPEFINRIDEVVVFEPLAREQIAGIAEIQLGRLR
Enterobacter 733 -----DYS----HMKDLVLGVVSHNFRPEFINRIDEVVVFHPLGEKHIASIAQIQLQRLY
P.falciparum 801 ---N-KKE----KIKEQVMKSVRETFRPEFYNRIDDHVIFDSLSKKELKEIANIEIRKVA
P.reichenowi 768 NSGTPIYKRVMEDVRLSLIKKCKKVFKPEFVNRIDKIGVFEPLNKKNLHKIVALRFKKLE

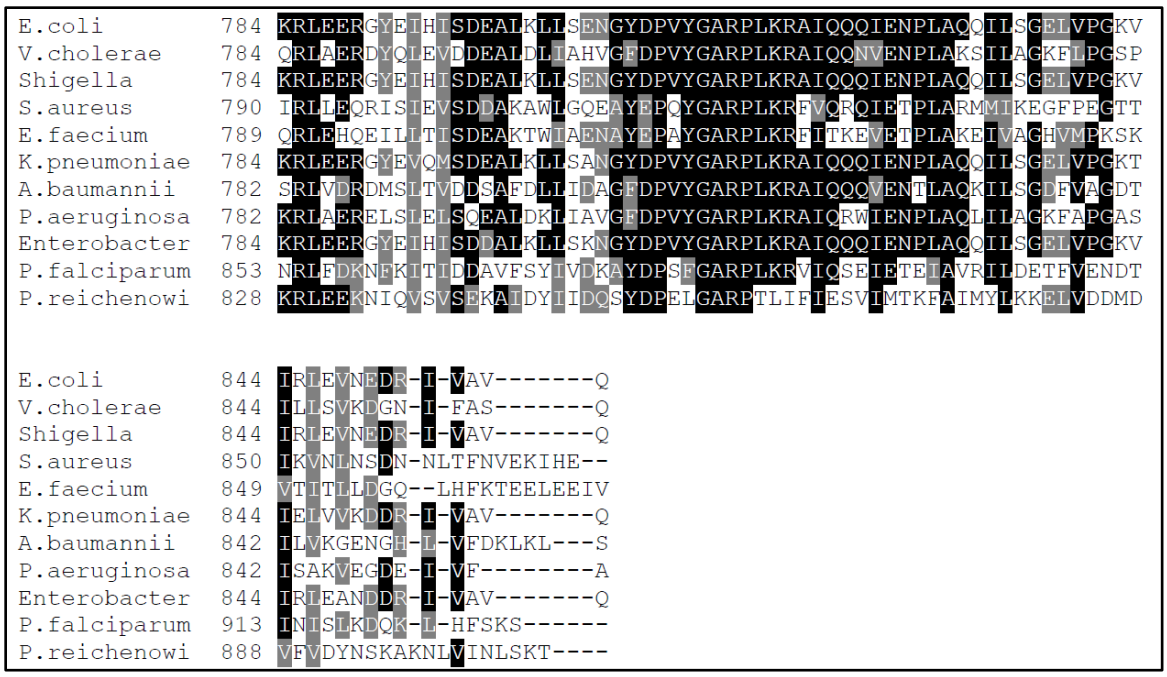


Figure 1.11. Sequence alignment of Hsp100 proteins from several human pathogens.

ClpB sequence from different pathogens have been compared using Clustal Omega tool provided by Bioinformatics Institute (EMBL-EBI, https://www.ebi.ac.uk/Tools/msa/clustalo/), followed by BOXSHADE 3.21 tool, from the Swiss Institute of Bioinformatics (https://embnet.vital-it.ch/software/BOX_form.html) which shaded multiple alignments. Black shaded boxes indicate conserved residues, grey indicates similar amino acids and white indicates different amino acids. (-) indicates the gaps introduce by Clustal Omega program when aligning these sequences.

E. coli- sp|P63284|CLPB_ECOLI, Escherichia coli (strain K12)

V. cholerae- sp|Q9KU18|CLPB_VIBCH, Vibrio cholerae serotype O1 (strain ATCC 39315 / El Tor Inaba N16961)

Shigella- tr|A0A656NWG3|A0A656NWG3_9ENTR, Shigella sp.

S. aureus- sp|Q7A6G6|CLPB_STAAN,Staphylococcus aureus (strain N315)

E. faecium- tr|A0A132PA02|A0A132PA02_ENTFC, Enterococcus faecium

K. pneumoniae- tr|W9B407|W9B407_KLEPN, Klebsiella pneumoniae

A. baumannii- tr|V5VGK1|V5VGK1_ACIBA, Acinetobacter baumannii

P. aeruginosa- sp|Q9HVN5|CLPB_PSEAE, Pseudomonas aeruginosa (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1)

Enterobacter- tr|A0A3S0FSR9|A0A3S0FSR9_9ENTR, Enterobacter sp. WCHEn090040

P. falciparum - tr|Q8IB03|Q8IB03_PLAF7, ClpB1, Plasmodium falciparum (isolate 3D7)

P. reichenowi - tr|A0A060RUC0|A0A060RUC0_PLARE, ClpB2, Plasmodium reichenowi

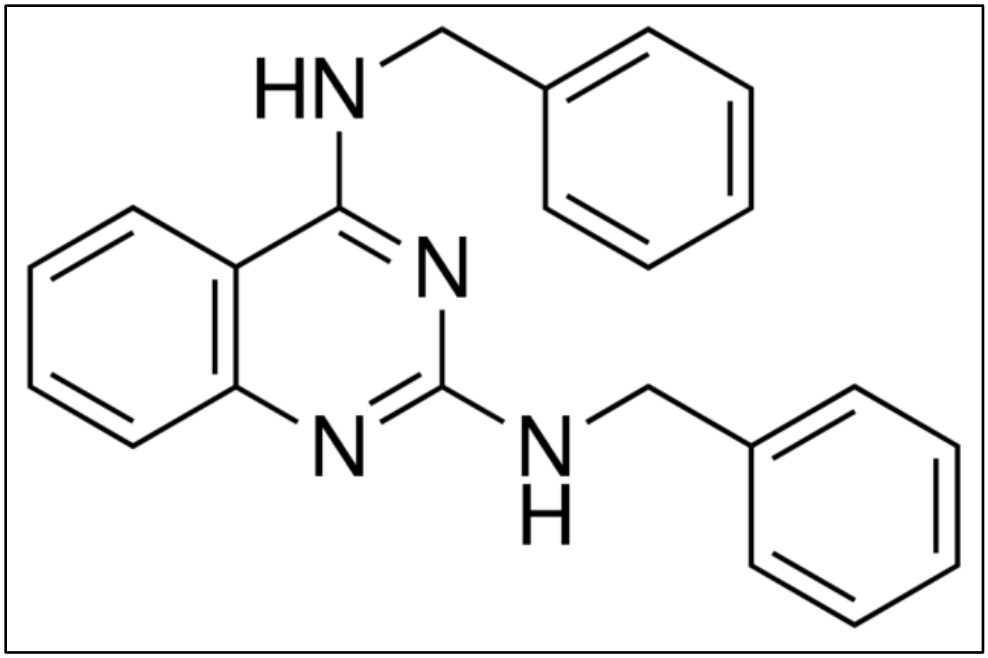


Figure 1.12. Structure of DBeQ.

N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ) structure is shown here. Based on the structures supplied by the manufacturer and redrawn using ChemDraw 12. DBeQ was initially described as inhibitor (an antitumor target) for the human AAA+ ATPase p97. DBeQ is a potent, reversible, inhibitor of p97 with IC_{50} of 1.5 μM (Chou, Brown et al. 2011). p97 (also known as Valosin-containing protein (VCP), Cdc48 in yeast and plants, CDC-48 in worms, Ter94 in flies) is a hexameric AAA+ protein. p97 is mainly involve in cellular protein degradation machinery and is not found in bacteria (Glaza, Ranaweera et al. 2021)

References.

Acebrón, S. P., I. Martín, U. del Castillo, F. Moro and A. Muga (2009). "DnaK-mediated association of ClpB to protein aggregates. A bichaperone network at the aggregate surface." FEBS Letters **583**(18): 2991-2996.

AhYoung, A. P., A. Koehl, D. Cascio and P. F. Egea (2015). "Structural mapping of the ClpB ATPases of *Plasmodium falciparum*: Targeting protein folding and secretion for antimalarial drug design." Protein science : a publication of the Protein Society **24**(9): 1508-1520.

Akoev, V., E. P. Gogol, M. E. Barnett and M. Zolkiewski (2004). "Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB." Protein Science **13**(3): 567-574.

Alksne, L. E. and P. M. Dunman (2008). "Target-based antimicrobial drug discovery." Methods Mol Biol **431**: 271-283.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(4096): 223-230.

Ashraf, G. M., N. H. Greig, T. A. Khan, I. Hassan, S. Tabrez, S. Shakil, I. A. Sheikh, S. K. Zaidi, M. Akram, N. R. Jabir, C. K. Firoz, A. Naeem, I. M. Alhazza, G. A. Damanhouri and M. A. Kamal (2014). "Protein misfolding and aggregation in Alzheimer's disease and type 2 diabetes mellitus." CNS & neurological disorders drug targets **13**(7): 1280-1293.

Avellaneda, M. J., K. B. Franke, V. Sunderlikova, B. Bukau, A. Mogk and S. J. Tans (2020). "Processive extrusion of polypeptide loops by a Hsp100 disaggregase." Nature **578**(7794): 317-320.

Barnett, M. E., A. Zolkiewska and M. Zolkiewski (2000). "Structure and activity of ClpB from *Escherichia coli*. Role of the amino-and -carboxyl-terminal domains." The Journal of biological chemistry **275**(48): 37565-37571.

Bukau, B., J. Weissman and A. Horwich (2006). "Molecular chaperones and protein quality control." Cell **125**(3): 443-451.

Cabrera, M., S. Boronat, L. Marte, M. Vega, P. Pérez, J. Ayté and E. Hidalgo (2020). "Chaperone-Facilitated Aggregation of Thermo-Sensitive Proteins Shields Them from Degradation during Heat Stress." Cell Reports **30**(7): 2430-2443.e2434.

Calloni, G., T. Chen, Sonya M. Schermann, H.-c. Chang, P. Genevaux, F. Agostini, Gian G. Tartaglia, M. Hayer-Hartl and F. U. Hartl (2012). "DnaK Functions as a Central Hub in the *E. coli* Chaperone Network." Cell Reports **1**(3): 251-264.

Camilloni, C., D. Bonetti, A. Morrone, R. Giri, C. M. Dobson, M. Brunori, S. Gianni and M. Vendruscolo (2016). "Towards a structural biology of the hydrophobic effect in protein folding." Scientific Reports **6**(1): 28285.

Carroni, M., E. Kummer, Y. Oguchi, P. Wendler, D. K. Clare, I. Sinning, J. Kopp, A. Mogk, B. Bukau and H. R. Saibil (2014). "Head-to-tail interactions of the coiled-coil domains regulate ClpB activity and cooperation with Hsp70 in protein disaggregation." eLife **3**: e02481.

Chow, I.-T., M. E. Barnett, M. Zolkiewski and F. Baneyx (2005). "The N-terminal domain of *Escherichia coli* ClpB enhances chaperone function." FEBS Letters **579**(20): 4242-4248.

Dobson, C. M., A. Šali and M. Karplus (1998). "Protein Folding: A Perspective from Theory and Experiment." Angewandte Chemie International Edition **37**(7): 868-893.

Doyle, S. M., O. Genest and S. Wickner (2013). "Protein rescue from aggregates by powerful molecular chaperone machines." Nature Reviews Molecular Cell Biology **14**(10): 617-629.

Doyle, S. M., J. R. Hoskins and S. Wickner (2012). "DnaK chaperone-dependent disaggregation by caseinolytic peptidase B (ClpB) mutants reveals functional overlap in the N-terminal domain and nucleotide-binding domain-1 pore tyrosine." J Biol Chem **287**(34): 28470-28479.

Forloni, G., L. Terreni, I. Bertani, S. Fogliarino, R. Invernizzi, A. Assini, G. Ribizzi, A. Negro, E. Calabrese, M. A. Volonté, C. Mariani, M. Franceschi, M. Tabaton and A. Bertoli (2002). "Protein misfolding in Alzheimer's and Parkinson's disease: genetics and molecular mechanisms." Neurobiology of Aging **23**(5): 957-976.

Gates, S. N. and A. Martin (2020). "Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation." Protein Science **29**(2): 407-419.

Gianni, S., N. R. Guydosh, F. Khan, T. D. Caldas, U. Mayor, G. W. N. White, M. L. DeMarco, V. Daggett and A. R. Fersht (2003). "Unifying features in protein-folding mechanisms." Proceedings of the National Academy of Sciences **100**(23): 13286-13291.

Glaza, P., C. B. Ranaweera, S. Shiva, A. Roy, B. V. Geisbrecht, F. J. Schoenen and M. Zolkiewski (2020). "Repurposing p97 inhibitors for chemical modulation of the bacterial ClpB/DnaK bi-chaperone system." J Biol Chem.

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

Harrison, C. (2003). "GrpE, a nucleotide exchange factor for DnaK." Cell stress & chaperones **8**(3): 218-224.

Hartl, F. U. (2017). "Protein Misfolding Diseases." Annual Review of Biochemistry **86**(1): 21-26.

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

Kampinga, H. H. and E. A. Craig (2010). "The HSP70 chaperone machinery: J proteins as drivers of functional specificity." Nature Reviews Molecular Cell Biology **11**(8): 579-592.

Karlin, S., J. Mrázek, J. Ma and L. Brocchieri (2005). "Predicted highly expressed genes in archaeal genomes." Proceedings of the National Academy of Sciences of the United States of America **102**(20): 7303-7308.

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

Kim, K. I., G.-W. Cheong, S.-C. Park, J.-S. Ha, K. M. Woo, S. J. Choi and C. H. Chung (2000). "Heptameric ring structure of the heat-shock protein ClpB, a protein-activated ATPase in *Escherichia coli* Edited by W. Baumeister." Journal of Molecular Biology **303**(5): 655-666.

Kuczynska-Wisnik, D., C. Cheng, R. R. Ganta and M. Zolkiewski (2017). "Protein aggregation in *Ehrlichia chaffeensis* during infection of mammalian cells." FEMS microbiology letters **364**(6): fnx059.

Lee, S., M. E. Sowa, Y.-h. Watanabe, P. B. Sigler, W. Chiu, M. Yoshida and F. T. F. Tsai (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.

Martin, I., J. Underhaug, G. Celaya, F. Moro, K. Teigen, A. Martinez and A. Muga (2013). "Screening and evaluation of small organic molecules as ClpB inhibitors and potential antimicrobials." J Med Chem **56**(18): 7177-7189.

Mattoo, R. U. H., S. K. Sharma, S. Priya, A. Finka and P. Goloubinoff (2013). "Hsp110 is a bona fide chaperone using ATP to unfold stable misfolded polypeptides and reciprocally collaborate with Hsp70 to solubilize protein aggregates." The Journal of biological chemistry **288**(29): 21399-21411.

Miot, M., M. Reidy, S. M. Doyle, J. R. Hoskins, D. M. Johnston, O. Genest, M.-C. Vitery, D. C. Masison and S. Wickner (2011). "Species-specific collaboration of heat shock proteins (Hsp) 70 and 100 in thermotolerance and protein disaggregation." Proceedings of the National Academy of Sciences **108**(17): 6915.

Mogk, A., T. Tomoyasu, P. Goloubinoff, S. Rüdiger, D. Röder, H. Langen and B. Bukau (1999). "Identification of thermolabile *Escherichia coli* proteins: prevention and reversion of aggregation by DnaK and ClpB." The EMBO Journal **18**(24): 6934-6949.

Parsell, D. A., A. S. Kowal, M. A. Singer and S. Lindquist (1994). "Protein disaggregation mediated by heat-shock protein Hspl04." Nature **372**(6505): 475-478.

Poole, K. (2012). "Bacterial stress responses as determinants of antimicrobial resistance." Journal of Antimicrobial Chemotherapy **67**(9): 2069-2089.

Ranaweera, C. B., P. Glaza, T. Yang and M. Zolkiewski (2018). "Interaction of substrate-mimicking peptides with the AAA+ ATPase ClpB from *Escherichia coli*." Arch Biochem Biophys **655**: 12-17.

Reidy, M., M. Miot and D. C. Masison (2012). "Prokaryotic chaperones support yeast prions and thermotolerance and define disaggregation machinery interactions." Genetics **192**(1): 185-193.

Rizo, A. N., J. Lin, S. N. Gates, E. Tse, S. M. Bart, L. M. Castellano, F. DiMaio, J. Shorter and D. R. Southworth (2019). "Structural basis for substrate gripping and translocation by the ClpB AAA+ disaggregase." Nature Communications **10**(1): 2393.

Rosenzweig, R., P. Farber, A. Velyvis, E. Rennella, M. P. Latham and L. E. Kay (2015). "ClpB N-terminal domain plays a regulatory role in protein disaggregation." Proceedings of the National Academy of Sciences of the United States of America **112**(50): E6872-E6881.

Schramm, F. D., K. Heinrich, M. Thüring, J. Bernhardt and K. Jonas (2017). "An essential regulatory function of the DnaK chaperone dictates the decision between proliferation and maintenance in Caulobacter crescentus." PLoS genetics **13**(12): e1007148-e1007148.

Torrente, M. P., L. M. Castellano and J. Shorter (2014). "Suramin Inhibits Hsp104 ATPase and Disaggregase Activity." PLOS ONE **9**(10): e110115.

Valastyan, J. S. and S. Lindquist (2014). "Mechanisms of protein-folding diseases at a glance." Disease Models & Mechanisms **7**(1): 9-14.

Winkler, J., A. Seybert, L. König, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A. S. Frangakis, A. Mogk and B. Bukau (2010). "Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing." The EMBO journal **29**(5): 910-923.

Zeymer, C., N. D. Werbeck, I. Schlichting and J. Reinstein (2013). "The molecular mechanism of Hsp100 chaperone inhibition by the prion curing agent guanidinium chloride." The Journal of biological chemistry **288**(10): 7065-7076.

Zhang, T., S. Kedzierska-Mieszkowska, H. Liu, C. Cheng, R. R. Ganta and M. Zolkiewski (2013). "Aggregate-Reactivation Activity of the Molecular Chaperone ClpB from *Ehrlichia chaffeensis*." PLOS ONE **8**(5): e62454.

Zolkiewski, M., T. Zhang and M. Nagy (2012). "Aggregate reactivation mediated by the Hsp100 chaperones." Arch Biochem Biophys **520**(1): 1-6.

Chapter 2 - Interaction of substrate-mimicking peptides with the AAA+ ATPase ClpB from *Escherichia coli*.

This chapter has been published as the following journal article:

Ranaweera CB, Glaza P, Yang T, Zolkiewski M. Interaction of substrate-mimicking peptides with the AAA+ ATPase ClpB from *Escherichia coli*. Archives of biochemistry and biophysics. 2018 Oct 1; 655:12-17. DOI: 10.1016/j.abb.2018.08.002

2.1 Abstract.

A molecular chaperone ClpB disaggregates and reactivates aggregated proteins in cooperation with DnaK, DnaJ, and GrpE. Within a cellular environment, ClpB must distinguish between properly folded and aggregated proteins by recognizing specific physical and/or chemical surface properties of the aggregates. However, the molecular mechanism of substrate binding to ClpB is poorly understood. We hypothesized that ClpB recognizes those polypeptide segments that promote protein aggregation because they are likely present at the surface of growing aggregates. We used an algorithm TANGO (Fernandez-Escamilla et al., Nat. Biotech. 2004, 22, 1302) to predict the aggregation-prone segments within the model ClpB-binding peptides and investigated interactions of the FITC-labeled peptides with ClpB using fluorescence anisotropy. We found that ClpB binds the substrate mimicking peptides with positive cooperativity, which is consistent with an allosteric linkage between substrate binding and ClpB oligomerization. The apparent affinity towards ClpB for peptides displaying different predicted aggregation propensities correlates with the peptide length. However, discrete aggregation-prone segments within the peptides are neither sufficient nor necessary for efficient interaction with ClpB. Our results suggest

that the substrate recognition mechanism of ClpB may rely on global surface properties of aggregated proteins rather than on local sequence motifs.

Keywords: Molecular chaperone ClpB, AAA+ ATPase, Protein-peptide interactions, Fluorescence anisotropy.

Abbreviations: FITC, fluorescein isothiocyanate; DMSO, dimethyl sulfoxide

2.2 Introduction.

ClpB is a bacterial member of the Hsp100 family of ATP dependent molecular chaperones that disaggregate and reactivate aggregated proteins in cooperation with Hsp70 and Hsp40 [1–3]. Hsp100s are found in bacteria, protozoa, fungi, and plants, but are absent in metazoan proteomes. Hsp100s belong to the superfamily of ATPases as- sociated with different activities (AAA+) [4]. In solution, ClpB forms a dynamic ensemble of homo oligomers with hexamers predominating in the presence of nucleotides [5–7]. Hexameric ClpB forms a cylinder-shaped particle with a narrow axial channel [8]. The cylindrical hexamer displays a significant conformational plasticity, as evidenced by its propensity to form semi open spiral like configurations [9]. The ClpB mediated aggregate reactivation is linked to extraction of single polypeptides from aggregated particles and their forced unfolding during translocation through the central channel in the hexamer [10]. A ratchet-like pulling mechanism of substrate translocation is supported by the conformational plasticity of the ClpB hexamer [11,12]. Extracted polypeptides are released from the ClpB channel after either partial translocation or complete unfolding [13–15].

While the substrate-translocation principle of the ClpB mediated protein disaggregation has been well documented and is indeed consistent with the mechanism of many AAA + ATPases, an upstream step of substrate recognition and binding to ClpB is poorly understood. In cells, the Hsp70/Hsp40 chaperones (DnaK/DnaJ/GrpE in bacteria) assist in targeting ClpB to the sites of

aggregate accumulation [16]. DnaK also stimulates the substrate extraction via a direct interaction with the coiled-coil domain of ClpB [17–20].

An essential role of DnaK notwithstanding, ClpB itself is capable of recognizing aggregated proteins and forming stable complexes with substrates in its ATP bound conformation [21–24]. An ATP hydrolysis deficient "substrate-trapping" ClpB variant became a useful tool for investigation of ClpB-substrate interactions [21]. Multiple evidence indicates that the ClpB-substrate complexes formed in the presence of ATP and absence of DnaK represent "on pathway" early intermediates of the disaggregation reaction. First, the DnaK independent ClpB aggregate interaction occurs in the ATP-bound, but not ADP bound or nucleotide free state, which is consistent with the well documented AAA + mechanism of substrate recognition [21,24,25]. Second, in the ATP-state, ClpB discriminates between the residue types in substrates [22] and between aggregates and native proteins [26]. Third, the residue substitutions in ClpB, which inhibit the DnaK independent substrate binding, also inhibit the disaggregation [23,24]. Fourth, the DnaK independent substrate binding is consistent with the co-chaperone independent reactivation of some aggregates by ClpB [27,28] and with a standalone disaggregase activity of the *Pseudomonas aeruginosa* ClpB variant, ClpG [29].

We hypothesized that ClpB recognizes those polypeptide segments that promote protein aggregation because they are likely present at the surface of growing aggregates and we decided to search for such sequence segments within the known peptide-based model substrates of ClpB [22]. We used an aggregation propensity prediction algorithm to identify the aggregation-prone segments in the model ClpB substrates and we investigated a correlation between the aggregation propensity and the ClpB binding affinity. We discovered that discrete aggregation prone segments

within the model ClpB substrates are neither sufficient nor necessary for strong interaction with ClpB.

2.3 Materials and methods.

2.3.1 Peptides and Proteins.

The peptides listed in Table 2.1 were purchased from Peptide 2.0 (Chantilly, Virginia). The peptides were N-terminally labeled with FITC and C-terminally amidated by the manufacturer. Peptide concentration was determined spectrophotometrically using the FITC molar extinction coefficient 77,000 $cm^{-1}M^{-1}$ at 494 nm. The ATP-hydrolysis deficient substrate-trapping ClpB variant (E279Q/E678Q) was produced as described before [23]. ClpB concentration was measured using the calculated extinction coefficient 0.38 cm^2/mg [30].

2.3.2 Aggregation-propensity prediction.

The TANGO algorithm (http://tango.crg.es) [31,32] was used to compute the aggregation propensity profiles for the peptide sequences (see Figure 2.1).

2.3.3 Peptide aggregation determination.

The turbidity of peptide solutions was measured using a BioTek Synergy H1 plate reader. A 96 well plate with 85 µl per well of buffer A (50 mM Tris-HCl, pH 7.5, 200 mM KCl, 20 mM $MgCl_2$) was equilibrated at 40 °C in the reader chamber. After 10 minutes of pre incubation, 15 µl of 1 mM peptide solution in DMSO was added to the wells with 15 µl of DMSO used as a control. Immediately after insertion into the reader, the plate was shaken inside the device for 10 seconds and the absorbance time course at 550 nm was recorded (see Figure 2.2).

2.3.4 ClpB-peptide interactions.

Fluorescence anisotropy measurements were performed using a PerkinElmer LS55 fluorescence spectrometer equipped with automatic polarizers. The FITC-labeled peptides were

dissolved in buffer A supplemented with 2 mM ATP or ADP. The peptide concentration was 18 nM (Figures 2.3A, 2.4, 2.5) or 80 nM (Figure 2.3B). The excitation and emission wavelengths were 494 nm and 521 nm, respectively. Before addition of ClpB, the peptide solutions were equilibrated at room temperature for 10 minutes and their fluorescence anisotropy was recorded. Next, the substrate trapping ClpB variant (E279Q/E678Q) was added in a stepwise manner to the cuvette with manual mixing. After each addition of ClpB, the samples were incubated for 2 minutes at room temperature, the anisotropy readings were collected during the following 2 minutes and averaged. Each titration was performed twice with the average anisotropy readings and their standard deviations calculated, using the GraphPad Prism software. The change of anisotropy, Δr, was obtained by subtracting the initial anisotropy reading for the ClpB-free peptides from the anisotropy measured during the titrations with ClpB. GraphPad Prism was used for nonlinear least-squares fitting of the model assuming cooperative binding of ClpB to the peptides according to the equation: $\Delta r = (\Delta r_{max} [ClpB]^h)/(Kdh + [ClpB]^h)$, where Δr is the observed change in anisotropy upon a ClpB addition, Δr_{max} is the fitting parameter corresponding to an anisotropy change at saturation, Kd is the apparent dissociation constant for the peptide-ClpB complex, and h is the Hill coefficient describing an extent of cooperativity in binding (see Table 2.2).

2.4 Results.

The 20 residue peptides B1 and B2 (see Table 2.1) are enriched in Lys and Arg and have been shown to mimic multiple hallmark features of the ClpB-aggregate interactions [22]. We used the TANGO algorithm [31,32] to test if the peptides B1 and B2 contain discrete sequence segments that could promote beta sheet driven aggregation. As shown in Figure. 2.1A and 2.1B, the peptide B1 contains one, and B2 two aggregation prone segments. Moreover, the predicted aggregation

scores for B2 are higher than for B1. Notably, the predicted aggregation prone segments in B1 and B2 are flanked by positively charged residues, a tendency found among the aggregation-promoting sequences across multiple proteomes [33].

We scrambled the sequences of B1 and B2 to modify their predicted aggregation profiles. First, we obtained the sequences B1N and B2N (see Table 2.1), for which the TANGO aggregation scores were insignificant (data not shown). Next, we obtained the sequence B2A, for which the aggregation scores were significantly higher than for B2 (see Figure 2.1C) and the sequence B2B with only one aggregation-prone segment and lower aggregation scores than for B2 (see Figure 2.1D). Thus, we designed a number of sequences with the same amino acid composition as B1 and B2, but with a range of different aggregation propensities, according to the TANGO predictions. The peptides B1, B1N, B2, B2N, B2A, and B2B have been synthesized and labeled with FITC at their N-termini. We also obtained the FITC-labeled peptides B1T, B2T1, and B2T2 (see Table 2.1), which correspond to the aggregation prone segments within B1 and B2, including their flanking basic residues.

We investigated if the calculated aggregation propensity of the peptides correlates with their solubility in aqueous solutions. All the labeled peptides were soluble in aqueous buffers at sub micromolar concentrations that were used in the binding experiments with ClpB (see below). However, we found that the peptides B2 and B2A caused a significant time dependent increase in solution turbidity when diluted to 150 μM at 40 °C. Notably, the turbidity increase for B2A was considerably faster than for its parent peptide B2 while those of B2N and B2B were insignificant (see Figure 2.2), which shows a correlation between the TANGO predictions of the peptide aggregation propensity (see Figure 2.1) and their behavior in solution. We did not detect a significant turbidity in the solutions of B1 and B1N, in agreement with their lower predicted

aggregation scores, as compared to B2 and B2A. Altogether, the results in Figure 2.2 validate the aggregation propensity ranking of the peptides based on the TANGO algorithm (see Figure 2.1).

We used fluorescence anisotropy to investigate the peptide interactions with ClpB. Binding of a FITC-labeled peptide to hexameric ClpB (~570 kDa) significantly decelerates the rotational diffusion of the fluorescein dye and produces an increase in fluorescence anisotropy. Indeed, a dose dependent increase in fluorescence anisotropy was observed during titration of B1 and B2 with the substrate trapping ClpB variant in the presence of ATP, but not when ADP was present (see Figure 2.3A). ClpB with ATP did not affect the anisotropy of a free fluorescein, which indicates that the interaction with ClpB is mediated by the peptide core and not by the label moiety (see Figure 2.3A). The ATP-dependent binding of B1 and B2 confirms that ClpB recognizes these peptides as pseudo substrates, as postulated before [22].

To estimate the binding stoichiometry, we conducted titrations of the peptides with ClpB using a ~4-fold higher peptide concentration, as compared to the conditions of Figure 2.3A. The fluorescence anisotropy increases for B1 and B2 reached an apparent saturation at ~1:1 ratio of the peptide/ClpB hexamer concentrations (see Figure 2.3B). This result indicates a single peptide binding site within the ClpB hexamer and is consistent with the substrate binding in the axial channel formed at the center of the hexamer, as determined before [22].

A sigmoidal shape of the binding isotherms for B1 and B2 in the presence of ATP (see Figure 2.3A), suggested a positive cooperativity in binding to ClpB. Fitting a cooperative binding model to the data in Figure 2.3A gave the Hill coefficient of ~2.5 for B1 and B2 (see Table 2.2). The apparent dissociation constant for B2 was approximately three fold lower than for B1. By using the N-terminally truncated ClpB, we confirmed a previous result [22] that the N-terminal

domain of ClpB, which supports binding to large aggregates [23], is not required for interactions with B1 and B2 (data not shown).

Next, we asked if the predicted aggregation prone segments in B1 and B2 (see Figure 2.1) are sufficient for producing a ClpB binding capability. As shown in Figure 2.4 and Table 2.2, ClpB showed much weaker affinity towards the shorter peptides B1T, B2T1, and B2T2 than their parent peptides B1 and B2. The binding of B1T, B2T1, and B2T2 to ClpB remained ATP-dependent (data not shown), which indicates a specific recognition of the TANGO predicted peptide segments by ClpB, but such recognition is insufficient to account for the binding affinity of B1 and B2.

We also asked if the predicted aggregation prone segments in B1 and B2 are necessary for recognition by ClpB. As shown in Figure 2.5A and 2.5B and Table 2 , the peptides B1N and B2N, which did not contain any predicted aggregating segments, showed a strong binding competency towards ClpB, equal to or exceeding that of the parent peptides B1 and B2. Moreover, ClpB interacted with the peptides B2A and B2B with a similar affinity (Figure 2.5B, Table 2.2), in spite of a significant difference between their predicted aggregation propensity profiles (see Figure 2.1) and aqueous solubility (see Figure 2.2).

2.5 Discussion.

TANGO is one of the algorithms for assessing aggregation propensity of polypeptides from their amino acid sequence [31,32,34]. TANGO accurately predicted aggregation of many peptide sequences, including the Alzheimer's disease linked variants of amyloid-β [31]. The predicted aggregation propensity is the strongest for sequence segments enriched in small hydrophobic and aromatic residues, but also Thr, Cys, Ser, Asn, and Gln. Sequences enriched in charged residues as well as Pro have the lowest tendency to aggregate [32]. A frequent occurrence of aggregation prone hydrophobicity enriched sequences is the price for maintaining stability of globular proteins

that rely on folding around hydrophobic cores. Interestingly, the TANGO analysis of multiple proteomes revealed that the aggregation-promoting sequences are preferentially capped with Arg or Lys [33]. This unexpected result suggested an electrostatic mechanism of counteracting protein aggregation that developed during a natural selection of protein sequences. Importantly, many molecular chaperones, including Hsp70 (DnaK) and Hsp100 (ClpB) show a preference towards predominantly basic sequences, in addition to the hydrophobic ones [22,35]. The affinity of ClpB towards positively charged residues is also consistent with the known capability of poly-Lys to activate the ClpB ATPase [36]. Thus, it could be postulated that the chaperones evolved to recognize aggregation-prone motifs, including their positively charged "gate- keepers". To our knowledge, the present study is the first attempt the test the above hypothesis for an Hsp100 chaperone.

The peptides B1 and B2 show a number of hallmark features attributed to the bona fide ClpB substrates: their interaction with ClpB requires ATP (see Figure 2.3A), they stimulate the ClpB ATPase, compete with aggregated proteins for binding to ClpB, and shift the self-association equilibrium of ClpB towards the hexamer [22]. B1 and B2 are enriched in basic residues that are essential for recognition by ClpB (see Table 2.1). Indeed, a replacement of the basic residues in B1 or B2 with Alanines or with acidic residues abolished their binding to ClpB [22]. The TANGO analysis of B1 and B2 revealed that these peptides contain discrete aggregation prone segments flanked by basic residues, consistent with the "gatekeeper" model (see Figure 2.1).

Fluorescence anisotropy titrations allowed us to obtain binding isotherms for the peptides interacting with ClpB in the ATP-bound state. We found that B1 and B2 become saturated with a hexameric ClpB at ~1:1 ratio (see Figure 2.3B), which is consistent with the single substrate binding site located at the entrance to the ClpB channel [11,22]. Interestingly, the peptide-ClpB

interaction showed a positive cooperativity (see Figure 2.3A). Since the cooperativity in binding of B1 and B2 to ClpB is not apparently associated with multiple binding sites (see Figure 2.3B), it is likely a manifestation of an allosteric linkage between the ClpB-peptide interaction and the ClpB oligomerization, which was observed before [22,37]. The nanomolar dissociation constants for the B1- ClpB and B2-ClpB complexes are consistent with the previous results [37].

We scrambled the sequences of B1 and B2 to modify their apparent aggregation promoting properties. Interestingly, a swap between Phe within the aggregation prone segment of B1 and Ala (see Table 2.1) was sufficient to reduce the predicted aggregation index for the peptide B1N. Similarly, a redistribution of residues within B2 produced either a decrease or an increase of the predicted aggregation index in B2N, B2A, and B2B (see Figure 2.1). We discovered that the peptides with the highest predicted aggregation index, i.e., B2A and B2 visibly aggregated in solution, as compared with the less aggregation-prone peptides (see Figure 2.2), which validated the computational predictions.

Since ClpB is targeted towards aggregated substrates, rather than soluble misfolded proteins that are recognized by DnaK and GroEL [38], we hypothesized that the aggregation prone sequences mediate interactions with ClpB. Unexpectedly, the peptides B1T, B2T1, and B2T2, which contain only the predicted aggregating sequences (see Table 2.1, Figure 2.1) bind to ClpB with much lower affinities than the full length B1 and B2 (see Figure 2.4 and Table 2.2). This result indicates that discrete aggregation prone sequences, including their flanking basic residues are not sufficient for efficient interactions with ClpB. Moreover, we found that the ClpB binding affinities of the scrambled sequences B1N, B2N, B2A, and B2B were similar to those of the parent peptides B1 and B2 (see Figure 2.5 and Table 2.2). Thus, the aggregation-prone segments in the peptides are not necessary for efficient binding to ClpB.

We did not find a direct correlation between the overall aggregation propensity of the investigated peptides and their affinity towards ClpB. In spite of containing the aggregation prone segments, the shorter peptides B1T, B2T1, and B2T2 showed a weak interaction with ClpB, which suggests that the length of the recognized peptide contributes to the binding to the chaperone, in agreement with a previously published result [37]. Quite unexpectedly, the peptides with a wide variability of the aggregation propensity profiles (see Figure 2.1) interacted with ClpB with similar affinities (see Figure 2.5 and Table 2.2). Notably, B1N interacted with ClpB with almost an order of magnitude lower Kd than B1, in spite of its insignificant aggregation score.

Our results suggest that the substrate recognition mechanism of ClpB may rely on multiple modes of interaction with broadly distributed motifs on the surface of aggregated particles, rather than on binding discrete local sequence segments. Possibly, the model peptides used in this study do not engage all substrate recognition sites in ClpB, such as those located in the N-terminal domain [23,39], which limits their utility as mechanistic probes.

2.6 Declarations of interest.

None.

2.7 Acknowledgements.

This study was supported by a grant from the National Institutes of Health AI121366 (to M.Z.) and by the Kansas Agricultural Experiment Station (contribution 18–623-J).

Table 2.1. Peptides used in this study.

Peptide	Sequence
B1	AHAWQHQGKTLFISRKTYRI
B1T	KTLFISR
B1N	FHAWQHQGKTLAISRKTYRI
B2	QRKLFFNLRKTKQRLGWFNQ
B2T1	KLFFNLR
B2T2	RLGWFNQ
B2N	QLWQGQFRNLRRKTFFKKLN
B2A	QGQFRNLRRKQLWTFFLNKK
B2B	QQFRNRRKLTFLKKLGWFNQ

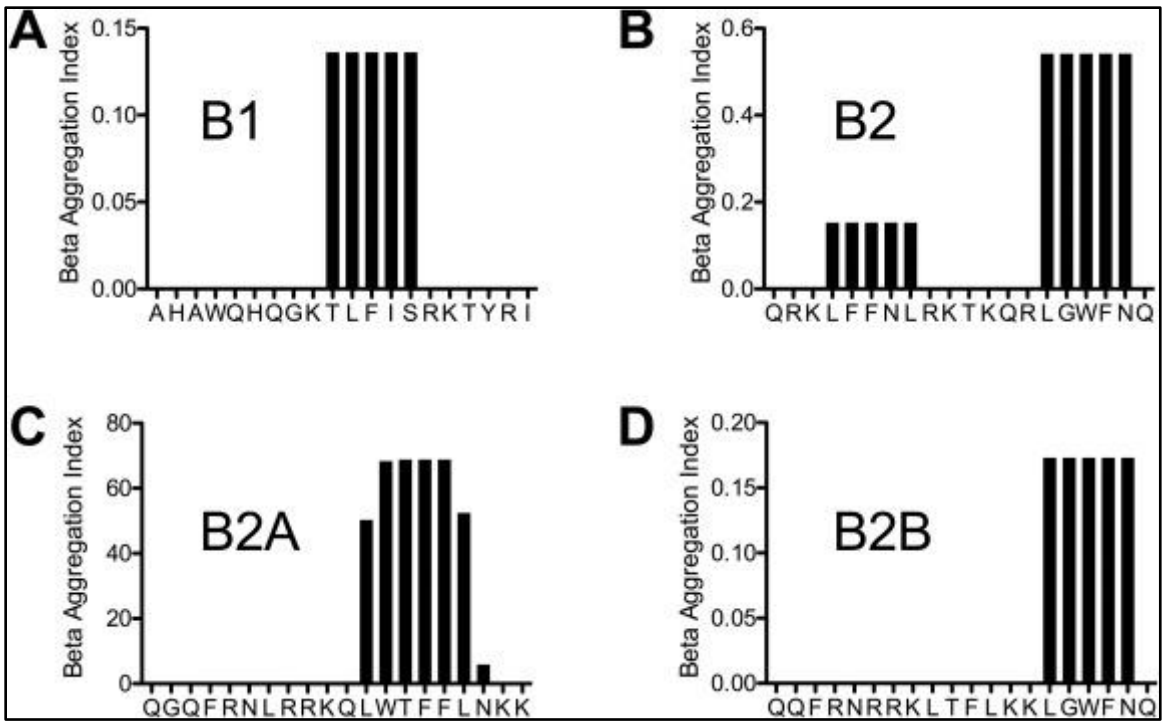


Figure 2.1. Prediction of the aggregation propensity of the peptides B1 (A), B2 (B), B2A (C), and B2B (D).

The aggregation index computed using the TANGO algorithm [31] is shown for the amino acid sequence of each peptide.

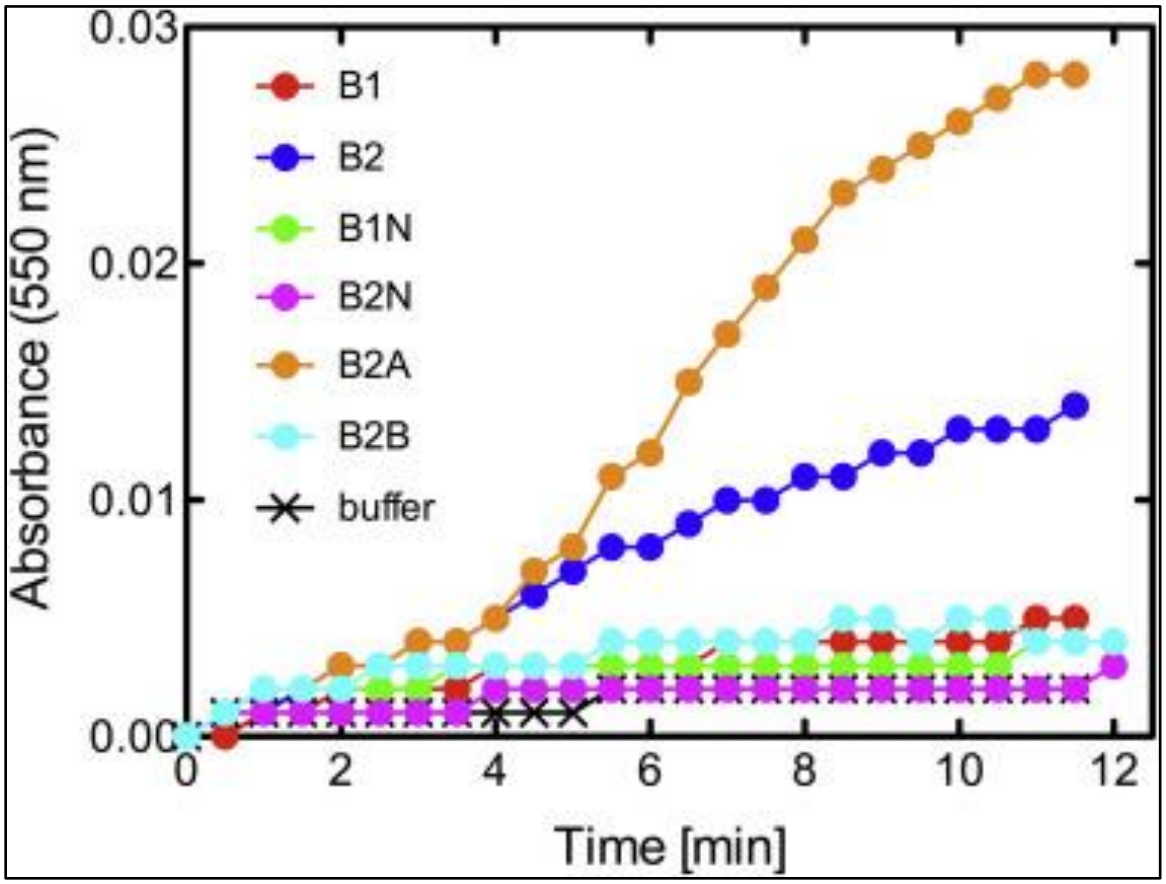


Figure 2.2. Aggregation propensity of the peptides in solution.

The solution turbidity was monitored at 40 °C after diluting the peptides to 150 μM (see Materials and Methods). This experiment was done by Dr Przemyslaw Glaza.

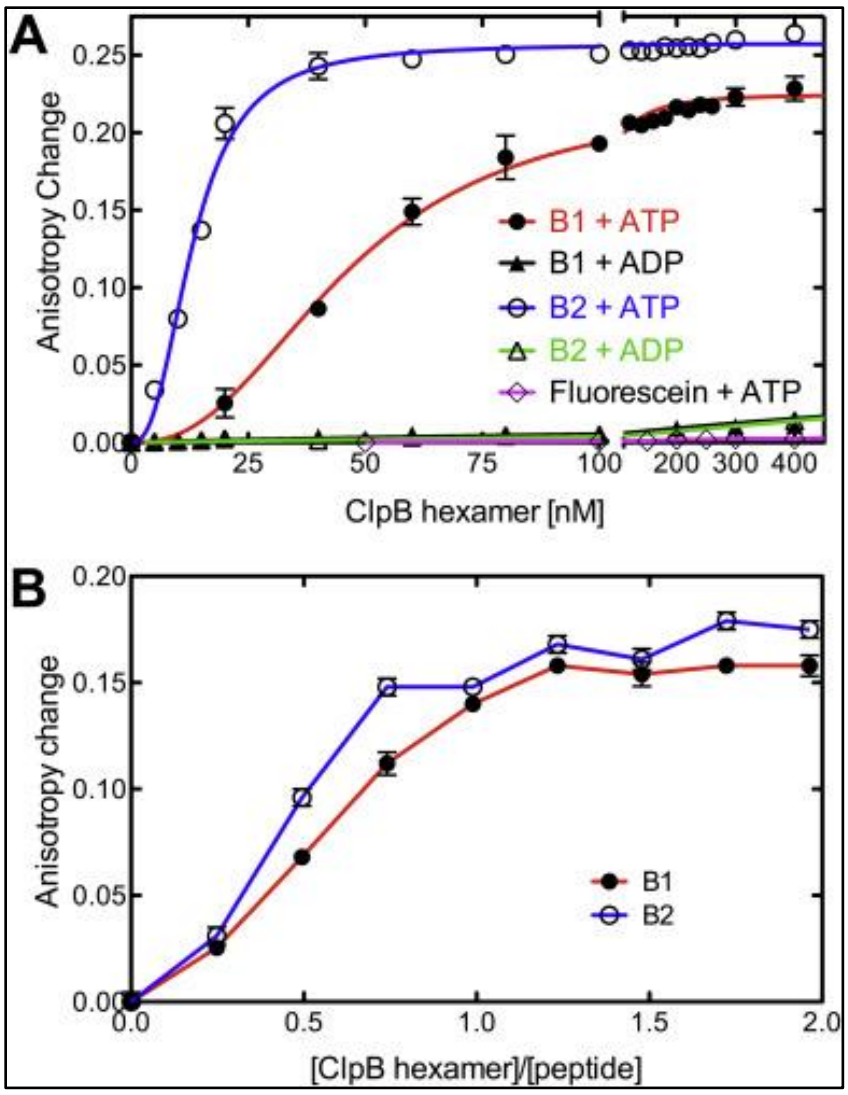


Figure 2.3. ATP dependent interaction between ClpB and the peptides B1 and B2.

The FITC fluorescence anisotropy changes were determined at room temperature upon titrating ClpB into the FITC-labeled peptide solutions. (A) Fluorescence anisotropy was determined for 18 nM peptides in 50 mM Tris-HCl, pH 7.5, 200 mM KCl, 20 mM $MgCl_2$ with 2 mM ATP or ADP. (B) Fluorescence anisotropy was determined for 80 nM peptides in 50 mM Tris-HCl, pH 7.5, 200 mM KCl, 20 mM $MgCl_2$ with 2 mM ATP (2.8.4 section B experiment was done by Taihao Yang).

Table 2.2. ClpB-peptide interaction parameters in the presence of ATP at 25 °C.

Peptide	Dissociation constant Kd [nM]	Hill coefficient h
B1	46.8 ± 1.5	2.40 ± 0.18
B1T	329 ± 39	1.72 ± 0.14
B1N	7.26 ± 0.32	1.64 ± 0.13
B2	13.24 ± 0.42	2.52 ± 0.22
B2T1	912 ± 231	1.13 ± 0.06
B2T2	>1000	not determined
B2N	17.54 ± 0.92	1.85 ± 0.19
B2A	10.17 ± 0.80	1.31 ± 0.17
B2B	11.51 ± 0.52	1.47 ± 0.11

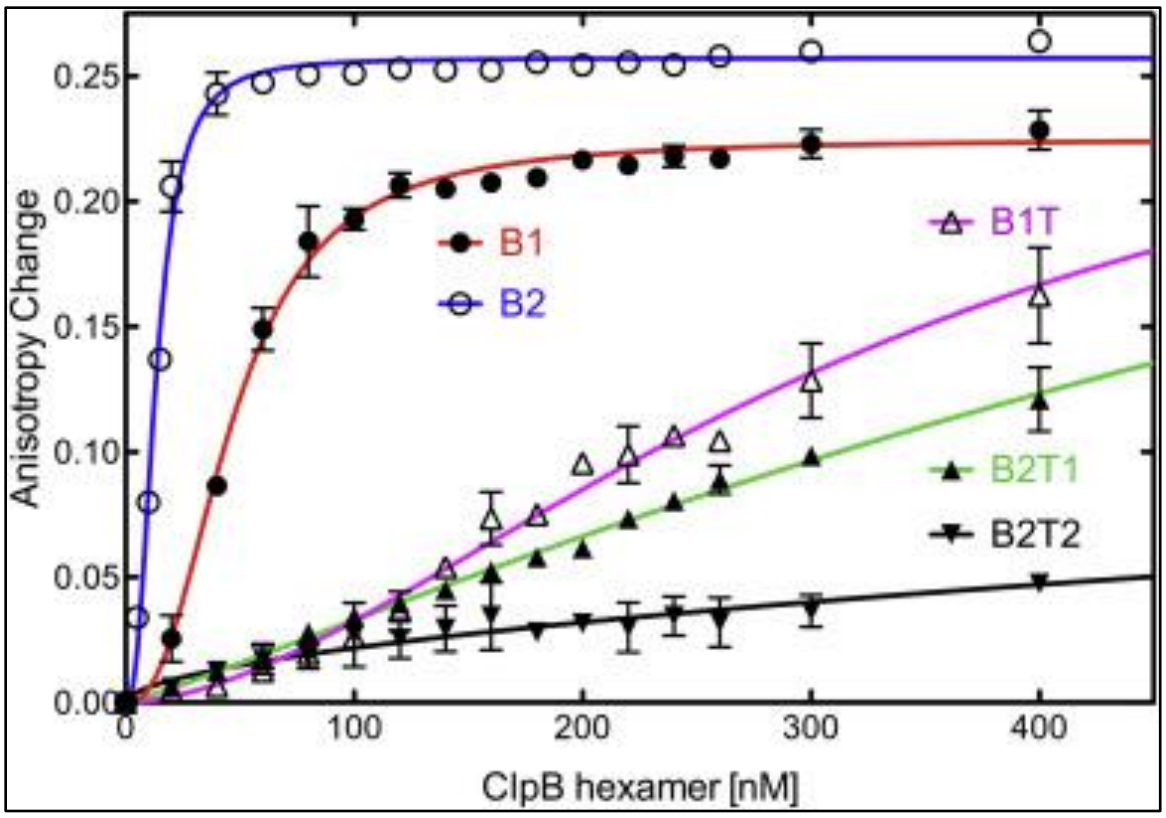


Figure 2.4. Interaction between ClpB and the predicted aggregation prone segments of the peptides B1 and B2.

The FITC fluorescence anisotropy changes were determined at room temperature upon titrating ClpB into the 18 nM peptide solutions in 50 mM Tris-HCl, pH 7.5, 200 mM KCl, 20 mM $MgCl_2$ with 2 mM ATP.

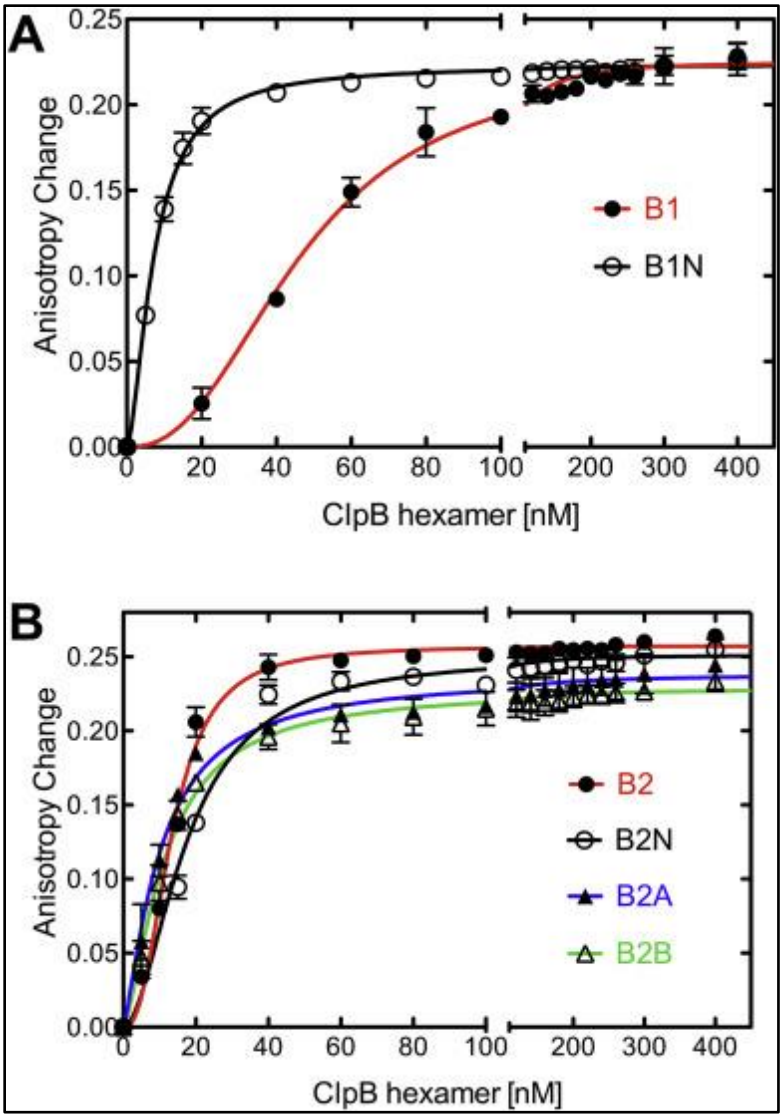


Figure 2.5. Interaction between ClpB and the peptides with different predicted aggregation propensities.

The FITC fluorescence anisotropy changes were determined at room temperature upon titrating ClpB into the 18 nM peptide solutions in 50 mM Tris-HCl, pH 7.5, 200 mM KCl, 20 mM $MgCl_2$ with 2 mM ATP.

2.9 References.

1. M. Zolkiewski, T. Zhang, M. Nagy, Aggregate reactivation mediated by the Hsp100 chaperones, Arch. Biochem. Biophys. 520 (2012) 1–6.

2. A. Mogk, E. Kummer, B. Bukau, Cooperation of Hsp70 and Hsp100 chaperone machines in protein disaggregation, Front. Mol. Biosci. 2 (2015) 22.

3. A. Mogk, B. Bukau, H.H. Kampinga, Cellular handling of protein aggregates by disaggregation machines, Mol. Cell. 69 (2018) 214–226.

4. A.F. Neuwald, L. Aravind, J.L. Spouge, E.V. Koonin, AAA+: a class of chaperone like ATPases associated with the assembly, operation, and disassembly of protein complexes, Genome Res. 9 (1999) 27–43.

5. M. Zolkiewski, M. Kessel, A. Ginsburg, M.R. Maurizi, Nucleotide dependent oligomerization of ClpB from *Escherichia coli*, Protein Sci. 8 (1999) 1899–1903.

6. V. Akoev, E.P. Gogol, M.E. Barnett, M. Zolkiewski, Nucleotide induced switch in oligomerization of the AAA+ ATPase ClpB, Protein Sci. 13 (2004) 567–574.

7. J. Lin, A.L. Lucius, Examination of ClpB quaternary structure and linkage to nucleotide binding, Biochemistry 55 (2016) 1758–1771.

8. S. Lee, M.E. Sowa, Y.H. Watanabe, P.B. Sigler, W. Chiu, M. Yoshida, F.T. Tsai, The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state, Cell 115 (2003) 229–240.

9. T. Uchihashi, Y.H. Watanabe, Y. Nakazaki, T. Yamasaki, H. Watanabe, T. Maruno, K. Ishii, S. Uchiyama, C. Song, K. Murata, R. Iino, T. Ando, Dynamic structural states of ClpB involved in its disaggregation function, Nat. Commun. 9 (2018) 2147.

10. J. Weibezahn, P. Tessarz, C. Schlieker, R. Zahn, Z. Maglica, S. Lee, H. Zentgraf, E.U. Weber-Ban, D.A. Dougan, F.T. Tsai, A. Mogk, B. Bukau, Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB, Cell 119 (2004) 653–665.

11. C. Deville, M. Carroni, K.B. Franke, M. Topf, B. Bukau, A. Mogk, H.R. Saibil, Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase, Adv. Sci. 3 (2017) e1701726.

12. S.N. Gates, A.L. Yokom, J. Lin, M.E. Jackrel, A.N. Rizo, N.M. Kendsersky, C.E. Buell, E.A. Sweeny, K.L. Mack, E. Chuang, M.P. Torrente, M. Su, J. Shorter, D.R. Southworth, Ratchet like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104, Science 357 (2017) 273–279.

13. C. Schlieker, I. Tews, B. Bukau, A. Mogk, Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides, FEBS Lett. 578 (2004) 351–356.

14. T. Haslberger, A. Zdanowicz, I. Brand, J. Kirstein, K. Turgay, A. Mogk, B. Bukau, Protein disaggregation by the AAA+ chaperone ClpB involves partial threading of looped polypeptide segments, Nat. Struct. Mol. Biol. 15 (2008) 641–650.

15. T. Li, C.L. Weaver, J. Lin, E.C. Duran, J.M. Miller, A.L. Lucius, *Escherichia coli* ClpB is a non-processive polypeptide translocase, Biochem. J. 470 (2015) 39–52.

16. J. Winkler, J. Tyedmers, B. Bukau, A. Mogk, Hsp70 targets Hsp100 chaperones to substrates for protein disaggregation and prion fragmentation, J. Cell Biol. 198 (2012) 387–404.

17. T. Haslberger, J. Weibezahn, R. Zahn, S. Lee, F.T. Tsai, B. Bukau, A. Mogk, M domains couple the ClpB threading motor with the DnaK chaperone activity, Mol. Cell. 25 (2007) 247–260.

18. Y. Oguchi, E. Kummer, F. Seyffer, M. Berynskyy, B. Anstett, R. Zahn, R.C. Wade, A. Mogk, B. Bukau, A tightly regulated molecular toggle controls AAA+ disaggregase, Nat. Struct. Mol. Biol. 19 (2012) 1338–1346.

19. F. Seyffer, E. Kummer, Y. Oguchi, J. Winkler, M. Kumar, R. Zahn, V. Sourjik, B. Bukau, A. Mogk, Hsp70 proteins bind Hsp100 regulatory M domains to activate AAA+ disaggregase at aggregate surfaces, Nat. Struct. Mol. Biol. 19 (2012) 1347–1355.

20. R. Rosenzweig, S. Moradi, A. Zarrine-Afsar, J.R. Glover, L.E. Kay, Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction, Science 339 (2013) 1080–1083.

21. J. Weibezahn, C. Schlieker, B. Bukau, A. Mogk, Characterization of a trap mutant of the AAA+ chaperone ClpB, J. Biol. Chem. 278 (2003) 32608–32617.

22. C. Schlieker, J. Weibezahn, H. Patzelt, P. Tessarz, C. Strub, K. Zeth, A. Erbse, J. Schneider Mergener, J.W. Chin, P.G. Schultz, B. Bukau, A. Mogk, Substrate recognition by the AAA+ chaperone ClpB, Nat. Struct. Mol. Biol. 11 (2004) 607–615.

23. M.E. Barnett, M. Nagy, S. Kedzierska, M. Zolkiewski, The amino-terminal domain of ClpB supports binding to strongly aggregated proteins, J. Biol. Chem. 280 (2005) 34940–34945.

24. M. Nagy, H.C. Wu, Z. Liu, S. Kedzierska Mieszkowska, M. Zolkiewski, Walker-A threonine couples nucleotide occupancy with the chaperone activity of the AAA+ ATPase ClpB, Protein Sci. 18 (2009) 287–293.

25. P.I. Hanson, S.W. Whiteheart, AAA+ proteins: have engine, will work, Nat. Rev. Mol. Cell Biol. 6 (2005) 519–529.

26. M. Nagy, I. Guenther, V. Akoyev, M.E. Barnett, M.I. Zavodszky, S. Kedzierska Mieszkowska, M. Zolkiewski, Synergistic cooperation between two ClpB isoforms in aggregate reactivation, J. Mol. Biol. 396 (2010) 697–707.

27. S.M. Doyle, J. Shorter, M. Zolkiewski, J.R. Hoskins, S. Lindquist, S. Wickner, Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein remodeling activity, Nat. Struct. Mol. Biol. 14 (2007) 114–122.

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

29. C. Lee, K.B. Franke, S.M. Kamal, H. Kim, H. Lunsdorf, J. Jager, M. Nimtz, J. Trcek, L. Jansch, B. Bukau, A. Mogk, U. Romling, Standalone ClpG disaggregase confers superior heat tolerance to bacteria, Proc. Natl. Acad. Sci. U.S.A. 115 (2018) E273–E282.

30. C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, Protein Sci. 4 (1995) 2411–2423.

31. A.M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence dependent and mutational effects on the aggregation of peptides and proteins, Nat. Biotechnol. 22 (2004) 1302–1306.

32. F. Rousseau, J. Schymkowitz, L. Serrano, Protein aggregation and amyloidosis: confusion of the kinds? Curr. Opin. Struct. Biol. 16 (2006) 118–126.

33. F. Rousseau, L. Serrano, J.W. Schymkowitz, How evolutionary pressure against protein aggregation shaped chaperone specificity, J. Mol. Biol. 355 (2006) 1037–1047.

34. A.P. Pawar, K.F. Dubay, J. Zurdo, F. Chiti, M. Vendruscolo, C.M. Dobson, Prediction of "aggregation-prone" and "aggregation susceptible" regions in proteins associated with neurodegenerative diseases, J. Mol. Biol. 350 (2005) 379–392.

35. S. Rudiger, L. Germeroth, J. Schneider Mergener, B. Bukau, Substrate specificity of the DnaK chaperone determined by screening cellulose bound peptide libraries, EMBO J. 16 (1997) 1501–1507.

36. Z. Liu, V. Tek, V. Akoev, M. Zolkiewski, Conserved amino acid residues within the Amino terminal domain of ClpB are essential for the chaperone activity, J. Mol. Biol. 321 (2002) 111–120.

37. T. Li, J. Lin, A.L. Lucius, Examination of polypeptide substrate specificity for *Escherichia coli* ClpB, Proteins 83 (2015) 117–134.

38. B. Bukau, J. Weissman, A. Horwich, Molecular chaperones and protein quality control, Cell 125 (2006) 443–451.

39. T. Zhang, E.A. Ploetz, M. Nagy, S.M. Doyle, S. Wickner, P.E. Smith, M. Zolkiewski, Flexible connection of the N-terminal domain in ClpB modulates substrate binding and the aggregate reactivation efficiency, Proteins 80 (2012) 2758–2768.

Chapter 3 - Repurposing p97 inhibitors for chemical modulation of the bacterial ClpB–DnaK bichaperone system.

This chapter has been published as the following journal article:

Glaza P, Ranaweera CB, Shiva S, Roy A, Geisbrecht BV, Schoenen FJ, Zolkiewski M. Repurposing p97 inhibitors for chemical modulation of the bacterial ClpB/DnaK bichaperone system. The Journal of biological chemistry. 2020 Nov 13:jbc.RA120.015413

DOI: https://doi.org/10.1074/jbc.ra120.015413

3.1 Abstract.

The ClpB–DnaK bichaperone system reactivates aggregated cellular proteins and is essential for survival of bacteria, fungi, protozoa, and plants under stress. AAA+ ATPase ClpB is a promising target for the development of antimicrobials because a loss of its activity is detrimental for survival of many pathogens and no apparent ClpB orthologs are found in metazoans. We investigated ClpB activity in the presence of several compounds that were previously described as inhibitor leads for the human AAA+ ATPase p97, an antitumor target. We discovered that N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ), the least potent among the tested p97 inhibitors, binds to ClpB with a Kd~60 μM and inhibits the casein-activated, but not the basal, ATPase activity of ClpB with an IC_{50} 5 μM. The remaining p97 ligands, which displayed a higher affinity toward p97, did not affect the ClpB ATPase. DBeQ also interacted with DnaK with a Kd~100 μM and did not affect the DnaK ATPase but inhibited the DnaK chaperone activity in vitro. DBeQ inhibited the reactivation of aggregated proteins by the ClpB–DnaK bichaperone system in vitro with an IC_{50}~5 μM and suppressed the growth of cultured *Escherichia coli*. The DBeQ induced loss of *E. coli* proliferation was exacerbated by heat shock but was nearly eliminated in a ClpB deficient *E.*

coli strain, which demonstrates a significant selectivity of DBeQ toward ClpB in cells. Our results provide chemical validation of ClpB as a target for developing novel antimicrobials. We identified DBeQ as a promising lead compound for structural optimization aimed at selective targeting of ClpB and/or DnaK.

Keywords: molecular chaperone, protein aggregation, small molecule inhibitor, antimicrobial compound, AAA+ ATPase, ClpB, DnaK

Abbreviations: DBeQ, N^2,N^4-dibenzylquinazoline-2,4-diamine, DMSO, dimethyl sulfoxide, G6PDH, glucose-6-phosphate dehydrogenase, Hsps, heat shock proteins, KJE, DnaK–DnaJ–GrpE, NBD, nucleotide binding domain, SPR, surface plasmon resonance, TCEP, tris(2-carboxyethyl)phosphine

3.2 Introduction.

On infection of a host, pathogens experience heat shock and oxidative stress, and their survival depends on molecular responses to these external conditions. The pathogen stress response has emerged as a critical mechanism for the development of novel antimicrobials (1–3). However, no successful inhibition of pathogen stress response machinery has been developed to date, mostly due to sequence conservation among heat-shock proteins (Hsps) across different domains of life.

The Hsp100 family offers a unique opportunity for inhibitor development among Hsps. The ATP dependent Hsp100 chaperones are essential for the survival of bacteria (where such proteins are called ClpB), lower eukaryotes (fungi and protozoa), and plants under stressful conditions. Unlike other heat shock families, Hsp100 chaperones are not found in animals or humans (4, 5). Hsp100 chaperones, in cooperation with Hsp70 and Hsp40 (bacterial DnaK and DnaJ), are uniquely responsible for reactivation of aggregated proteins in microbial cells (6–8).

Hsp100 chaperones are required for invasiveness and/or in host survival of multiple significant bacterial and protozoan pathogens including the ESKAPE bacteria (9–11), *Mycobacterium tuberculosis* (12, 13), *Mycoplasma pneumoniae* (14), *Salmonella typhimurium* (15), *Shigella dysenteriae* (16, 17), *Porphyromonas gingivalis* (18), *Francisella tularensis* (19), *Leptospira interrogans* (20), *Leishmania donovani* (21), and the tick transmitted rickettsia *Ehrlichia chaffeensis* (22, 23). In the malaria parasite *Plasmodium falciparum*, there are two apparent Hsp100 isoforms and each one supports a different essential element of Plasmodium infectivity (24–27). Thus, inhibition of Hsp100 may suppress infectivity and survival of a broad range of clinically relevant pathogenic microorganisms. Unfortunately, no high affinity Hsp100 selective inhibitors are known.

Hsp100 chaperones are a subgroup of the AAA+ family of ATPases associated with different cellular activities (28). Like most AAA+ ATPases, Hsp100s form cylinder shaped hexamers and use energy from ATP hydrolysis to induce conformational rearrangements in protein substrates (29). Hsp100 mediated reactivation of aggregates is linked to a forced extraction of polypeptides from aggregated particles and their subsequent unfolding by translocation through a narrow channel at the center of the Hsp100 hexamer (30, 31). Small molecule inhibitors have been recently developed for the human AAA+ ATPase, p97, a promising antitumor target (32–34), which led to clinical evaluation and validated the drugability of AAA+ proteins. We reasoned that inhibitors of AAA+ ATPases distantly related to Hsp100 might serve as prototype scaffolds for the development of Hsp100 selective ligands. In this study, we investigated the effects of several p97 inhibitor leads on the activity of bacterial ClpB. We discovered that one of the known p97 inhibitors, N^2,N^4- dibenzylquinazoline-2,4-diamine (DBeQ), is a promising candidate compound for targeting Hsp100 chaperones.

3.3 Experimental procedures.

3.3.1 DNA constructs.

The nucleotide sequence encoding the full-length *E. coli* DnaK (residues 1–638) was amplified from plasmid pTTQ19dnaK^{+} (61) using the following primers:

DnaK-NcoI (5'ATATACCATGGGTAAAATAATTGGTATCGACC 3') and

DnaK-XhoI (5' TATATCTCGAGTTTTTTGTCTTT GACTTCTTC 3'), where the engineered restriction sites are underlined. The PCR product did not contain the DnaK STOP codon. Subsequently, the DnaK sequence was cloned into the NcoI/XhoI restriction site of pET28a. The STOP codon was introduced between the DnaK coding sequence and the C-terminal His-tag sequence in pET28a by site directed mutagenesis with the following primers:

DnaK_ STOPf (5' GTCAAAGACAAAAAATAAGAGCACCACCA CCACCACC 3') and DnaK_STOPr (5' GGTGGTGGTG GTGGTGCTCTTATTTTTTGTCTTTGAC 3'), where the STOP codon is underlined. The final construct (plasmid pET28-DnaK) was verified by DNA sequencing and was used for production of DnaK.

The DNA sequence encoding the NBD of *E. coli* DnaK (residues 1–388) was PCR amplified from plasmid pTTQ19dnaK^{+} (61) using the following primers:

NBD-NcoI (5' ATATACCATGGGTAAAATAATTGGTATCGACC 3') and

NBD-XhoI (5' TATATCTCGAGGTCTTTTACGTCACCAGTCAGAAC 3'), where the engineered restriction sites are underlined. Subsequently, the NBD sequence was cloned into the NcoI/XhoI restriction site of pET28a. The construct (pET28-NBD1) was verified by DNA sequencing and used for production of the DnaK NBD.

The plasmids pHSG575-ClpB-YFP and pBAD-YFP- Luciferase (44) were provided by Bernd Bukau and Axel Mogk (ZMBH, University of Heidelberg, Germany).

3.3.2 Proteins.

E. coli ClpB and its variants E279Q, E678Q, E279Q/E678Q, W462F, W543F, and W462F/W543F were produced as described previously (52, 62, 63). To produce *E. coli* DnaK, the *E. coli* strain BL21(DE3) was transformed with pET28-DnaK. Single transformants were used to prepare an overnight culture in the LB media supplemented with kanamycin (50 μg/ml). On the following day, the overnight culture was diluted 50-fold in 1 l of LB with kanamycin (50 μg/ml). The culture was incubated at 30°C. When absorbance at 600 nm reached 0.4, IPTG was added to the final concentration of 0.5 mM and the culture was incubated for 2 h at 30°C with shaking at 200 RPM. Then, bacterial cells were collected by centrifugation (Beckman JA-14, 3000 relative centrifugal force, 15 min, 4 °C). The pellet was suspended in 20 ml of cold buffer KA (25 mM Hepes-KOH, pH 7.6, 50 mM KCl, 2.5 mM $MgCl_2$, 1mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol) by vortexing. Cells were disrupted by sonication followed by centrifugation (rotor Beckman JA-20 11,500 RPM, 60 min, 4°C). The supernatant was loaded on a Q Sepharose column (3.2 cm × 18 cm) equilibrated with buffer KA. After washing with buffer KA (150 ml), proteins were eluted with an increasing gradient of buffer KB (25 mM Hepes-KOH, pH 7.6, 550 mM KCl, 2.5 mM $MgCl_2$, 1 mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol). The collected fractions, which contained DnaK, were pooled and then diluted 10 fold in buffer KC (50 mM Tris-HCl, pH 7.8, 1 mM EDTA, 5 mM β-mercaptoethanol, 10% glycerol). The sample was loaded on a DEAE Sephacel column (3.2 × 12 cm) equilibrated with buffer KC. After washing the column with 100 ml of buffer KD (50 mM Tris-HCl pH 7.8, 50 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol, 5% glycerol), proteins were eluted with an increasing gradient of buffer KE (50 mM Tris-HCl, pH 7.8, 400 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol, 5% glycerol). The collected fractions were concentrated with a centrifugal filter device (10000 molecular weight cutoff, 3500 RPM,

4°C). Concentrated samples were resolved on a Superdex 200 gel filtration column equilibrated with buffer KA. Fractions containing DnaK with purity greater than 95% were collected and concentrated to 2 mg/ml with a centrifugal filter device. Protein concentration was determined by the Bradford method. The aliquots of DnaK were stored at −20 °C.

To produce the DnaK NBD, *E. coli* strain BL21(DE3) was transformed with pET28-NBD1. Single transformants were used to prepare overnight culture in the LB media supplemented with kanamycin (50 μg/ml). The overnight culture was diluted 50 fold in 1 liter of LB with kanamycin and incubated at 30°C. When absorbance at 600 nm reached 0.4, IPTG was added to the final concentration of 1 mM and the culture was incubated for 3 h at 30 °C with shaking at 200 RPM. The bacterial cells were collected by centrifugation (Beckman JA-14, 3000 relative centrifugal force, 15 min, 4 °C). The pellet was suspended in 20 ml of cold T10 buffer (40 mM Tris-HCl, pH 8.0, 100 mM KCl, 10 mM imidazole) by vortexing. Cells were disrupted by sonication. The soluble fraction was obtained by centrifugation (Beckman JA-20, 11,500 RPM, 60 min, 4°C). The supernatant was loaded on a column with 5 ml of nickel nitrilotriacetic acid resin (Ni-NTA) equilibrated with buffer T10. The column was washed with 50 ml of buffer T20 (40 mM Tris-HCl, pH 8.0, 100 mM KCl, 20 mM imidazole). Proteins were eluted with buffer T250 (40 mM Tris-HCl, pH 8.0, 100- mM KCl, 250 mM imidazole). Fractions containing the NBD were pooled and concentrated with a centrifugal filter device (10000 molecular weight cut-off, 3500 RPM, 4°C). The concentrated NBD sample was resolved on a Superdex 200 gel filtration column equilibrated with buffer B (50 mM Tris-HCl, pH 7.5, 0.2 M KCl, 20 mM $MgCl_2$, 10% glycerol, 1 mM EDTA, 1 mM DTT). Fractions containing the NBD with purity greater than 95% were collected and concentrated to 5 mg/ml. Protein concentration was determined by the Bradford method. The aliquots of the NBD were stored at −20°C.

Firefly luciferase, G6PDH from *Leuconostoc mesenteroides*, DnaJ, and GrpE were obtained as described before (25, 52, 62, 64). Peptide B2 and its FITC-labeled variant were obtained as described before (65). κ-Casein and FITC–casein were obtained from Sigma-Aldrich (St Louis, MO).

3.3.3 Bacterial strains.

The following strains of *E. coli* were used: MC4100, MC4100ΔclpB::kan (40), dnaK103 (41), and dnaK756 (66).

3.3.4 Chemicals.

Candidate inhibitor compounds DBeQ, ML240, and NMS-873 were purchased from Sigma Aldrich (St Louis, MO) and used directly without further purification; C3 and C6 were obtained from TimTec (Newark, DE). The compounds in a powder form were dissolved in DMSO at 10 mM and used in biochemical studies after further dilutions.

3.3.5 ATPase activity assays.

To determine the ATPase activity, ClpB was diluted to 28 μg/ml (49 nM hexamer) in buffer C (50 mM Tris-HCl, pH 7.4, 20 mM $MgCl_2$, 1 mM EDTA, 0.5 mM tris(2-carboxyethyl) phosphine [TCEP]). In some experiments, the ClpB solution was supplemented with 17.4 μg/ml (10 μM) κ-casein. The ClpB aliquots (18 μl) were mixed with 1 μl of the investigated compounds at different concentrations in 100% DMSO or with DMSO as a control. A sample without ClpB was used as a baseline control. After a 10 minutes preincubation at 37 or 45 °C, the ATP hydrolysis reaction was initiated by adding 1 μl of 100 mM ATP. The samples were incubated for 60 minutes (basal ClpB activity in the absence of κ-casein) or 15 min (in the presence of κ-casein) at 37 °C or 30 min at 45 °C in the absence of casein. For the E678Q ClpB variant, the incubation time in the presence of κ-casein was 60 minutes. After incubation, 15 μl of each sample was mixed with 200

µl of the ammonium molybdate/malachite green reagent (67) dispensed into a 96 well plate, followed by an addition of 30 µl of 34% sodium citrate (68). The plate was agitated inside a Synergy H1 reader (BioTek, Winooski, VT) for 15 minutes at room temperature, and the samples' absorbance was measured at 630 nm. The readouts for ClpB containing samples were corrected for absorbance of samples without ClpB, to account for nonenzymatic production of inorganic phosphate. A standard curve obtained with different inorganic phosphate concentrations was used to determine the amount of phosphate produced from ATP in the presence of ClpB.

To determine the ATPase activity of DnaK, 18 µl of DnaK solution or DnaK NBD solution prepared in buffer C was mixed with 1 µl of the investigated compounds or DMSO as a control. Samples without DnaK were used as control blanks. The samples were preincubated for 10 minutes at 37 °C. The reaction was initiated by adding 1 µl of 100 mM ATP bringing the final concentration of DnaK to 1.1 µM or DnaK NBD to 3.4 µM. The samples were incubated at 37 °C for 70 minutes for DnaK or 75 minutes for NBD. The concentration of inorganic phosphate produced from ATP in the presence of DnaK was determined as described above for ClpB.

3.3.6 Aggregate reactivation assays.

G6PDH was chemically denatured by mixing 5 µl of the G6PDH stock solution (441 µM) with 5 µl of buffer A (10 M urea, 16% glycerol, 40 mM DTT) preheated at 47 °C. The samples of G6PDH in buffer A were incubated for 5 minutes at 47 °C. Subsequently, 5 µl of denatured G6PDH was mixed with 95 µl of buffer B (50 mM Tris-HCl, pH 7.4, 20 mM $Mg(OAc)_2$, 30 mM KCl, 1 mM EDTA, 1 mM β-mercaptoethanol) preheated at 47 °C to initiate protein refolding, which, under these conditions, is inefficient and leads to misfolding and aggregation of G6PDH (69). G6PDH aggregate production was arrested after 1 minute (for refolding by KJE) or 15 minutes (for refolding by KJE with ClpB) by transferring into ice for 2 minutes. To observe the

aggregate reactivation, aggregated G6PDH was diluted 10 fold in buffer C (50 mM Tris-HCl, pH 7.4, 20 mM $MgCl_2$, 1 mM EDTA, 0.5 mM TCEP) containing the chaperones KJE (5.5 μM DnaK, 1.1 μM DnaJ, 1.1 μM GrpE) or KJE ClpB (5.5 μM DnaK, 1.1 μM DnaJ, 1.1 μM GrpE, 1.65 μM) ClpB (hexamer)). G6PDH in buffer C without chaperones was used as a control for a spontaneous aggregate reactivation. The chaperone G6PDH samples (18 μl) were supplemented with 1 μl of DBeQ in DMSO or pure DMSO as a control. The samples were incubated for 2 minutes at 30 °C. The aggregate reactivation was initiated by adding 1 μl of 100 mM ATP, followed by incubation for 60 minutes at 30 °C. To determine the enzymatic activity of G6PDH, a 5 μl aliquot from the reactivated sample was mixed with 195 μl of buffer D (50 mM Tris-HCl, pH 7.8, 5 mM $MgCl_2$, 2 mM glucose-6-phosphate, 1 mM $NADP^+$) in a 96 well plate that had been preheated at 30 °C. The plate was inserted into Synergy H1 plate reader (BioTek, Winooski, VT) with the chamber temperature at 30 °C, and the sample absorbance at 340 nm was measured after 10 minutes. The amount of NADPH produced was calculated based on the NADPH molar extinction coefficient, $\text{Ɛ}340 = 6220\ [M^{-1} \bullet cm^{-1}]$.

Firefly luciferase (220 μM) was diluted 300 fold in PBS buffer with 1 mg/ml bovine serum albumin and then incubated for 12 minutes at 45 °C. The chaperone solutions were prepared in buffer C (50 mM Tris-HCl, pH 7.4, 20 mM $MgCl_2$, 1 mM EDTA, 0.5 mM TCEP): KJE (1.15 μM DnaK, 1.15 μM DnaJ, 0.57 μM GrpE) and KJE-ClpB (1.15 μM DnaK, 1.15 μM DnaJ, 0.57 μM GrpE, 1.72 μM ClpB (hexamer)). Then, 17 μl of a chaperone solution was mixed with 1 μl of DBeQ in DMSO or pure DMSO as a control and preincubated for 10 min at 25 °C. Next, 1 μl of the heat denatured luciferase (0.7 μM) and 1 μl of ATP (100 mM) were added to the samples, followed by incubation for 20 minutes at 25 °C. To determine the luciferase activity, 5 μl of each sample was transferred into 100 μl of the luciferase substrate (Promega), previously dispensed into

a white 96 well plate. Luminescence was measured using Synergy H1 plate reader (BioTek, Winooski, VT).

3.3.7 Surface plasmon resonance.

ClpB, DnaK, and the NBD of DnaK were extensively dialyzed against the SPR buffer (20 mM Hepes, pH 7.4, 150 mM NaCl, 20 mM $MgCl_2$, 1-mM EDTA, 1 mM DTT). After dialysis, the protein concentration was determined by the Bradford method. The DBeQ protein interactions were studied using a Biacore T200 instrument (GE Healthcare) at 25 °C. The proteins were diluted to 300 μg/ml in 10 mM sodium acetate buffer (pH 5 for ClpB DnaK or pH 4 for NBD) and immobilized on HC1500M sensor chips (Xantec) by standard amine coupling chemistry as described previously (70). A reference flow cell was created by ethyl(dimethylaminopropyl) carbodiimide/N-hydroxysuccinimide activation followed by quenching with 1 M ethanolamine (pH 9.0). Solvent correction curves were obtained at the beginning and end and after every 50 injection cycles by injecting varying DMSO concentrations (4.0, 4.4, 4.6, 4.8, 5.0, 5.2, 5.4, and 5.9% [v/v]). All experiments were performed using a flow rate of 30 μl/min.

Binding isotherms were obtained with the ligand diluted to different concentrations in the SPR running buffer HBSMTD (20 mM Hepes, pH 7.4, 140 mM NaCl, 20 mM $MgCl_2$, 5% (v/ v) DMSO, 0.005% (v/v) Tween-20). Replicate injections were performed for all samples, and a buffer only injection was performed for every 12 injection cycles. To test the structural integrity of ClpB after immobilization on the sensor chips, we performed control titrations with ATPγS and adenosine. The dissociation constants determined from those experiments were 20 μM for ATPγS and 2 mM for adenosine. These preliminary experiments indicated that the immobilization on the SPR chip preserved the oligomeric structure of ClpB, because monomeric ClpB has low affinity toward ATP (Kd > 1 mM (62)). Moreover, the immobilized ClpB discriminated

between a nucleotide analog and a nucleoside, which indicated an adequate structural integrity of ClpB on the sensor chips.

The resulting sensorgrams were analyzed using Biacore T200 Evaluation Software v3.0 according to the following procedure. Solvent correction curves were generated and applied to all data sets, and all sensorgrams were double referenced by subtracting the most recent buffer blank injection. The signal immediately before injection stop of these corrected sensorgrams was treated as the binding response. GraphPad Prism was used for nonlinear least squares fitting of the model assuming either noncooperative or cooperative binding of DBeQ to DnaK and ClpB, respectively.

3.3.8 Bacterial viability.

E. coli MC4100, MC4100ΔclpB, dnaK103, and dnaK756 strains were used to determine an influence of DBeQ on bacterial growth and viability. Bacteria were maintained in the LB media, and in the case of the MC4100ΔclpB strain, LB was supplemented with 30 μg/ml kanamycin. All experiments were initiated by preparing overnight cultures inoculated from single colonies and grown at 37 °C. The dnaK103 and dnaK756 overnight cultures were grown at 30 °C. On the following day, the cultures were diluted 100 fold in 10 ml of LB without antibiotics and incubated at 37 °C. The culture optical density was monitored at 600 nm. When absorbance at 600 nm reached 0.4, the culture was divided into 1 ml aliquots and supplemented with 10 μl of DMSO or different concentrations of DBeQ in DMSO. The samples were immediately transferred into an incubator/shaker at temperature 37, 45, or 50 °C. Bacteria were cultured for up to 4 hours with shaking (200 RPM). At specific time points, 100 μl of each culture was withdrawn and serially diluted in sterile 0.9% NaCl up to 10^6 fold dilution. To estimate the number of viable cells, 5 μl or 10 μl of the diluted cultures was spread on the LB agar plate without an antibiotic. After a complete adsorption of liquid on the LB agar surface, the plates were incubated overnight at 37 °C. On the

following day, the bacterial colonies were counted, and the colony forming units/ml were calculated.

3.3.9 Chemical library screening.

An absorbance based assay for ClpB ATPase activity was developed by using BioMol Green Reagent (Enzo Life Sciences, Farmington, NY) for screening in a 384 microplate based format. The optimized assay conditions were as follows: ClpB (150 nM), ATP (200 mM), and κ-casein (25 μM) in 100 mM Tris/HCl, pH 8, 10 mM $MgCl_2$, 1 mM DTT, and 1 mM EDTA. The samples were incubated for 10 minutes at 37 °C, and the ATP hydrolysis reaction was terminated by adding 80 μl of BIOMOL Green. Absorbance at 620 nm was measured with a PerkinElmer EnSpire. The optimized assay was used to screen compounds drawn from Selleck Bioactive library (1900 compounds: Selleck Chemicals, Houston, TX) in the presence or absence of κ-casein.

3.3.10 Analytical ultracentrifugation.

Protein samples (0.5 ml) of wt ClpB and its variants W462F and W543F were dialyzed against 3 changes of 1 l of buffer AUC (50 mM Tris-HCl, pH 7.4, 200 mM KCl, 20 mM $MgCl_2$, 1 mM EDTA, 0.5 mM TCEP). After dialysis, the samples were centrifuged (15 min, 13,000 RPM, 4°C). The supernatants were collected, and the protein concentration was determined with the Bradford assay. Sedimentation velocity experiments were performed in Beckman Optima XL-I analytical ultracentrifuge equipped with a 4-cell rotor and 2-sector cells. The ClpB samples were diluted to 1 mg/ml in the AUC buffer supplemented with DMSO (5% (v/v) final concentration) with 1mM ATPɤS, with or without 50 μM DBeQ. Centrifugation was performed at 48,000 RPM at 20 °C. The sedimentation profiles were collected using the interference detection system and analyzed with the Origin software supplied by the instrument manufacturer using the time derivative method of Stafford (71). Observed sedimentation coefficients were corrected to values

corresponding to the density and viscosity of water (s20,w) using Sednterp software (www.jphilo.mailway.com) with additional corrections for the presence of 5% DMSO (72).

3.3.11 Fluorescence anisotropy.

To monitor interactions between casein and ClpB, we used FITC-casein and the substrate-trapping ClpB E279Q/E678Q variant (52, 73). The titration of FITC casein with ClpB E279Q/E678Q in the presence of ATP or ADP and the fluorescence anisotropy measurements were performed as previously described (65).

To monitor interactions between the peptide B2 and DnaK, the DnaK stock was diluted in buffer C (50 mM Tris-HCl, pH 7.4, 20 mM $MgCl_2$, 1 mM EDTA, 0.5 mM TCEP) to a final concentration of 1 µM. DBeQ solutions in DMSO (20 µl) were mixed with 372 µl of 1 µM DnaK. After 10 minutes of incubation at room temperature, the samples were supplemented with 8 µl of 1 µM FITC-B2 in buffer C (FITC-Ahx- QRKLFFNLRKTKQRLGWFNQ-NH2) (51). The FITC-B2 sample without DnaK was used as a control. After a 30 minutes incubation at room temperature, fluorescence anisotropy of FITC-B2 was measured with a PerkinElmer LS55 (λexc = 485 nm (slit: 15 nm), λem = 521 (slit: 20 nm)).

3.3.12 Western blotting.

The ClpB protein level in *E. coli* cells was analyzed by Western blotting. Overnight culture of each strain was diluted 100-fold and incubated at 37 °C with shaking (200 RPM). When the absorbance at 600 nm reached 0.4, the culture was split into two: for one, the incubation at 37 °C continued, while the second one was transferred to 45 °C. After 2 hours of incubation at the target temperatures, the cells were collected by centrifugation. The supernatant was discarded, and cell pellets were suspended in 160 µl of 2x Laemmli sample buffer. Sixty microliters of each sample was resolved in duplicates in 8% SDS-PAGE gel. The gel was cut into two parts, one part was

used for blotting onto a nitrocellulose membrane and the second part of gel was stained in Coomassie R-250. The membrane was blocked with 5% nonfat dried milk dissolved in TBST buffer (19 mM Tris-HCl, pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.2% Tween-20) overnight at 4 °C. Next, the membrane was incubated with rabbit polyclonal anti-ClpB IgG (1:50,000 in milk) for 1 hour at room temperature. After incubation with the primary antibody, the membrane was washed in TBST buffer (4 times, 10 minutes). Next, the membrane was incubated with anti-rabbit horseradish peroxidase conjugated secondary antibodies for 1 hour at room temperature and washed in TBST (4 times, 10 minutes) and then TBST without Tween-20 (10 minutes). Signal detection was performed using SuperSignal West Pico Chemiluminescent Substrate (Pierce) and an Azure c500 digital imaging system (Azure Biosystems, Dublin, CA).

3.3.13 Confocal microscopy.

ClpB-YFP and Luciferase-YFP were expressed at low levels in *E. coli* strains ΔclpB, dnaK103, and dnaK756 as described before (44). The expression of ClpB-YFP and YFP-Luciferase was initiated by addition of 200 μM IPTG or 0.02% (w/v) arabinose, respectively. The bacterial cells were cultured at 30°C with 100 μM DBeQ or DMSO only. Protein translation was stopped by addition of erythromycin (30 μg/ml), and each culture was divided into two samples. One sample remained at 30 °C (control), whereas the other sample was transferred to 45 °C for 30 minutes (heat shock conditions). A thin layer of 1% agarose (TopVision Low Melting Point Agarose, Thermo Scientific) was prepared on a clean microscope glass slide. A small drop of 1% agarose was placed near the frosted end of the glass slide while bringing a second glass slide at a 30 to 45 ° angle from above, allowing the agarose drop to spread along the contact edge. An *E. coli* culture sample (1 μl) was immediately placed on the agarose-covered glass slide, covered with a cover slip, and used for fluorescence imaging. Fluorescence imaging was performed using a Carl

Zeiss LSM 880 Airyscan confocal microscope equipped with a Plan- Apochromat 63x/1.40 Oil DIC M27 objective coupled with a YFP Filter (excitation at 514 nm, emission at 561 nm, laser strength 0.98%). Zen 2.3 lite software was used for image capture and processing.

3.3.14 Data availability.

All relevant data are available within this manuscript and the associated Supporting Information.

3.4 Results.

3.4.1 DBeQ inhibits the casein activated ATPase activity of ClpB.

We investigated the ATPase activity of *Escherichia coli* ClpB in the presence of three previously described p97 inhibitors: DBeQ (32), its p97-optimized derivative ML240 (35), and an alkylsulfanyl-1,2,4-triazole NMS-873 (36). We also tested two ClpB inhibitor candidates previously identified through a high throughput ClpB interaction screen: C3 and C6 (37) (Appendix A Fig. A.1). None of the above compounds, except C3, showed a significant inhibition of the basal ClpB ATPase at 100 μM (Fig.3.1A). However, only C3 and DBeQ strongly inhibited the ClpB ATPase in the presence of casein, a known pseudo substrate of ClpB and an activator of the ClpB ATPase (Fig.3.1, A–B). We also identified DBeQ, but not NMS-873, as a candidate ClpB ATPase inhibitor in the presence of casein in the pilot screen of 388 bioactive compounds (Appendix A Fig. A.2). We therefore focused the following studies on DBeQ as a novel ClpB inhibitor candidate.

The solubility of DBeQ in aqueous buffers is limited (32), but we found that the DBeQ absorbance obeys the Beer–Lambert law and the solution turbidity does not significantly increase for concentrations of up to 100 μM DBeQ in the presence of 1% dimethyl sulfoxide (DMSO) (Appendix A Fig. A.3). DBeQ suppressed the ClpB ATPase in the presence of casein to the basal

level with an apparent IC_{50} 5 μM (Fig.3.1B) but did not inhibit the ATP dependent binding of casein to ClpB (Fig.3.1C). We introduced E-to-Q substitutions in the Walker B motif in each of the two ATP binding domains of ClpB (D1 and D2, see Appendix A Fig. A.4) to disable ATP hydrolysis activity in each domain. Of the two ATP binding domains of ClpB, only the C-terminal one (i.e., D2) showed casein-induced activation and was affected by DBeQ (Fig.3.1, D–E).

ClpB contains two Trp residues, W543 located between D1 and D2 and W462 in the coiled-coil middle domain (see Appendix A Fig. A.4). Remarkably, the ClpB variants with each of the two Trp residues substituted with Phe became sensitive to inhibition by DBeQ even in the absence of casein (Fig.3.1F). The sedimentation coefficient distributions for wt ClpB and the Trp/Phe variants (Appendix A Fig. A.5) showed a dominant ~ 13 to 15 S molecular species, which corresponds to a hexameric ClpB in equilibrium with monomers (38). Importantly, no shifts of the sedimentation coefficient distribution maximum toward a lower s20,w were observed in the presence of DBeQ (Appendix A Fig. A.5), which indicates that DBeQ does not induce dissociation of the ClpB hexamers. Overall, the different parameters tested above showed that DBeQ directly affects ClpB but does not compromise its structural integrity, with the basal ATPase unaffected by DBeQ even at 45 °C (Appendix A Fig. A.6).

3.4.2 DBeQ inhibits the aggregate reactivation activity of ClpB–DnaK.

We tested the effect of DBeQ on the reactivation of aggregates prepared from two model substrates: firefly luciferase and bacterial glucose-6-phosphate dehydrogenase (G6PDH), whose native activity was not affected by DBeQ (Appendix A Fig. A.7). Either substrate can be reactivated by DnaK–DnaJ–GrpE (KJE), albeit with a significantly lower efficiency than by ClpB with KJE (Fig.3.2, A and C), which allowed us to test the effect of DBeQ on both components of the ClpB–DnaK bichaperone system. Unexpectedly, we found that DBeQ inhibited both ClpB

dependent and ClpB-independent reactivation of aggregated luciferase with a similar IC_{50} value (Fig.3.2B), which suggests that the rate of luciferase reactivation is limited by the activity of KJE. For aggregated G6PDH, however, ClpB dependent reactivation was inhibited by DBeQ with a several fold lower IC_{50} than ClpB independent reactivation by KJE alone (Fig.3.2D). These results indicate that DBeQ inhibits the activity of not only ClpB but also DnaK. However, we did not detect inhibition of the DnaK ATPase activity by DBeQ (Appendix A Fig. A.8A) or suppression of DnaK binding to a substrate mimicking peptide (Appendix A Fig. A.8B).

3.4.3 DBeQ interacts with ClpB and DnaK.

We used surface plasmon resonance (SPR) to verify a direct interaction between DBeQ and ClpB (Fig.3.3) or DnaK (Appendix A Fig. A.9). The binding isotherms for DBeQ and ClpB showed positive cooperativity (Fig.3.3B) with an apparent Kd 60 μM and a Hill coefficient 2.3 (Appendix A Supplementary Table A.1). The DBeQ binding isotherms were virtually identical for wt ClpB and its 2 Trp/Phe variants (Fig.3.3). In contrast to ClpB, the DBeQ interaction with DnaK was noncooperative (Appendix A Fig. A.9, A–C) with an apparent Kd 100 μM (Appendix A Supplementary Table A.1). The above apparent binding affinity of DnaK for DBeQ should be considered approximate, as it is close to the ligand solubility limit.

We determined that DBeQ interacted with the isolated nucleotide binding domain (NBD) of DnaK with a similar affinity as with the full-length protein (Appendix A Fig. A.9, D–F, Supplementary Table A.1). As observed before (39), the ATPase activity of the DnaK NBD is similar to that of the full length DnaK (Appendix A Fig. A.8A). We found that DBeQ did not inhibit the ATPase activity of the isolated NBD of DnaK (Appendix A Fig. A.8A).

3.4.4 DBeQ suppresses the proliferation and survival of *E. coli*.

ClpB is produced in *E. coli* cultured at 37°C and is strongly upregulated during heat shock (Appendix A Fig. A.10). The ΔclpB strain, which does not produce ClpB (40), does not show growth defects at 37°C or during a mild heat shock at 45°C but rapidly loses viability at 50°C (Fig.3.4, B, D and F). DBeQ inhibited the growth of *E. coli* at 37°C in a concentration dependent manner (Fig.3.4A), but the DBeQ effects were eliminated in the ΔclpB strain (Fig.3.4B). The viability of *E. coli* at 45°C (but not the ΔclpB strain) was strongly inhibited by DBeQ (Fig.3.4,C–D and Fig.3.5). On expression of a ClpB–YFP fusion protein in the ΔclpB strain (for the experiments shown in Fig.3.6), the cells' susceptibility to DBeQ was restored and was exacerbated at higher levels of ClpB expression (Appendix A Fig. A.11). A critical role of ClpB in *E. coli* survival under severe stress manifested at 50°C, as shown by a loss of viability of the ΔclpB strain even without DBeQ (Fig. 3.4F). *E. coli* with an intact clpB gene survived the heat shock at 50°C for several hours but lost viability at the lowest DBeQ concentrations tested (Fig.3.4E). These experiments demonstrate that a treatment of *E. coli* with DBeQ is toxic to the cells, but the DBeQ toxicity at a moderate stress of 45°C is more severe than that caused by a loss of ClpB in the ΔclpB strain.

To test the effects of DBeQ on the cross talk between ClpB and DnaK, we investigated the effects of DBeQ on viability of two *E. coli* strains with defective DnaK: dnaK103, which produces a truncated inactive DnaK that does not bind to its substrates (41, 42), and dnaK756, in which substrate release from DnaK is inefficient (43). We found that DBeQ suppressed viability of dnaK103, but surprisingly, viability of dnaK756 was not affected by DBeQ (Fig. 3.5).

To further explore the effects of DBeQ on the function of ClpB in *E. coli* cells, we tracked the ClpB localization with confocal fluorescence microscopy using a ClpB construct with a

C-terminally fused YFP. As demonstrated earlier (42, 44), heat induced aggregates of thermosensitive proteins, such as *Photinus pyralis* luciferase, accumulate at the poles of *E. coli* cells (Appendix A Fig. A.12). ClpB and DnaK colocalize with the aggregates ((44) and Fig. 3.6). Importantly, the absence of a functional DnaK in the dnaK103 strain prevents ClpB from accumulating with the aggregates (Fig. 3.6 and (42)), but in the presence of DBeQ, ClpB colocalizes with the aggregates even in dnaK103 (Fig. 3.6). The heat shock induced localization of ClpB at the cellular poles in the dnaK756 strain with or without DBeQ was indistinguishable from that in wt *E. coli* (Appendix A Fig. A.13). The results with dnaK103 *E. coli* show that DBeQ affects localization of ClpB in cells exposed to heat shock. The fluorescent foci observed in *E. coli* in the presence of DBeQ are more diffuse than those found in the absence of the ligand, an effect exacerbated in the dnaK103 strain (Fig. 3.6). Because the foci localization was attributed to nucleoid occlusion (44), an altered foci appearance in the presence of DBeQ may reflect a modified size of the nucleoid free space in cells whose proliferation was challenged with DBeQ and/or a loss of a functional DnaK.

3.5 Discussion.

We hypothesized that small molecule ligands of one AAA+ ATPase, p97, might also interact with another one, ClpB. We tested three inhibitors that show an increasing potency toward p97 (45): DBeQ (p97 IC_{50}~3 μM), ML240 (p97 IC_{50}~0.1 μM), and NMS-873 (p97 IC_{50} 0.02 μM). We discovered that among the three ligands, only DBeQ, the least potent one toward p97, affected ClpB (Fig.3.1) with an apparent IC_{50} close to that displayed toward p97. ML240 is a p97 optimized DBeQ derivative, and its loss of potency toward ClpB suggests that a structural diversity among AAA+ ATPases is sufficient to allow discrimination between different ligands from the same

chemical family. Thus, it may be feasible to orthogonally modify DBeQ and achieve an enhanced selectivity toward ClpB with a lower potency toward p97 and the other AAA+ ATPases.

Antimicrobial activity of 2,4-diaminoquinazolines has been reported before (46, 47), but without a clear identification of their cellular targets. We have now shown that ClpB is the main target of DBeQ in *E. coli* under both permissive conditions and during heat stress (Figs.3.4 and 3.5). It is remarkable that the DBeQ induced inhibition of *E. coli* viability depends on production of a single protein, ClpB. Although the ΔclpB strain does not respond to DBeQ, it becomes susceptible upon expression of ClpB (Appendix A Fig. A.11), which indicates that ClpB is required for the sensitivity of *E. coli* toward DBeQ. p97, the known DBeQ target, is not produced in bacteria. Interestingly, eukaryotic parasites, including *Plasmodium* and *Leishmania*, produce p97 as well as ClpB. An antiparasitic activity of DBeQ has been reported (48, 49), and it remains to be determined if the compound's preferred in vivo target is p97 or ClpB.

As shown by SPR, DBeQ binds to ClpB with a positive cooperativity (Hill coefficient >2), which suggests multiple binding sites (Fig.3.3 and Appendix A supplementary Table A.1). Detection of multiple binding sites for DBeQ is consistent with the oligomeric structure of ClpB. The apparent IC_{50} for DBeQ is an order of magnitude lower than the apparent Kd (Figs. 3.1–3.3, and Appendix A Supplementary Table A.1), which suggests that partial saturation of the DBeQ sites in the ClpB hexamer is sufficient for a strong inhibition. Indeed, it has been observed that incorporation of a single inactive subunit into a hexameric ClpB blocks aggregate reactivation (50).

ClpB mediated reactivation of protein aggregates depends on several molecular processes: the ClpB hexamer assembly, ATP dependent substrate binding, DnaK dependent substrate engagement, and ATP hydrolysis dependent substrate unfolding/translocation (4). Importantly,

there is an allosteric linkage between substrate binding and the ATPase engine of ClpB, as demonstrated by activation of the ATPase in the presence of substrates (7). We found that DBeQ does not inhibit the ClpB hexamer assembly, substrate binding, or the basal ATPase of ClpB, but it does affect the linkage between substrate binding and the ATPase of the C-terminal D2 module (Fig. 3.1). Indeed, whereas substrate binding primarily occurs at the N-terminal region of ClpB (51, 52), a partial insertion of the substrate into the ClpB channel and its contact with the D2 module have been observed even without ATP hydrolysis (53). Moreover, D2 is fully responsible for the casein induced activation of the ClpB ATPase because the activation is lost upon disabling D2 (Fig.3.1E).

DBeQ mediated modulation of the allostery within ClpB is further demonstrated by the apparent uncoupling of the DBeQ effects from substrate binding in the two Trp-to-Phe ClpB variants (Fig.3.1F). Trp462 is located within the mobile coiled- coil domain of ClpB (Appendix A Fig. A.4), which controls the ATPase activity of the ClpB hexamer (54, 55). Trp543 is located at the interface between D1 and D2 (Appendix A Fig. A.4). Both Trp462 and Trp543 side chains are exposed at the inter subunit interface within the ClpB hexamer (Appendix A Fig. A.14). A similarity between the DBeQ binding isotherms for wt ClpB and its Trp/Phe variants (Fig.3.3 and Appendix A Supplementary Table A.1) indicates that the Trp/Phe substitutions do not produce additional DBeQ binding sites in ClpB. Altogether, the enhanced sensitivity of the Trp/Phe ClpB variants toward DBeQ suggests that the ligand binding site(s) may be located between the hexamer subunits and in the vicinity of the D1–D2 junction.

We found that aggregate reactivation with the ClpB–DnaK bichaperone system becomes inefficient in the presence of DBeQ (Fig.3.2). Because the energy generated from ATP in the AAA+ engine is directly coupled with substrate translocation/ unfolding (56, 57), our results

suggest that DBeQ affects a linkage between the ClpB ATPase and substrate translocation by suppressing the acceleration of ATPase in response to a substrate, which decelerates substrate translocation and makes its reactivation inefficient.

An unexpected result of our study is a discovery of interactions between DBeQ and DnaK (Appendix A Fig. A.9) as well as inhibition of the DnaK mediated reactivation of protein aggregates in vitro (Fig. 3.2). Unlike ClpB, DnaK is monomeric in solution and supports noncooperative binding of DBeQ (Appendix A Fig. A.9, Supplementary Table A.1). DBeQ binds to the NBD of DnaK (Appendix A Fig. A.9) but does not inhibit its ATPase (Appendix A Fig. A.8A). Our results are consistent with a recent report on the binding of amino quinazolines within the ATP site of Hsp70 (58) but may also indicate a nonspecific interaction due to a predominance of hydrophobic groups in DBeQ. Interestingly, C3 also inhibits KJE (37) as well as ClpB and even luciferase (Fig.3.1A, Appendix A Fig. A.7A). These results demonstrate a challenge in designing selective inhibitors for different families of molecular chaperones.

In vitro binding of DBeQ to DnaK notwithstanding, ClpB appears as the main target of DBeQ in *E. coli* because the compound's effects manifest even in the absence of functional DnaK in the dnaK103 strain but do not manifest in the ΔclpB strain (Fig.3.5). Remarkably, at 45°C, inhibition of ClpB with DBeQ is more detrimental for bacterial survival than a lack of ClpB in the ΔclpB strain. Thus, the phenotype of chemical inhibition of ClpB transcends a loss of function, at least under moderate stress. Toxicity of ClpB in bacteria and its ortholog Hsp104 in yeast has been observed for "hyperactive" protein variants with mutations within the coiled-coil middle domain (55, 59, 60). However, unlike the DBeQ treated ClpB, the toxic hyperactive variants display elevated ATPase and disaggregase activities.

The apparent toxic gain of function of the DBeQ inhibited ClpB in bacterial cells could be blamed on DBeQ induced aggregation of ClpB. However, we have not detected a propensity of ClpB to misfold and/or aggregate in the presence of DBeQ in vitro (Appendix A Fig. A.5), which is also supported by a lack of DBeQ effects on the basal ATPase activity of ClpB (Fig.3.1 and Appendix A Fig. A.6). Moreover, the survival of dnaK756 *E. coli* upon treatment with DBeQ does not support a premise that the compound's toxicity in cells is linked to misfolding or aggregation of its main target, ClpB, because, in such a case, a viability rescue should not be expected in a strain with a nonproductive DnaK chaperone (43).

Chaperone-deficient strains of *E. coli*, ΔclpB, and dnaK103 survive moderate heat stress of 45°C, thanks to other protein quality control factors: chaperones and proteases whose highly promiscuous and redundant activities are sufficient for maintaining proteostasis in those strains (Fig.3.5). Because DBeQ apparently targets ClpB and no other chaperones in *E. coli*, the ΔclpB strain is therefore resistant to DBeQ under moderate stress conditions.

Physiologically, DnaK recruits ClpB to the aggregates and hands the substrates over to ClpB for disaggregation. In wt *E. coli* cells, a recruitment of ClpB to the aggregated proteins located at the cell poles strictly depends on DnaK (42). A striking polar localization of ClpB in dnaK103 cells in the presence of DBeQ suggests that the compound allows ClpB to overcome a deficiency of its recruiter cochaperone (Fig.3.6). However, ClpB interactions with protein aggregates become nonproductive in the presence of DBeQ (Fig.3.2). A nonproductive binding of ClpB to protein aggregates may become dominant and could suppress viability of *E. coli* because it may hinder access of other protein quality control factors to their aggregated substrates and irreversibly disturb cellular proteostasis. The above explanation of the DBeQ induced phenotype in *E. coli* is corroborated by an unexpected rescue of cellular viability in the dnaK756 strain

(Fig.3.5). The DnaK756 variant binds to aggregates but cannot release them (43). Apparently, a substrate trapping capability of DnaK756 counteracts the nonproductive interactions of ClpB with the aggregates in the presence of DBeQ and is sufficient to preserve cellular viability. Thus, the apparent gain-of-function effect of DBeQ may be due to disturbing the balance of cellular proteostasis by stimulating nonproductive interactions of ClpB that suppress effectiveness of the remaining chaperones. Importantly, regardless of its exact mode of function in bacterial cells, DBeQ is a potent and highly selective molecular probe that targets and disrupts protein quality control in *E. coli*.

Molecular chaperones have not been previously explored as targets for novel antimicrobials. In this work, we demonstrated that the AAA+ disaggregase ClpB can be selectively targeted with a small-molecule ligand in bacterial cells and that such a treatment could produce a loss of bacterial viability. This result is significant because mammalian cells do not contain orthologues of ClpB (5), whereas many pathogenic microorganisms require the ClpB activity for infectivity and survival. We have also shown that owing to a complexity of the cellular protein control machinery, understanding the mechanism of action of chemical inhibitors and the cellular phenotypes they produce requires multiple orthogonal biochemical and biological tests, such as those used in this study.

3.6 Acknowledgments.

The authors thank Bernd Bukau, Axel Mogk, Sabina Kedzierska Mieszkowska, and Francois Baneyx for kindly providing the reagents and Joel Sanneman from the Kansas State University Confocal Microscopy and Microfluorometry Core for help with confocal microscopy imaging.

3.7 Author contributions.

M. Z. and F. J. S. conceptualization, P. G.,C. B. R., S. S., A. R., and B. V. G. investigation; P. G., C. B. R., S. S.,A. R., B. V. G., and M.Z. data analysis; M.Z. writing original manuscript draft; P. G., C. B. R., S. S., A. R., B. V. G., F. J. S., and M.Z. writing manuscript review and editing.

3.8 Funding and additional information.

This study was supported by grants from the National Institutes of Health, AI121366 and AI141586 (to M. Z.) and GM121511 (to B. V. G.), and by the Kansas Agricultural Experiment Station (contribution 19-063-J). The con- tent is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

3.9 Conflict of interest.

The authors declare that they have no conflicts of interest with the contents of this article.

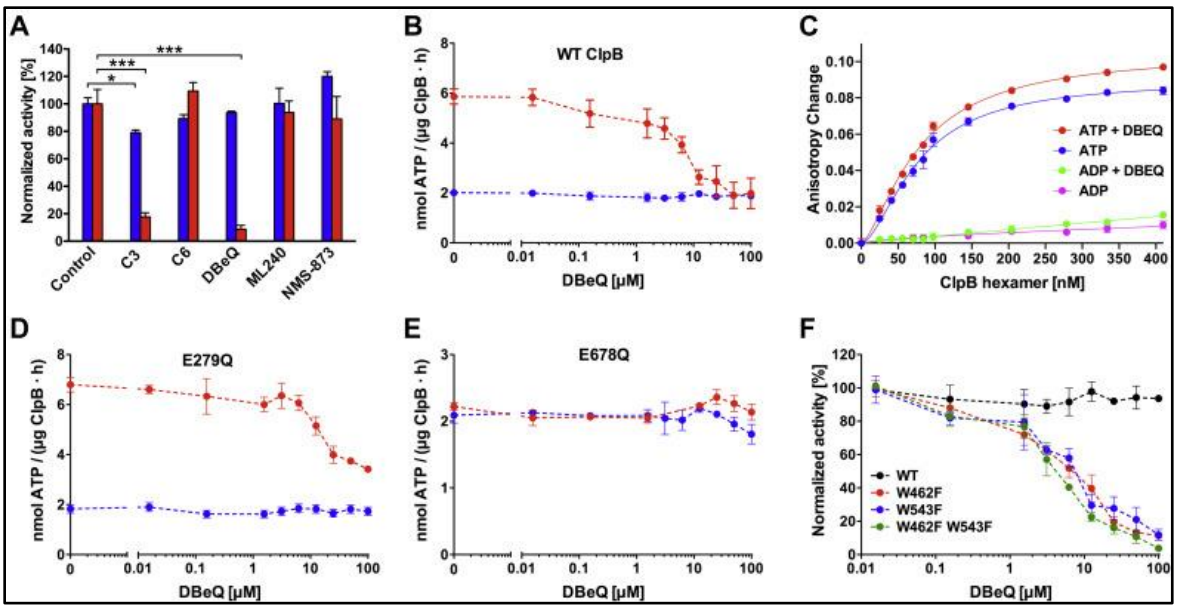


Figure 3.1. DBeQ inhibits the casein-activated ATPase activity of ClpB.

Effects of the inhibitor candidates on the ATPase activity of ClpB at 37 °C. *A*, the ClpB ATPase activity was determined in the absence (*blue*) and presence (*red*) of the pseudo substrate κ-casein without (control) and with 100-µM inhibitor candidates (see Fig. S1). The average activity values from three measurements normalized to the control are shown with SDs; *$p < 0.05$; ***$p < 0.001$. *B*, *D*, *and E*, the ClpB ATPase activity was measured in the absence (*blue*) and presence (*red*) of κ-casein for wt ClpB (*B*), E279Q (*D*), and E678Q (*E*). *C*, FITC fluorescence anisotropy changes were determined on titrating ClpB E279Q/E678Q into 100 nM FITC casein with 5 mM ATP or ADP and with or without 80 µM DBeQ. The average values from two independent measurements are shown with SDs. *Solid lines* show the results of fitting of the binding isotherm equation (see Experimental procedures), which gave the following values for ATP with DBeQ K_d, 78.5 nM; saturation signal, 0.11; Hill number, 1.48; r^2, 0.997 and for ATP without DBeQ K_d, 77.3 nM; saturation signal, 0.09; Hill number, 1.68; r^2, 0.994. *F*, a comparison of the basal ATPase activity of wt ClpB (*black*) with W462F (*red*), W543F (*blue*), and the double mutant W462F/W543F (*green*). The average values from three measurements are shown with

SDs. DBeQ, N^2, N^4-dibenzylquinazoline-2, 4-diamine. Experiments 3.10.1 A,D,E,F were performed by Dr Przemyslaw Glaza.

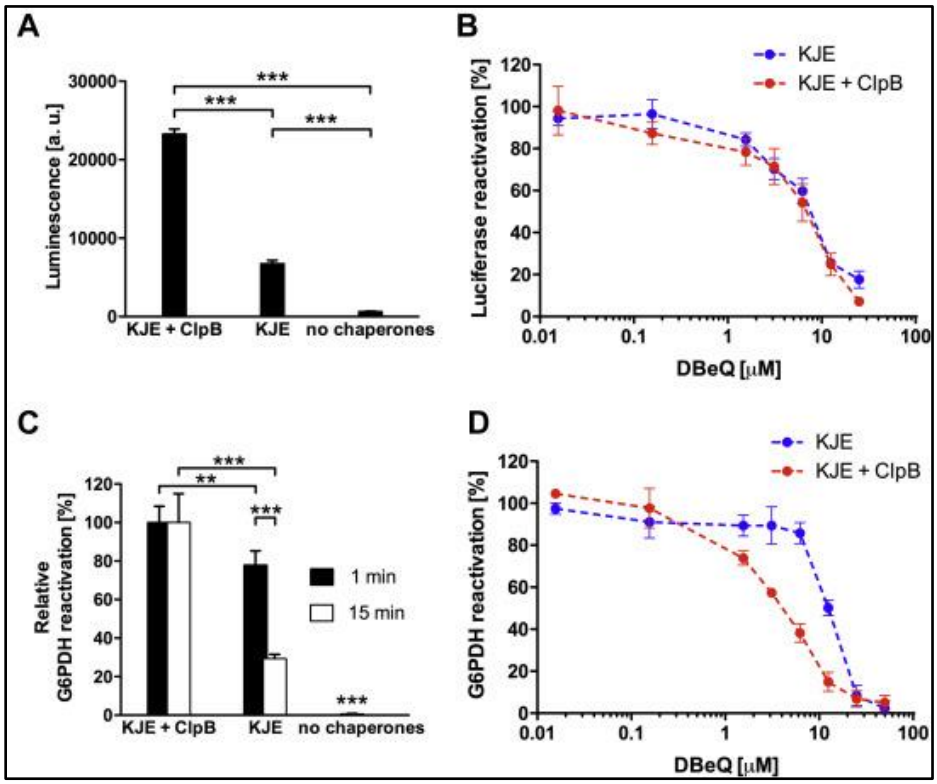


Figure 3.2. Effects of DBeQ on the reactivation of protein aggregates mediated by ClpB and DnaK–DnaJ–GrpE (KJE).

A, the activity of the thermally aggregated firefly luciferase after 20 min of reactivation at 25 °C in the presence of KJE, KJE + ClpB, or without the chaperones. The average values from three measurements are shown with SDs; ***$p < 0.001$. B, the effects of DBeQ on the reactivation of aggregated luciferase in the presence of KJE (blue) or KJE + ClpB (red). The average values from three measurements are shown with SDs. C, the activity of the aggregated glucose-6-phosphate dehydrogenase (G6PDH) after 60 min of reactivation at 30 °C in the presence of KJE, KJE + ClpB, or without the chaperones. The aggregation of G6PDH was arrested after 1 min (black

bars) or 15 min (open bars) to produce two samples with a different average size of aggregates. The average values from three measurements are shown with SDs; ***$p < 0.001$; **$p < 0.01$. D, the effects of DBeQ on the reactivation of aggregated G6PDH. G6PDH samples aggregated for 1 min were used as substrates for KJE (blue). G6PDH samples aggregated for 15 min were used as substrates for KJE + ClpB (red). The average values from three measurements are shown with SDs. The data in panels C and D were normalized to the activity levels measured in the absence of DBeQ. DBeQ, N^2, N^4 -dibenzylquinazoline-2, 4-diamine. Experiments 3.10.2 A,B,C,D were performed by Dr Przemyslaw Glaza.

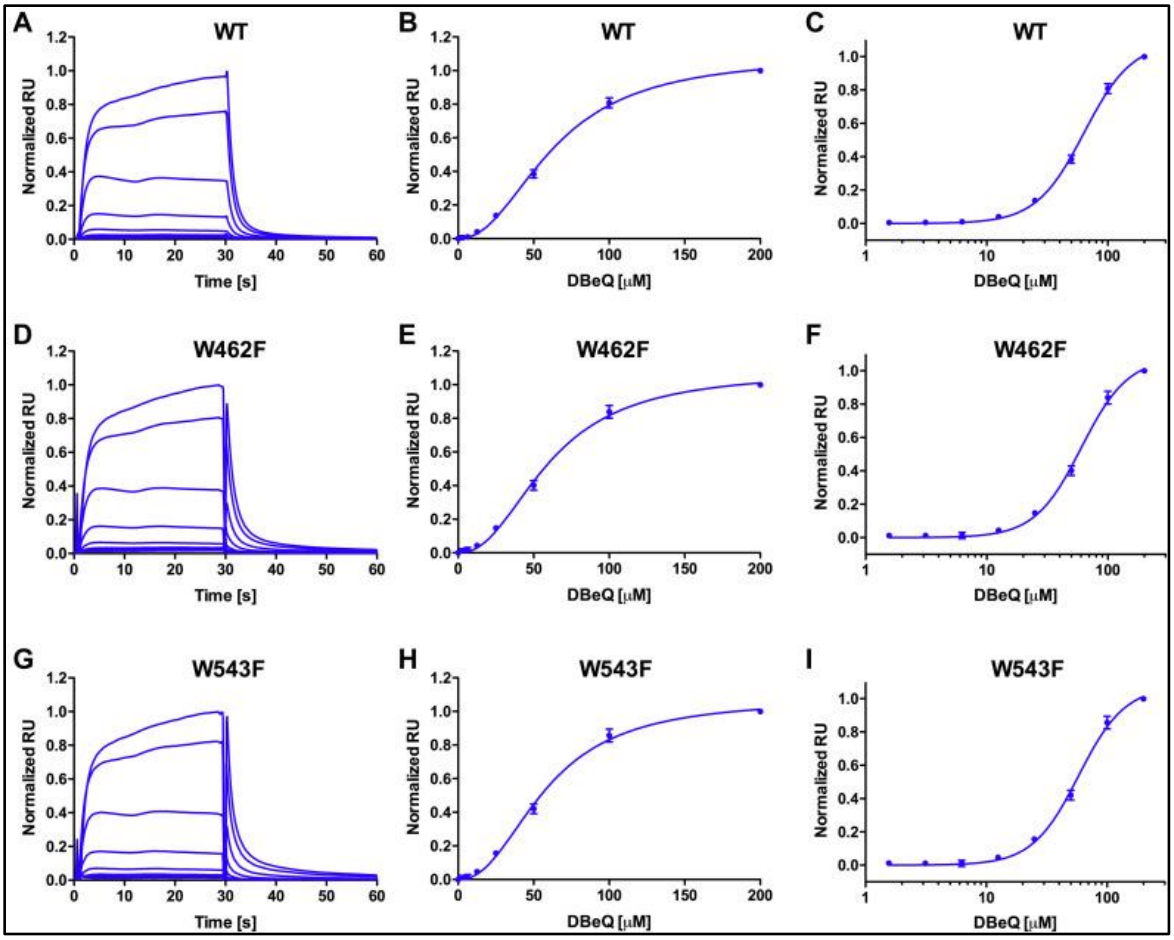


Figure 3.3. Surface plasmon resonance (SPR) analysis of the binding of DBeQ to wt ClpB (A-C), W462F (D-F), and W543F (G-I).

A, D, and G, representative SPR sensograms for three ClpB variants. Each solid line represents a sensogram obtained with a given DBeQ concentration; a higher DBeQ concentration produces a higher SPR response. DBeQ binding isotherms are shown with the linear ligand concentration scale (B, E, and H) or the logarithmic scale (C, F, and I). Shown are the averages with SDs from three repeated experiments. Solid lines in the panels B, C, E, F, H, and I represent the fits of the cooperative binding model with the parameters listed in appendix A , Table A.1. DBeQ, N^2, N^4-dibenzylquinazoline-2, 4-diamine. Experiments 3.10.3 A,B,C,D,E,F,G,H,I were performed with Dr Brian V. Geisbrecht, Dr. Przemyslaw Glaza.

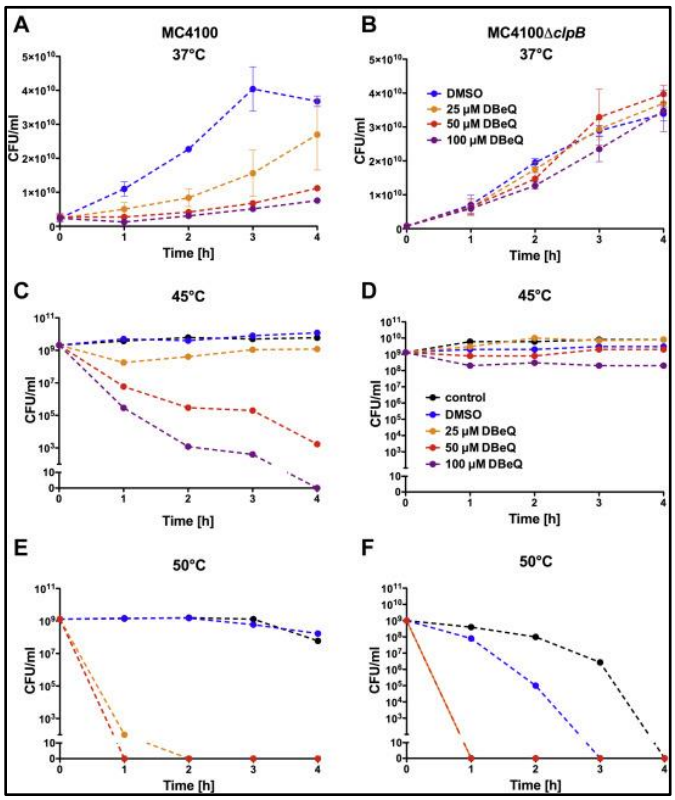


Figure 3.4. Effects of DBeQ on the growth and viability of *E. coli* and its ClpB deficient strain.

Growth of *E. coli* MC4100 (A) and MC4100ΔclpB (B) at 37 °C in the presence of an increasing concentration of DBeQ. Shown are the averages with SDs from three independent experiments. C–F, survival time courses for MC4100 (C and E) and MC4100ΔclpB (D and F) at 45 °C (C and D) and 50 °C (E and F) in the presence of an increasing concentration of DBeQ. Shown are representative results from 3 repeated experiments, including those in Figure 3.5. The legend in panel B applies also to panel A. The legend in panel D applies also to panels C, E, and F. DBeQ, N^2, N^4 -dibenzylquinazoline-2, 4-diamine. Experiments 3.10.4 A,B,C,D,E,F were performed by Dr Przemyslaw Glaza.

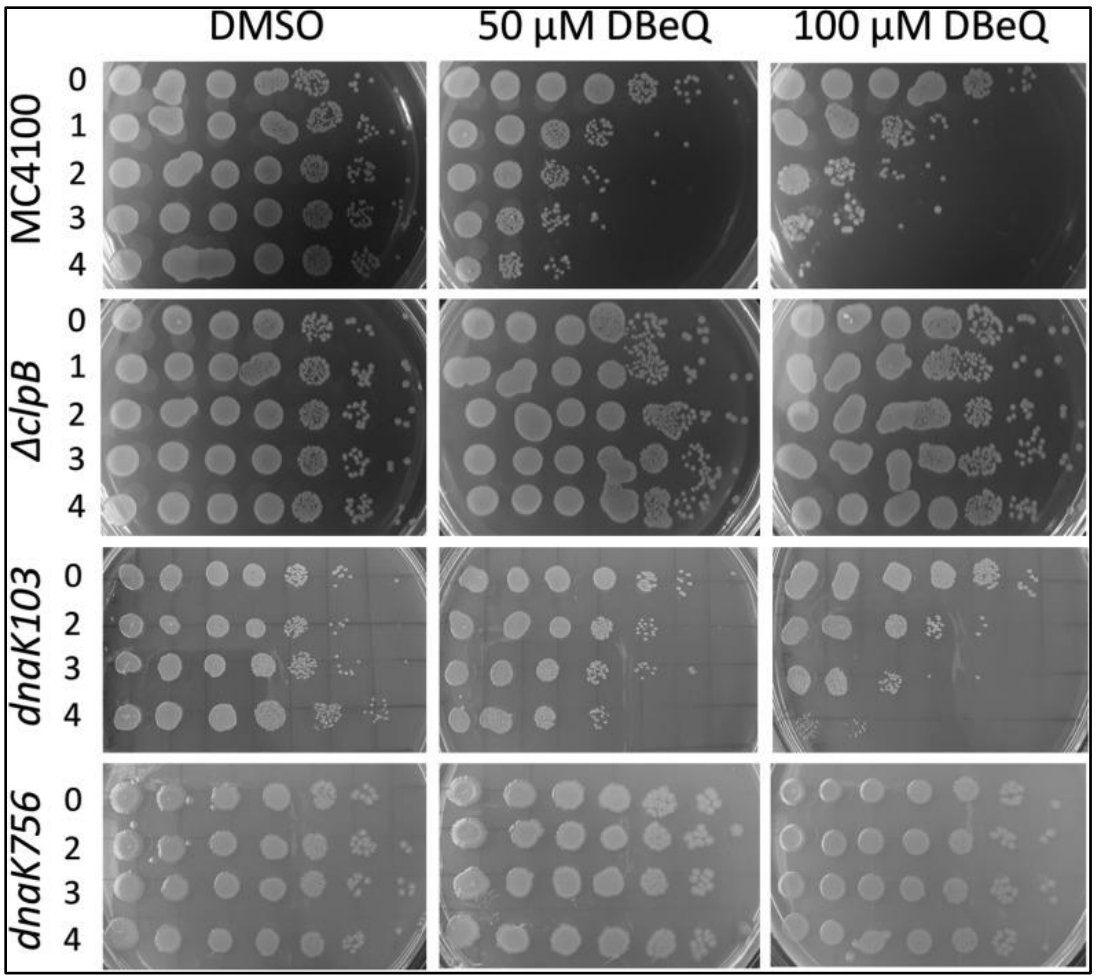


Figure 3.5. Viability of *E. coli* strains after exposure to increasing concentrations of DBeQ at 45 °C.

The indicated strains of *E. coli* were incubated at 45 °C in the presence of DBeQ or with only DMSO for the period of time indicated on the left (in hours) and then spotted on agar plates and incubated overnight at 37 °C. Each spot on the agar plates represents a viable culture after a 10 fold serial dilution (from left to right). Shown are representative results from 3 repeated experiments. DBeQ, N^2, N^4 -dibenzylquinazoline-2, 4-diamine. DMSO, dimethyl sulfoxide.

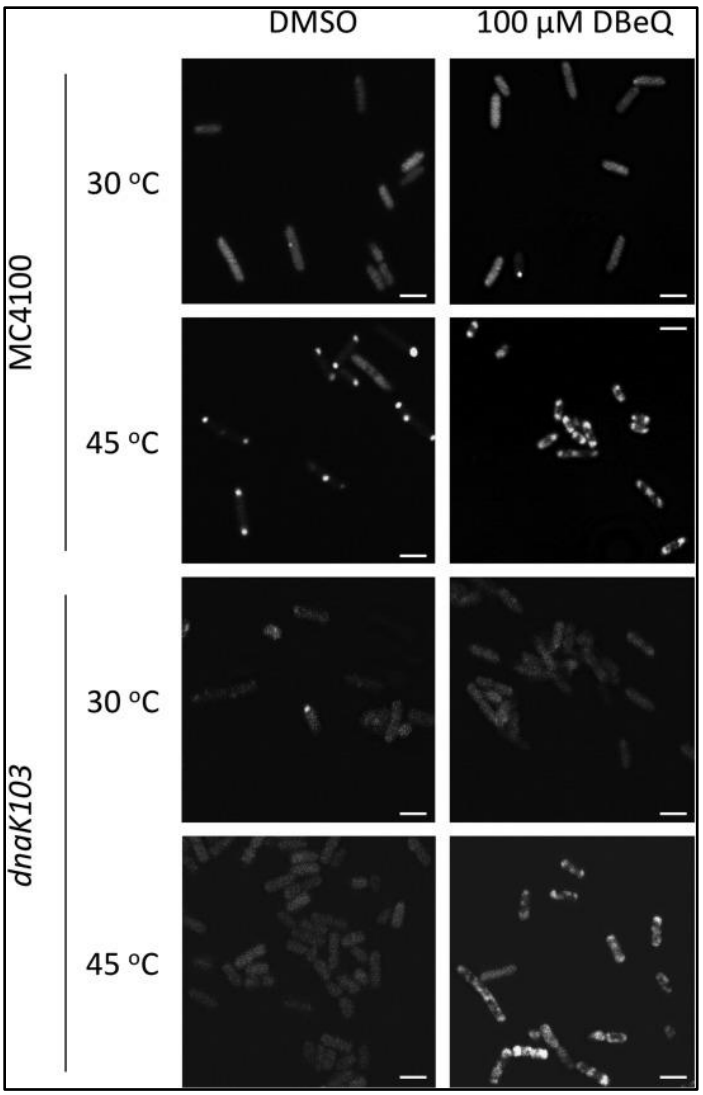


Figure 3.6. Localization of ClpB-YFP in *E. coli*.

ClpB-YFP was expressed in the indicated *Escherichia coli* strains. The cells were grown at 30 °C and then shifted to 45 °C for 30 min (see Experimental Procedures). The images show YFP fluorescence signal. Representative images from 3 independent experiments are shown. The white bar in each panel corresponds to 2 µm.

3.11 References.

1. Stewart, G. R., and Young, D. B. (2004) Heat-shock proteins and the host-pathogen interaction during bacterial infection. Curr. Opin. Immunol. 16, 506–510.

2. Henderson, B., Allan, E., and Coates, A. R. (2006) Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. Infect. Immun. 74, 3693–3706.

3. Neckers, L., and Tatu, U. (2008) Molecular chaperones in pathogen virulence: emerging new targets for therapy. Cell Host Microbe. 4, 519–527.

4. Zolkiewski, M., Zhang, T., and Nagy, M. (2012) Aggregate reactivation mediated by the Hsp100 chaperones. Arch. Biochem. Biophys. 520, 1–6.

5. Erives, A. J., and Fassler, J. S. (2015) Metabolic and chaperone gene loss marks the origin of animals: evidence for Hsp104 and Hsp78 chaperones sharing mitochondrial enzymes as clients. PLoS One 10, e0117192.

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

7. 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, 28083–28086.

8. Goloubinoff, P., Mogk, A., Zvi, A. P., Tomoyasu, T., and Bukau, B. (1999) Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. Proc. Natl. Acad. Sci. U. S. A. 96, 13732–13737.

9. de Oliveira, N. E., Abranches, J., Gaca, A. O., Laport, M. S., Damaso, C. R., Bastos Mdo, C., Lemos, J. A., and Giambiagi-deMarval, M. (2011) clpB, a class III heat-shock gene regulated by CtsR, is involved in thermotol-erance and virulence of *Enterococcus faecalis*. Microbiology 157, 656–665.

10. Frees, D., Chastanet, A., Qazi, S., Sorensen, K., Hill, P., Msadek, T., and Ingmer, H. (2004) Clp ATPases are required for stress tolerance, intra-cellular replication and biofilm formation in *Staphylococcus aureus*. Mol. Microbiol. 54, 1445–1462.

11. Lehoux, D. E., Sanschagrin, F., and Levesque, R. C. (2002) Identification of in vivo essential genes from *Pseudomonas aeruginosa* by PCR-based signature tagged mutagenesis. FEMS Microbiol. Lett. 210, 73–80.

12. Estorninho, M., Smith, H., Thole, J., Harders-Westerveen, J., Kierzek, A., Butler, R. E., Neyrolles, O., and Stewart, G. R. (2010) ClgR regulation of chaperone and protease systems is

essential for *Mycobacterium tuberculosis* parasitism of the macrophage. Microbiology 156, 3445–3455.

13. Vaubourgeix, J., Lin, G., Dhar, N., Chenouard, N., Jiang, X., Botella, H., Lupoli, T., Mariani, O., Yang, G., Ouerfelli, O., Unser, M., Schnappinger, D., McKinney, J., and Nathan, C. (2015) Stressed mycobacteria use the chaperone ClpB to sequester irreversibly oxidized proteins asymmetrically within and between cells. Cell Host Microbe. 17, 178–190.

14. Kannan, T. R., Musatovova, O., Gowda, P., and Baseman, J. B. (2008) Characterization of a unique ClpB protein of *Mycoplasma pneumoniae* and its impact on growth. Infect. Immun. 76, 5082–5092.

15. Turner, A. K., Lovell, M. A., Hulme, S. D., Zhang-Barber, L., and Barrow,P. A.(1998) Identification of *Salmonella typhimurium* genes required for colonization of the chicken alimentary tract and for virulence in newly hatched chicks. Infect. Immun. 66, 2099–2106.

16. Pieper, R., Zhang, Q., Parmar, P. P., Huang, S. T., Clark, D. J., Alami, H., Donohue-Rolfe, A., Fleischmann, R. D., Peterson, S. N., and Tzipori, S. (2009) The *Shigella dysenteriae* serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins. Proteomics. 9, 5029–5045.

17. Kuntumalla, S., Zhang, Q., Braisted, J. C., Fleischmann, R. D., Peterson, S. N., Donohue-Rolfe, A., Tzipori, S., and Pieper, R. (2011) In vivo versus in vitro protein abundance analysis of *Shigella dysenteriae* type 1 reveals changes in the expression of proteins involved in virulence, stress and energy metabolism. BMC Microbiol. 11, 147.

18. Capestany, C. A., Tribble, G. D., Maeda, K., Demuth, D. R., and Lamont, R.J. (2008) Role of the Clp system in stress tolerance, biofilm formation, and intracellular invasion in *Porphyromonas gingivalis*. J. Bacteriol. 190, 1436–1446.

19. Meibom, K. L., Dubail, I., Dupuis, M., Barel, M., Lenco, J., Stulik, J., Golovliov, I., Sjostedt, A., and Charbit, A. (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, 1384–1401.

20. Lourdault, K., Cerqueira, G. M., Wunder, E. A., Jr., and Picardeau, M. (2011) Inactivation of ClpB in the pathogen *Leptospira interrogans* reduces virulence and resistance to stress conditions. Infect. Immun. 79, 3711–3717.

21. Krobitsch, S., and Clos, J. (1999) A novel role for 100 kD heat shock proteins in the parasite *Leishmania donovani*. Cell Stress Chaperones. 4, 191–198.

22. Zhang, T., Kedzierska-Mieszkowska, S., Liu, H., Cheng, C., Ganta, R. R., and Zolkiewski, M. (2013) Aggregate-reactivation activity of the molecular chaperone ClpB from *Ehrlichia chaffeensis*. PLoS One 8, e62454.

23. Kuczynska-Wisnik, D., Cheng, C., Ganta, R. R., and Zolkiewski, M. (2017) Protein aggregation in *Ehrlichia chaffeensis* during infection of mammalian cells. FEMS Microbiol. Lett 364, fnx059.

24. Kalanon, M., and McFadden, G. I. (2010) Malaria, *Plasmodium falciparum* and its apicoplast. Biochem. Soc. Trans. 38, 775–782.

25. Ngansop, F., Li, H., Zolkiewska, A., and Zolkiewski, M. (2013) Biochemical characterization of the apicoplast-targeted AAA+ ATPase ClpB from *Plasmodium falciparum*. Biochem. Biophys. Res. Commun. 439, 191–195.

26. de Koning-Ward, T. F., Gilson, P. R., Boddey, J. A., Rug, M., Smith, B. J., Papenfuss, A. T., Sanders, P. R., Lundie, R. J., Maier, A. G., Cowman, A. F., and Crabb, B. S. (2009) A newly discovered protein export machine in malaria parasites. Nature 459, 945–949.

27. Beck, J. R., Muralidharan, V., Oksman, A., and Goldberg, D. E. (2014) PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. Nature 511, 592–595.

28. Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res. 9, 27–43.

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

30. Weibezahn, J., Tessarz, P., Schlieker, C., Zahn, R., Maglica, Z., Lee, S., Zentgraf, H., Weber-Ban, E. U., Dougan, D. A., Tsai, F. T., Mogk, A., and Bukau, B. (2004) Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. Cell 119, 653–665.

31. Avellaneda, M. J., Franke, K. B., Sunderlikova, V., Bukau, B., Mogk, A., and Tans, S. J. (2020) Processive extrusion of polypeptide loops by a Hsp100 disaggregase. Nature 578, 317–320.

32. Chou, T. F., Brown, S. J., Minond, D., Nordin, B. E., Li, K., Jones, A. C., Chase, P., Porubsky, P. R., Stoltz, B. M., Schoenen, F. J., Patricelli, M. P., Hodder, P., Rosen, H., and Deshaies, R. J. (2011) Reversible inhibitor of p97, DBeQ, impairs both ubiquitin-dependent and autophagic protein clearance pathways. Proc. Natl. Acad. Sci. U. S. A. 108, 4834–4839.

33. Anderson, D. J., Le Moigne, R., Djakovic, S., Kumar, B., Rice, J., Wong, S., Wang, J., Yao, B., Valle, E., Kiss von Soly, S., Madriaga, A., Soriano, F., Menon, M. K., Wu, Z. Y., Kampmann, M., et al. (2015) Targeting the AAA ATPase p97 as an approach to treat cancer through disruption of protein homeostasis. Cancer Cell 28, 653–665.

34. Vekaria, P. H., Home, T., Weir, S., Schoenen, F. J., and Rao, R. (2016) Targeting p97 to disrupt protein homeostasis in cancer. Front. Oncol. 6, 181.

35. Chou, T. F., Li, K., Frankowski, K. J., Schoenen, F. J., and Deshaies, R. J. (2013) Structure-activity relationship study reveals ML240 and ML241 as potent and selective inhibitors of p97 ATPase. ChemMedChem. 8, 297–312.

36. Magnaghi, P., D'Alessio, R., Valsasina, B., Avanzi, N., Rizzi, S., Asa, D., Gasparri, F., Cozzi, L., Cucchi, U., Orrenius, C., Polucci, P., Ballinari, D., Perrera, C., Leone, A., Cervi, G., et al. (2013) Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. Nat. Chem. Biol. 9, 548–556.

37. Martin, I., Underhaug, J., Celaya, G., Moro, F., Teigen, K., Martinez, A., and Muga, A. (2013) Screening and evaluation of small organic molecules as ClpB inhibitors and potential antimicrobials. J. Med. Chem. 56, 7177–7189.

38. Akoev, V., Gogol, E. P., Barnett, M. E., and Zolkiewski, M. (2004) Nucleotide induced switch in oligomerization of the AAA+ ATPase ClpB. Protein Sci. 13, 567–574.

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

40. Squires, C. L., Pedersen, S., Ross, B. M., and Squires, C. (1991) ClpB is the *Escherichia coli* heat shock protein F84.1. J. Bacteriol 173, 4254–4262.

41. Spence, J., Cegielska, A., and Georgopoulos, C. (1990) Role of *Escherichia coli* heat shock proteins DnaK and HtpG (C62.5) in response to nutritional deprivation. J. Bacteriol. 172, 7157–7166.

42. Winkler, J., Tyedmers, J., Bukau, B., and Mogk, A. (2012) Hsp70 targets Hsp100 chaperones to substrates for protein disaggregation and prion fragmentation. J. Cell Biol. 198, 387–404.

43. Buchberger, A., Gassler, C. S., Buttner, M., McMacken, R., and Bukau, B. (1999) Functional defects of the DnaK756 mutant chaperone of *Escherichia coli* indicate distinct roles for amino- and carboxyl-terminal residues in substrate and co-chaperone interaction and interdomain communication. J. Biol. Chem. 274, 38017–38026.

44. Winkler, J., Seybert, A., Konig, L., Pruggnaller, S., Haselmann, U., Sourjik, V., Weiss, M., Frangakis, A. S., Mogk, A., and Bukau, B. (2010) Quantitative and spatiotemporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. EMBO J. 29, 910–923.

45. Chapman, E., Maksim, N., de la Cruz, F., and La Clair, J. J. (2015) Inhibitors of the AAA+ chaperone p97. Molecules. 20, 3027–3049.

46. DeGraw, J. I., Brown, V. H., and Colwell, W. T. (1974) Potential anti-leprotic agents. Inhibition of mycobacterial dihydrofolic reductase by 2,4-diamino-5-methyl-6-alkylquinazolines. J. Med. Chem. 17, 762–764.

47. Van Horn, K. S., Burda, W. N., Fleeman, R., Shaw, L. N., and Manetsch, R. (2014) Antibacterial activity of a series of N2,N4-disubstituted quinazo-line-2,4-diamines. J. Med. Chem. 57, 3075–3093.

48. Harbut, M. B., Patel, B. A., Yeung, B. K., McNamara, C. W., Bright, A. T., Ballard, J., Supek, F., Golde, T. E., Winzeler, E. A., Diagana, T. T., and Greenbaum, D. C. (2012) Targeting the ERAD pathway via inhibition of signal peptide peptidase for antiparasitic therapeutic design. Proc. Natl. Acad. Sci. U. S. A. 109, 21486–21491.

49. Van Horn, K. S., Zhu, X., Pandharkar, T., Yang, S., Vesely, B., Vanaerschot, M., Dujardin, J. C., Rijal, S., Kyle, D. E., Wang, M. Z., Werbovetz, K. A., and Manetsch, R. (2014) Antileishmanial activity of a series of N(2), N(4)-disubstituted quinazoline-2,4-diamines. J. Med. Chem. 57, 5141–5156.

50. Hoskins, J. R., Doyle, S. M., and Wickner, S. (2009) Coupling ATP utilization to protein remodeling by ClpB, a hexameric AAA+ protein. Proc. Natl. Acad. Sci. U. S. A. 106, 22233–22238.

51. Schlieker, C., Weibezahn, J., Patzelt, H., Tessarz, P., Strub, C., Zeth, K., Erbse, A., Schneider Mergener, J., Chin, J. W., Schultz, P. G., Bukau, B., and Mogk, A. (2004) Substrate recognition by the AAA+ chaperone ClpB. Nat. Struct. Mol. Biol. 11, 607–615.

52. Barnett, M. E., Nagy, M., Kedzierska, S., and Zolkiewski, M. (2005) The amino terminal domain of ClpB supports binding to strongly aggregated proteins. J. Biol. Chem. 280, 34940–34945.

53. Rizo, A. N., Lin, J., Gates, S. N., Tse, E., Bart, S. M., Castellano, L. M., DiMaio, F., Shorter, J., and Southworth, D. R. (2019) Structural basis for substrate gripping and translocation by the ClpB AAA+ disaggregase. Nat. Commun. 10, 2393.

54. Haslberger, T., Weibezahn, J., Zahn, R., Lee, S., Tsai, F. T., Bukau, B., and Mogk, A. (2007) M domains couple the ClpB threading motor with the DnaK chaperone activity. Mol. Cell 25, 247–260.

55. Oguchi, Y., Kummer, E., Seyffer, F., Berynskyy, M., Anstett, B., Zahn, R., Wade, R. C., Mogk, A., and Bukau, B. (2012) A tightly regulated molecular toggle controls AAA+ disaggregase. Nat. Struct. Mol. Biol. 19, 1338–1346.

56. Sen, M., Maillard, R. A., Nyquist, K., Rodriguez-Aliaga, P., Presse, S., Martin, A., and Bustamante, C. (2013) The ClpXP protease unfolds substrates using a constant rate of pulling but different gears. Cell 155, 636–646.

57. Rodriguez-Aliaga, P., Ramirez, L., Kim, F., Bustamante, C., and Martin, A. (2016) Substrate translocating loops regulate mechanochemical coupling and power production in AAA+ protease ClpXP. Nat. Struct. Mol. Biol. 23, 974–981.

58. Jones, A. M., Westwood, I. M., Osborne, J. D., Matthews, T. P., Cheeseman, M. D., Rowlands, M. G., Jeganathan, F., Burke, R., Lee, D., Kadi, N., Liu, M., Richards, M., McAndrew, C., Yahya, N., Dobson, S. E.,et al. (2016) A fragment based approach applied to a highly flexible target: insights and challenges towards the inhibition of HSP70 isoforms. Sci. Rep. 6, 34701.

59. Lipinska, N., Zietkiewicz, S., Sobczak, A., Jurczyk, A., Potocki, W., Morawiec, E., Wawrzycka, A., Gumowski, K., Slusarz, M., Rodziewicz-Motowidlo, S., Chrusciel, E., and Liberek, K. (2013) Disruption of ionic interactions between the nucleotide binding domain 1 (NBD1) and middle (M) domain in Hsp100 disaggregase unleashes toxic hyperactivity and partial independence from Hsp70. J. Biol. Chem. 288, 2857–2869.

60. Chamera, T., Klosowska, A., Janta, A., Wyszkowski, H., Obuchowski, I., Gumowski, K., and Liberek, K. (2019) Selective hsp70 dependent docking of Hsp104 to protein aggregates protects the cell from the toxicity of the disaggregase. J. Mol. Biol. 431, 2180–2196.

61. Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177, 4121–4130.

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

63. Nagy, M., Akoev, V., and Zolkiewski, M. (2006) Domain stability in the AAA+ ATPase ClpB from *Escherichia coli*. Arch. Biochem. Biophys. 453, 63–69.

64. Zhang, T., Ploetz, E. A., Nagy, M., Doyle, S. M., Wickner, S., Smith, P. E., and Zolkiewski, M. (2012) Flexible connection of the N-terminal domain in ClpB modulates substrate binding and the aggregate reactivation efficiency. Proteins 80, 2758–2768.

65. Ranaweera, C. B., Glaza, P., Yang, T., and Zolkiewski, M. (2018) Inter-action of substrate-mimicking peptides with the AAA+ ATPase ClpB from *Escherichia coli*. Arch. Biochem. Biophys. 655, 12–17.

66. Georgopoulos, C. P. (1977) A new bacterial gene (groPC) which affects lambda DNA replication. Mol. Gen. Genet 151, 35–39.

67. Hess, H. H., and Derr, J. E. (1975) Assay of inorganic and organic phosphorus in the 0.1-5 nanomole range. Anal. Biochem 63, 607–613.

68. Lanzetta, P. A., Alvarez, L. J., Reinach, P. S., and Candia, O. A. (1979) An improved assay for nanomole amounts of inorganic phosphate. Anal. Biochem. 100, 95–97.

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

70. Garcia, B. L., Skaff, D. A., Chatterjee, A., Hanning, A., Walker, J. K., Wyckoff, G. J., and Geisbrecht, B. V. (2017) Identification of C3b-binding small molecule complement inhibitors using cheminformatics. J. Immunol. 198, 3705–3718.

71. 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. Anal. Biochem. 203, 295–301.

72. LeBel, R. G., and Goring, D. A. I. (1962) Density, viscosity, refractive index, and hygroscopicity of mixtures of water and dimethyl sulfoxide. J. Chem. Eng. Data 7, 100–101.

73. Weibezahn, J., Schlieker, C., Bukau, B., and Mogk, A. (2003) Characterization of a trap mutant of the AAA+ chaperone ClpB. J. Biol. Chem. 278, 32608–32617.

Chapter 4 - Protein aggregation in *Anaplasma phagocytophilum* during infection of mammalian cells & Biochemical characterization of ClpB and DnaK from *Anaplasma phagocytophilum.*

Data presented in this chapter will be used for manuscript preparation later.

4.1 *Anaplasma* and Anaplasmosis.

Anaplasma phagocytophilum is an obligate, Gram negative intracellular pathogen measuring 0.5 to 1.0 μm in size. Importantly, *A.phagocytophilum* lacks peptidoglycan and lipopolysaccharide biosynthetic machinery (Lin and Rikihisa 2003) . *Anaplasma phagocytophilum* is one of the most important tick-borne bacteria which causes infections in veterinary animals and humans (Dumler, Choi et al. 2005). The disease caused by *Anaplasma phagocytophilum* is known as Anaplasmosis. Anaplasma bacteria are transmitted to animals and humans by tick bites mainly from *Ixodes scapularis* (black legged ticks) and *Ixodes pacificus* (western black legged tick) ticks. Human Granulocytic Anaplasmosis (HGA), an infection of neutrophils caused by *Anaplasma phagocytophilum* was first identified only in 1990 despite the fact the pathogen has been identified in veterinary practice since 1932. Now, it is estimated that at least one billion people of worldwide population are at a risk of contracting infections from obligate intracellular pathogens (McClure, Chávez et al. 2017). New England, upper Midwest, northern California, and many parts of Europe have high number of HGA incidents because *Ixodes* tick bites are quite frequent (Dumler, Choi et al. 2005). Even though Anaplasmosis has already become a major veterinary and public health concern, there is a limited amount of information available about bacterial physiology, pathogen and host interaction, and bacterial virulence factors. It has been observed that genetic manipulation

of obligate intracellular pathogens is extremely difficult as they require a host cell for proliferation and only a handful of genetic tools have been developed to date (Cheng, Nair et al. 2013, Atif 2015).

4.2 Life cycle of *Anaplasma phagocytophilum*.

Generally, infected ticks would transmit *Anaplasma phagocytophilum* to mammals during a blood meal (figure 4.1) taken from mammals. The infected mammals include raccoons (*Procyon lotor*), gray foxes (*Urocyon cinereoargenteus*), white-tailed deer (*Odocoileus virginianus*), white-footed mice (*Peromyscus leucopus*) and gray squirrels (*Sciurus carolinensis*). At the same time, if these mammals are already infected with *Anaplasma*, then an uninfected tick could acquire *Anaplasma* during a blood meal as well. *Anaplasma* bacteria are found from later larval, nymph to the adult stage of a tick life cycle. So far there is no evidence to indicate that bacteria are transmitted from infected adults to tick eggs or from eggs to an immediate larval stage. As adults will lay eggs in the environment and as the eggs hatch, larvae will emerge from the hatched eggs. In order to survive, these larvae must get a blood meal from a suitable host for survival and development into a nymph. Once they acquire a blood meal from a host, these larvae will drop off from the earlier host and will prepare themselves to become nymphs. Humans are considered a dead-end host for *Anaplasma phagocytophilum* (Rikihisa 2011).

Once eukaryotic cells are infected with *Anaplasma phagocytophilum*, bacteria undergo cyclical rounds of infectious and replicative forms of *Anaplasma phagocytophilum (McClure, Chávez et al. 2017)*. The infectious form is known as Dense Core (DC) cells and the replicative form (noninfectious form) is known as Reticulate Cells (RC). When *Anaplasma phagocytophilum* encounters a suitable eukaryotic host cell, Dense Core cells get internalized (step 1 in figure 4.2). Once internalized, the DC cells differentiate into Reticulate Cells (step 2 in figure 4.2) in a vacuole

and the bacteria in these inclusions are known as morula. The RC cells multiply by binary fission inside the vacuoles (step 3 in figure 4.2). Once the RC cells are multiplied, they differentiate back to DC cells and exit the host cells by inducing lysis (step 4 in figure 4.2).

To invade and replicate inside host cells, *Anaplasma* must first attach to the host cell. *Anaplasma phagocytophilum* has three main adhesin components which are upregulated during host cell invasion. These adhesin molecules are the *Anaplasma* cell membrane components found on Dense Core (DC) cell membrane and are transcriptionally upregulated during tick blood meal which transmits *Anaplasma* to mammals. These adhesin molecules are Outer membrane protein A (OmpA, 19.9 Kda in size, known as peptidoglycan associated lipoprotein,), *Anaplasma phagocytophilum* surface protein (Asp14, 14 kDa in size) and *Anaplasma phagocytophilum* invasion protein A (AipA , 36.9 kDa in size) (Ojogun, Kahlon et al. 2012, Naimi, Gumpf et al. 2020). The Outer membrane protein A (OmpA) binds to sialyl Lewis x (sLe^x) and other sialylated glycans that decorate P selectin glycoprotein 1 (PSGL-1) on the host cell. Asp14 interacts with host cell surface Protein Disulfide Isomerase (PDI) to induce host cell invasion. AipA also binds to an unknown host cell receptor for cellular invasion, but this receptor has not been identified yet (Naimi, Gumpf et al. 2020). However, despite the identification of these *Anaplasma* adhesins, their total contribution to *in vivo* infection is still unknown.

4.2 Anaplasmosis treatment and development of novel therapeutics.

As *Anaplasma phagocytophilum* is an obligate intracellular pathogen , these bacteria must invade a suitable host cell for survival and replication. This intracellular habitat protects these organisms from potentially hostile challenging environments. To survive under these hostile conditions, *Anaplasma phagocytophilum* must adapt itself by altering/changing its virulence factors, gene expression patterns and perhaps even modulate host response as well. Even though

there are compelling data to support an idea that host-pathogen stress response could become a novel antibiotic target (Zhang, Kedzierska-Mieszkowska et al. 2013, Kuczynska-Wisnik, Cheng et al. 2017, Glaza, Ranaweera et al. 2020), so far that idea has not been tested in obligate intracellular pathogens. Despite the fact that there were more than 5000 cases of Human granulocytic anaplasmosis (HGA) in 2017 and over 4000 cases in 2018 (https://www.cdc.gov/anaplasmosis/stats/index.html, Accessed on 02/23/2021), there are no prophylactic treatments against HGA. Doxycycline is the only available treatment against Human granulocytic anaplasmosis (HGA) as there are no vaccines available for HGA (Naimi, Gumpf et al. 2020). It is worth stating here that doxycycline treatment has become difficult to administer against HGA due to nonspecific onset of the disease. Thus, early detection of anaplasmosis has been difficult, which reduces the efficacy of the antibiotic treatment.

There are compelling evidence to suggest that heat shock proteins play a vital role in host-pathogen interactions. The data indicate that heat shock proteins from the pathogen also bind to host cells and alter host cell cellular metabolism. In *Anaplasma phagocytophilum* major surface protein 4 (MSP4) and Heat shock protein 70 (HSP70/DnaK) were shown to be localized on the bacterial membrane and to assist the bacterium in invading tick cells (Villar, Ayllón et al. 2010, Busby, Ayllón et al. 2012, Contreras, Alberdi et al. 2017). According to the findings of Busby and colleagues , cultured tick cells tend to upregulate Heat Shock Protein 20 (HSP20) and HSP70 as a cellular response to blood feeding stress and *Anaplasma* infection. Importntly, (HSP70/ DnaK) is upregulated in both host and the pathogen which suggests that HSP70 is one of the most important heat shock proteins in host pathogen interaction, stress response and even progression of the infection. As HSP20 is upregulated differentially in tick cells in response to *Anaplasma phagocytophilum* , Busby and colleagues hypothesized that HSP20 provides protection for tick

cells from further infections by *Anaplasma phagocytophilum.* With the intention of developing a vaccine, Contreras and colleagues (Contreras, Alberdi et al. 2017) studied the protective capacity of MSP4 and MSP4-HSP70 antigens of *Anaplasma phagocytophilum.* They used sheep as the animal model and found that MSP4-HSP70 recombinant protein was only partially protective against pathogen infection in this model.

It is well-established that bacterial DnaK works together with HSP100/ bacterial ClpB to maintain proper protein homeostasis. Sequence alignments of *Anaplasma* ClpB and DnaK reveal that key structural and functional arrangements are highly conserved among different species (figure 4.9 and Appendix B supplementary figure B.7). However, to date there are no biochemical studies on ClpB and DnaK in *Anaplasma phagocytophilum.* Perhaps one reason for a lack of such studies on chaperones from obligate intracellular pathogens, could be due to the difficulties in manipulating the genome of these bacteria. In order to address this knowledge gap, we decided to perform a full *in vitro* biochemical characterization of ClpB and DnaK from *Anaplasma phagocytophilum*. Our main goal of this study was to assess the extent of protein aggregation in *Anaplasma phagocytophilum* during infection of mammalian cells and gain insights into substrate recognition mechanisms of ClpB and DnaK in *Anaplasma phagocytophilum*. We were also hoping to translate our novel findings on substrate detection and chaperone interaction in *Anaplasma* into developing and testing small-molecule inhibitors as potential lead candidates for treatment of Anaplasmosis.

4.3 Material and Methods.

4.3.1 Cultivation and purification of *Anaplasma phagocytophilum* in HL60 cells.

The human promyelocytic cell lines HL-60 (ATCC CCL-240, Manassas, VA) and *Anaplasma phagocytophilum* strain HZ infected HL-60 were used for *Anaplasma* genomic DNA isolation. The cells were cultured in complete RPMI 1640 (Gibco/ThermoFisher Scientific) where RPMI media had been supplemented with 10% fetal bovine serum (Invitrogen/ThermoFisher Scientific, Waltham, MA) and 2 mM L-glutamine (Mediatech, Manassas, VA) . Media preparation and cells were handled under sterile conditions all the time. We followed previously described protocols from following groups (Chen, Dumler et al. 1994, Zhang, Chen et al. 2021). The cells (infected and uninfected) were maintained in T25 tissue culture flasks (TPP Techno Plastic Products AG). For certain experiments culture flasks were treated either with or without 0.5 mM Guanidine hydrochloride (GuHCl). At specific time intervals, 25 ml cultures were used for *A.phagocytophilum* purification and analysis of protein aggregation. Infection status of each culture was checked by hematoxylin staining as described by (Cheng and Ganta 2008). HL60 live cell counting was done using, Invitrogen Countess II Automated Cell Counter (ThermoFisher Scientific, USA).

To isolate host cell free *Anaplasma phagocytophilum* , when HL-60 cells were infected about 80-90%, cells were collected from T25 flasks and were harvested by centrifugation at 400 × g for 10 min at 4°C. The pellet was resuspended in 5 ml of RPMI media , and the cells were disrupted by sonication (Fisher Scientific, 60 Sonic Dismembrator). Sonicated two times for 30 seconds, each time, at a setting of 6.5 to release Anaplasma organisms from host HL-60 cells. The lysate was Centrifuged for 5 min at 100 × g, 4°C, to pellet host nucleus and other host cell debris. The supernatant was filtered by passing it through 2.7μm sterile isopore membrane filters

(Whatman, Pittsburgh, PA). This filtrate then centrifuged for 15 minutes at 15,560 × g, 4°C to isolate host cell free *Anaplasma*.

4.3.2 Isolation of A*naplasma* protein aggregates.

Once host cell free Anaplasma were isolated, cells were lysed by sonication as mentioned in (Kwiatkowska, Matuszewska et al. 2008). We used equal number of *Anaplasma* cells from all samples for protein isolation experiments. Equal number of *Anaplasma* were taken as described by (Sirigireddy and Ganta 2005) .Unbroken bacterial cells were removed by centrifugation (2000 × g, at 4°C for 20 minutes). The supernatant from earlier step was incubated with 2% Triton X-100 for 30 minutes at room temperature. Next , Triton treated sample was centrifuged at 21000 × g, at 4°C for 30 minutes to separate soluble protein and insoluble aggregates from *Anaplasma*. The Protein concentration of the aggregated and soluble fractions were measured by Bradford method (Bio-Rad ,USA).

4.3.3 Cloning of *Anaplasma* ClpB and DnaK.

Anaplasma phagocytophilum genomic DNA was isolated using Wizard® Genomic DNA Purification Kit (Promega ,France) following manufacturers protocol. In order to confirm isolation of *Anaplasma* genomic DNA ,we had used Anaplasma specific primers for detection of Anaplasma 16S rRNA gene (0.48-kb in size ,Forward - 5′ ctcagaacgaacgctgg 3', Reverse - 5′catttctagtggctatccc 3'). Isolated *Anaplasma* genomic DNA was used as the template and Polymerase Chain Reaction (PCR) was carried out as described in the (Sirigireddy and Ganta 2005) and PCR products were resolved on a 1% agarose gel in 1× Tris-acetate EDTA buffer (40 mmol/L Tris-acetate, 1 mmol/L EDTA, pH 8.0) supplemented with 0.1 μg/ml of Ethidium Bromide, and were visualized under UV light. Our PCR results confirmed presence of *Anaplasma* genomic DNA.

We decided to use pET-28a(+) vector (Novagen ,5369bp in size) for our cloning of *Anaplasma* ClpB and DnaK. Amino acid sequence for *A. phagocytophilum* ClpB was downloaded from protein data base of National Center for Biotechnology Information (NCBI) with the following GeneBank id: GenBank: ABD44425.1 (ATP-dependent chaperone protein ClpB *Anaplasma phagocytophilum* str. HZ). We designed ClpB specific primers having restriction endonuclease sites included in the 5' tail of each forward and reverse primers. We used NheI and XhoI restriction enzymes (From New England Biolabs) to digest pET-28a(+) vector and our ClpB forward and reverse primers had same restriction enzymes sites in 5' tails of each primer (restriction sites underlined).

Aph ClpB-Forward: 5'ACTGA GCTAGC ATG GAC TTG AAT AAG TTT ACT GA 3'

Aph ClpB-Reverse: 5' TCAGT CTCGAG TTA TGT AGC CTC TTT GAT AAC TAG 3'

The entire protein coding sequence of *A. phagocytophilum* ClpB was amplified using above mentioned gene specific primers and Q5® High-Fidelity DNA Polymerase (NEB) following manufacturer's (NEB) instructions. The PCR products and pET-28a(+) vector were digested with NheI and XhoI, and purified using QIAquick PCR Purification (QIAGEN ,USA) to remove the digested overhangs and other small fragments. Later purified digested vector and PCR primer products were ligated using T4 DNA Ligase (NEB) following manufacturer's instructions. This new plasmid was name as pCBR_AphClpB (figure 4.4). Correct insertion of *Anaplasma* ClpB gene was confirmed by DNA sequencing (MCLAB, South San Francisco, CA 94080) using general T7 promoter primer, general T7 terminator primer and following *Anaplasma* ClpB specific internal primers.

AphClpB Internal 1 Forward 5' CTC GCT AGG AGA TTT CAA CCA G 3'

AphClpB Internal 1 Reverse 5' GCT GCA TTA CCT GAC GAT CAA TAC 3'

AphClpB Internal 2 Forward 5’ ACA AAA AAC AAT CCT GCC CTA ATT G 3’

AphClpB Internl 2 Reverse 5’ CTA ATA CCG TGG TGC AGC TC 3’

AphClpB Internal 3 Forward 5’ CAG GAG AAT TAA TGT ACG GAA TCA TTC 3’

AphClpB Internl 3 Reverse 5’ GCG CTC CTA TCA GTT TTG C 3’

This new recombinant plasmid (pCBR_AphClpB, figure 4.4) got transformed into *E. coli* BL21(DE3) cells (EMD Millipore).

Amino acid sequence for *A. phagocytophilum* DnaK was downloaded from protein data base of National Center for Biotechnology Information (NCBI) with the following GeneBank id: GenBank: ABD44082.1 (chaperone protein DnaK *Anaplasma phagocytophilum* str. HZ). We designed DnaK specific primers having restriction endonuclease sites included in the 5’ tail of each forward and reverse primers. We used NheI and XhoI restriction enzymes (From New England Biolabs) to digest pET-28a(+) vector and our ClpB forward and reverse primers had same restriction enzymes sites in 5’ tails of each primer (restriction sites underlined).

Aph DnaK-Forward: 5 ‘ ACTGA GCTAGC ATGGGGGTAGTCATGG 3’

Aph DnaK-Reverse: 5’ TCAGT CTCGAG CTAAGTATTCTTCTTGTCCTCG 3’

The entire protein coding sequence of *A. phagocytophilum* DnaK was amplified using above mentioned gene specific primers and Q5® High-Fidelity DNA Polymerase (NEB) following manufacturer’s (NEB) instructions. The PCR products and pET-28a(+) vector were digested with NheI and XhoI, and purified using QIAquick PCR Purification (QIAGEN ,USA) to remove the digested overhangs and other small fragments. Later, purified digested vector and PCR primer products were ligated using T4 DNA Ligase (NEB) following manufacturer’s instructions. This new plasmid was name as pCBR_AphDnaK (figure 4.5). Correct insertion of *Anaplasma* DnaK gene was confirmed by DNA sequencing (MCLAB, South San Francisco, CA 94080) using

general T7 promoter primer, general T7 terminator primer and following *Anaplasma* DnaK specific internal primers.

AphDnaK Internal 1 Forward 5’ GCT GGT ACG ATT GCT GGC CTA G 3’

AphDnaK Internal 1 Reverse 5’ ATC ACT CAA AGC CTT CTT ACA TGG CTC TAT AGT 3’

AphDnaK Internal 2 Forward 5’ GCG GGT ATT AAG GAT AAC AGT AAG GTC GAC 3’

AphDnaK Internl 2 Reverse 5’ GCC TTA TCC TTT GCT GAA ACG TGC A 3’

AphDnaK Internal 3 Forward 5’ TTC AGG TAA GGA GCA GAC TAT TAA G 3’

AphDnaK Internl 3 Reverse 5’ AGA CAA ACG CGC CAA TCT GAA C 3’

This new recombinant plasmid (pCBR_AphDnaK, figure 4.5) got transformed into *E. coli* BL21(DE3) cells (EMD Millipore).

4.3.4 Purification of recombinant *Anaplasma* proteins.

Once new recombinant plasmids (pCBR_AphClpB & pCBR_AphDnaK) got transformed into *E. coli* BL21(DE3) cells (EMD Millipore) separately. Single transformants were used to prepare an overnight culture in the LB media supplemented with kanamycin (50 μg/ml). On the following day, the overnight culture was diluted 50-fold in 1 l of LB with kanamycin (50 μg/ml). The cultures were incubated at 18°C until the optical density reached ~0.5 (at 600 nm). When optical density reached ~0.5, Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1mM and left for overnight incubation (with shaking at 250 RPM) at 18°C. Following day, bacterial cells were collected by centrifugation (Beckman JA-14, 3000 relative centrifugal force, 15 min, 4 °C). The pellet was suspended in 20 ml of cold T10 buffer (40-mM Tris-HCl, pH 8.0, 100 mM KCl, 10 mM imidazole) by vortexing. Cells were disrupted by sonication. The soluble fraction was obtained by centrifugation (Beckman JA-20, 11,500 RPM, 60 min, 4°C). The supernatant was loaded on a column with 5 ml of nickel-nitrilotriacetic acid resin

(Ni-NTA, QIAGEN) equilibrated with buffer T10. The column was washed with 50 ml of buffer T20 (40 mM Tris-HCl, pH 8.0, 100 mM KCl, 20 mM imidazole). Proteins were eluted with buffer T250 (40 mM Tris-HCl, pH 8.0, 100 mM KCl, 250 mM imidazole).

Fractions containing *Anaplasma* ClpB or *Anaplasma* DnaK were pooled separately and concentrated with a centrifugal filter device (50000 molecular weight cut-off for *Anaplasma* ClpB, 30000 molecular weight cut-off for *Anaplasma* DnaK, 3500 RPM, 4°C). The concentrated *Anaplasma* ClpB or *Anaplasma* DnaK samples were resolved on a Superdex 200 gel filtration column equilibrated with buffer B (50 mM Tris-HCl, pH 7.5, 0.2 M KCl, 20 mM $MgCl_2$, 10% glycerol, 1 mM EDTA, 1 mM DTT). Fractions containing *Anaplasma* ClpB or *Anaplasma* DnaK with purity greater than 95% (SDS-PAGE followed by Coomassie staining) were collected and concentrated.

Next, Thrombin cleavage (to remove N-terminal Histidine tag) was carried out on concentrated *Anaplasma* ClpB and *Anaplasma* DnaK fractions using Thrombin CleanCleave™ Kit (Millipore Sigma) following manufacturer's instructions. Later, Thrombin cleaved *Anaplasma* ClpB and *Anaplasma* DnaK fractions collected, and protein concentration was determined by the Bradford method. The aliquots of *Anaplasma* ClpB and *Anaplasma* DnaK were stored at −20°C. It is important to mention here that the post Thrombin cleavage N-terminal sequence of the recombinant *Anaplasma* ClpB and *Anaplasma* DnaK contained following three (Glycine-Serine-Histidine) additional amino acids.

E. coli ClpB, DnaK and *E. chaffeensis* ClpB were purified as described earlier (Zhang, Kedzierska-Mieszkowska et al. 2013, Glaza, Ranaweera et al. 2020).

4.3.5 Plasmids and bacterial strains.

Plasmid pCBR_AphClpB (this study , figure 4.4), pCBR_AphDnaK (this study, figure 4.5) and pCS6 (Schmidt, Shis et al. 2012) used. pCS6 was a gift from Matthew Bennett (Addgene plasmid number 55752 ; http://n2t.net/addgene:55752 ; RRID:Addgene_55752)

Anaplasma phagocytophilum (strain HZ) and following strains of *E. coli* were used: MC4100, MC4100ΔclpB::kan,(Squires, Pedersen et al. 1991), and DnaK103 (Spence, Cegielska et al. 1990).

4.3.6 Peptides and aggregation-propensity prediction.

Peptides listed in Appendix B supplementary figure B.6 were purchased from Peptide 2.0 (Chantilly, Virginia). The peptides were N-terminally labeled with FITC and C-terminally amidated by the manufacturer. Peptide concentration was determined spectrophotometrically using the FITC molar extinction coefficient 77,000 $cm^{-1}M^{-1}$ at 494 nm. The TANGO algorithm (http://tango.crg.es/) was used to compute the aggregation-propensity profiles for the peptide sequences (Fernandez-Escamilla, Rousseau et al. 2004, Rousseau, Schymkowitz et al. 2006). Our previous publication (chapter 2) has full details of Tango scores for aggregate formation tendency and location of predicted aggregate prone segments of peptides used (Ranaweera, Glaza et al. 2018). Sequences of the peptides are shown in Appendix B supplementary figure B.6.

4.3.7 ATPase activity assays.

To determine the ATPase activity, either ClpB or DnaK was mixed in buffer C (50 mM Tris-HCl, pH 7.4, 20 mM $MgCl_2$, 1 mM EDTA, 0.5 mM tris(2-carboxyethyl) phosphine [TCEP]). In some experiments, the ClpB or DnaK solution was supplemented with total of 0.004 mg of κ-casein. Additionally, 2.2μg of DnaJ and 0.8μg of GrpE was added for all ATPase measurements of DnaK.

After a 10 minutes preincubation of relevant protein at 37 °C, the ATP hydrolysis reaction was initiated by adding 2 µl of 100 mM ATP. For all ATPase activity readings of *Anaplasma* ClpB or *Anaplasma* DnaK, 15µl of *Anaplasma* ClpB or DnaK (55µg of *Anaplasma* ClpB or DnaK) was mixed with 200 µl of the ammonium molybdate/malachite green reagent dispensed into a 96- well plate and left for 1 minute at room temperature. Next 30 µl of 34% sodium citrate was added and mixed well and left for 15 minutes for color development. Absorbance readings were taken using microplate reader ((BioTek, Winooski, VT) at 630nm after 15 minutes. The plate was agitated inside the microplate reader (BioTek, Winooski, VT) for 15 min at room temperature.

The samples were incubated for 60 minutes before addition of ammonium molybdate/malachite green reagent (basal ClpB or DnaK ATPase activity in the absence of κ-casein) or 45 minutes before addition of ammonium molybdate/malachite green reagent (for ATPase activity in presence of κ-casein) at 37 °C. The readouts for ClpB or DnaK containing samples were corrected for absorbance of samples without ClpB or DnaK, to account for nonenzymatic production of inorganic phosphate. A standard curve obtained with different inorganic phosphate concentrations was used to determine the amount of phosphate produced from ATP in the presence of ClpB or DnaK.

4.3.8 *Anaplasma* ClpB and DnaK interaction with FITC labelled peptides.

Fluorescence anisotropy measurements were performed using a PerkinElmer LS55 fluorescence spectrometer equipped with automatic polarizers. The FITC-labeled casein was dissolved in buffer (50 mM Tris-HCl, pH 7.4, 20 mM $MgCl_2$, 1 mM EDTA, 0.5 mM tris(2-carboxyethyl) phosphine [TCEP]) and supplemented with 2 mM ATP or ADP. Peptide concentration was100 nM. The excitation and emission wavelengths were 494 nm and 521 nm, respectively. Before addition of ClpB or DnaK, the peptide solutions were equilibrated at room

temperature for 10 minutes and their fluorescence anisotropy was recorded. Next, either *Anaplasma* ClpB or DnaK was added in a stepwise manner to the cuvette with manual mixing. After each addition of ClpB or DnaK, the samples were incubated for 2 minutes at room temperature, the anisotropy readings were collected during the following 2 minutes and averaged. Each titration was performed twice with the average anisotropy readings and their standard deviations calculated, using the GraphPad Prism software. The change of anisotropy, Δr, was obtained by subtracting the initial anisotropy reading for either ClpB-free or DnaK-free peptides from the anisotropy measured during the titrations with ClpB or DnaK. GraphPad Prism was used for preparation of binding isotherms.

For all peptide interactions (*Anaplasma* ClpB-peptide, *Anaplasma* DnaK-peptide or *E. coli* DnaK-peptide), GraphPad Prism was used with nonlinear least-squares one site specific binding model. One site specific binding model equation was $\Delta r = \Delta r_{max} * X/(Kd + X)$. Δr is the observed change in anisotropy upon addition of relevant protein, Δr_{max} is the fitting parameter corresponding to an anisotropy change at saturation, X is protein concentration and Kd is the apparent dissociation constant for the peptide-ClpB or peptide-DnaK interaction.

4.3.9 Bacterial viability.

E. coli MC4100, MC4100ΔclpB, DnaK103 strains were used to determine bacterial growth and viability. Bacteria were maintained in the LB media, and in the case of the MC4100ΔclpB strain, LB was supplemented with 30 μg/ml kanamycin. All experiments were initiated by preparing overnight cultures inoculated from single colonies and grown at 37 °C. The DnaK103 cultures was grown at 30 °C. On the following day, the cultures were diluted 100-fold in 10 ml of LB without antibiotics and incubated at 37 °C. The culture optical density was monitored at 600 nm. When absorbance at 600 nm reached 0.4, samples were immediately transferred into an

incubator/shaker at temperature 30, 45, or 50 °C. Bacteria were cultured for up to 4 h with shaking (200 RPM). At specific time points, 100 µl of each culture was withdrawn and serially diluted in sterile 0.9% NaCl up to 10^6 fold dilution. After a complete adsorption of liquid on the LB agar surface, the plates were incubated overnight at 37 °C.

4.3.10 Confocal Microscopy.

ClpB-YFP (Yellow Fluorescent Protein) was expressed at low levels in *E. coli* strains of ΔclpB, DnaK103, as described by (Winkler, Seybert et al. 2010) . The expression of ClpB-YFP was initiated by addition of 200 µM IPTG. The bacterial cells were cultured at 30°C. Protein translation was stopped by addition of erythromycin (30 µg/ml), and each culture was divided into two samples. One sample remained at 30 °C (control), whereas the other sample was transferred to 45 °C for 30 min (heat-shock conditions). A thin layer of 1% agarose (TopVision Low Melting Point Agarose, Thermo Scientific) was prepared on a clean microscope glass slide. A small drop of 1% agarose was placed near the frosted end of the glass slide while bringing a second glass slide at a 30 to 45 ° angle from above, allowing the agarose drop to spread along the contact edge. An *E. coli* culture sample (1 µl) was immediately placed on the agarose-covered glass slide, covered with a cover slip, and used for fluorescence imaging. Fluorescence imaging was performed using a Carl Zeiss LSM 880 Airyscan confocal microscope equipped with a Plan- Apochromat 63x/1.40 Oil DIC M27 objective coupled with a YFP Filter (excitation at 514 nm, emission at 561 nm, laser strength 0.98%). Zen 2.3 lite software was used for image capture and processing.

4.3.11 Western blotting.

The ClpB and DnaK protein levels in *A. phagocytophilum* cells were analyzed by Western blotting. Host cell free *A. phagocytophilum* cells were collected at different time post infection time points as described in materials and methods. The cells were suspended in 160 µl of 2x

Laemmli sample buffer. Sixty microliters of each sample was resolved in duplicates in 10% SDS-PAGE gel. One gel was used for blotting onto a nitrocellulose membrane and the second part of gel was stained with silver staining. The membrane was blocked with 3% Bovine serum albumin (BSA) dissolved in TBST buffer (19 mM Tris-HCl, pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.2% Tween-20) overnight at 4 °C. Next, the membrane was incubated with rabbit polyclonal anti-ClpB IgG or anti-DnaK IgG (1:50,000 in BSA) for 1 h at room temperature. After incubation with the primary antibody, the membrane was washed in TBST buffer (4 times, 10 min). Next, the membrane was incubated with anti-rabbit horseradish peroxidase conjugated secondary antibodies for 1 h at room temperature and washed in TBST (4 times, 10 min) and then TBST without Tween-20 (10 min). Signal detection was performed using SuperSignal West Pico Chemiluminescent Substrate (Pierce) and image acquiring was done using an Invitrogen iBright FL1500 Imaging System (ThermoFisher Scientific, USA).

4.3.12 Molecular detection of *Anaplasma phagocytophilum* by RT-PCR assay.

Real-time quantitative PCR (RT-PCR) was performed using TaqMan probes ,species specific primers, SuperScript™ III Platinum™ One-Step qRT-PCR Kit (Invitrogen, USA) and Applied Biosciences StepOnePlus thermocyclers. Followed the exact protocol as described by (Sirigireddy and Ganta 2005) using species specific primers and common TaqMan probes . Bacterial numbers were estimated by real-time quantitative PCR targeting the16S rRNA gene segment and preparing a standard curve as described by (Sirigireddy and Ganta 2005)

4.4 Results.

4.4.1 Guanidine hydrochloride (GuHCl) induces protein aggregation in *A. phagocytophilum.*

To study role of ClpB in *A. phagocytophilum* replication in mammalian cells, we used a known inhibitor of ClpB, Guanidine hydrochloride (GuHCl) (Zeymer, Werbeck et al. 2013, Kuczynska-Wisnik, Cheng et al. 2017). Previous studies demonstrated that the GuHCl-induced protein aggregation in *E.chaffeensis*, mediated by interaction of the guanidinium cation with ClpB (Kuczynska-Wisnik, Cheng et al. 2017). Equal number of HL-60 cells were infected with *A. phagocytophilum* (infected with equal number of bacteria) in T25 flasks, with or without 0.5 mM Guanidine hydrochloride (GuHCl). Appendix B supplementary figure B.1 shows images of *A. phagocytophilum* replication in HL 60 cells at different time points. Host cells were lysed at different time points, and host cell free bacteria were isolated. Soluble and aggregated protein fraction isolation was carried out as described in (Kwiatkowska, Matuszewska et al. 2008, Kuczynska-Wisnik, Cheng et al. 2017) (figure 4.6) and gels were stained by Silver staining method (Pierce™ Silver Stain Kit , Thermo Scientific™). Total protein concentration was measured by Bradford method and equal amounts of protein were loaded in the gel for comparison. A percentage of aggregated proteins found in *A. phagocytophilum* at different time points is shown in figure 4.6. Our data showed that the amounts of aggregated proteins found in *Anaplasma* increased up to 48 hours with 0.5 mM Guanidine hydrochloride, while untreated samples maintained the amount of aggregated proteins for up to 96 hours without any significant changes. We performed a separate control experiment to demonstrate that HL 60 cell growth was not significantly affected by 0.5 mM Guanidine hydrochloride (figure 4.6).

4.4.2 *A. phagocytophilum* ClpB is preferentially targeted towards aggregates during GuHCl treatment.

A. phagocytophilum (with or without 0.5 mM GuHCl) was isolated from the host cells after different post infection time points, bacteria were lysed, and soluble and aggregated fractions were isolated as described in Materials and Methods. Equal amounts of proteins were loaded on SDS-PAGE and immune-detection of ClpB in the soluble and aggregated fractions was performed. We observed a significant enrichment of ClpB in the aggregated fraction from the 0.5 mM GuHCl treated samples. We observed a slight increase of ClpB levels in the soluble fraction (after 96 hrs) of *A. phagocytophilum* treated with 0.5 mM GuHCl (based on band intensity, figure 4.7). We also performed Western blotting to detect *Anaplasma* DnaK levels in the same samples (figure 4.8) . We observed that higher amounts of DnaK accumulated within the first 48 hours in the samples where no GuHCl was present (based on signal intensity in figure 4.8).

4.4.3 *A. phagocytophilum* ClpB partially complements cellular functions of *E. coli* ClpB in *E. coli* cells.

E. coli strain where the ClpB gene is disrupted (MC4100ΔclpB) was transformed with a cloned *Anaplasma* ClpB plasmid (pCBR_AphClpB, ClpB gene was regulated by T7 promoter in this construct) and plasmid pCS6 (Schmidt, Shis et al. 2012) for production of T7 RNA polymerase in *E. coli* (Addgene, 55752). Cells were grown under severe heat shock conditions at 50°C for the period of time indicated in figure 4.10 and then spotted on agar plates and incubated overnight at 37 °C. Each spot on the agar plates represents a viable culture after a 10 fold serial dilution (from left to right). Shown are representative results from 3 repeated experiments (figure 4.10). ClpB is an essential gene for the *E. coli* survival under severe heat shock and the ClpB-deficient *E. coli* cannot withstand a severe heat shock for an extended time (refer to Appendix B supplementary

figure B.2). Our results demonstrated that *Anaplasma* ClpB was produced in *E. coli* cells (Appendix B supplementary figure B.3) and that MC4100ΔclpB strain managed to replicate and survived longer at 50°C upon expression of *Anaplasma* ClpB.

4.4.4 *A. phagocytophilum* DnaK is unable to complement cellular functions of *E. coli* DnaK in *E. coli* cells.

E. coli strain DnaK103 which produces only an inactive fragment of DnaK, was transformed with recombinant *Anaplasma* DnaK plasmid (pCBR_AphDnaK) and plasmid pCS6 for production of T7 RNA polymerase. Once these two plasmids were present in DnaK103, a third plasmid ClpB-YFP (*E.coli* ClpB gene with the C-terminal fusion with Yellow Fluorescent Protein, (Glaza, Ranaweera et al. 2020)) was introduced. The cells were grown under moderate heat shock conditions at 45°C for the period indicated on the left (in hours) and then spotted on agar plates and incubated overnight at 37 °C. Each spot on the agar plates represents a viable culture after a 10-fold serial dilution (from left to right). Shown are representative results from 3 repeated experiments (figure 4.11). Our results demonstrated that even though *Anaplasma* DnaK was produced in *E. coli* Dnak103 cells (Appendix B supplementary figure B.4), *Anaplasma* DnaK could not complement cellular functions of native *E. coli* Dnak for heat shock response and did not enhance viability of *E. coli* Dnak103 cells.

Next, we used the Dnak103 *E. coli* cells in confocal microscopy experiments (see Materials and Methods section) and tested whether *Anaplasma* DnaK could drive ClpB-YFP towards the cellular poles where aggregated proteins accumulate in response to heat shock (Winkler, Seybert et al. 2010, Glaza, Ranaweera et al. 2020). Our confocal microscope images demonstrated that *Anaplasma* DnaK cannot guide *E. coli* ClpB-YFP towards aggregated proteins in heat-shocked cells (figure 4.12 and Appendix B supplementary figure B.5).

4.4.5 ATPase activity of purified *Anaplasma* ClpB and DnaK.

ATPase assays were performed as described in Materials and Methods. ATPase activity was measured without any substrate (basal activity) or with casein as a pseudo-substrate. *Anaplasma* ClpB and *Ehrlichia* ClpB both had similar basal ATPase activity and this basal ATPase activity was two-fold higher than the basal ATPase activity observed for *E .coli* ClpB (figure 4.13). When casein was used as substrate, the ATPase activity was stimulated for all ClpB proteins tested. However, for *E. coli* ClpB, substrate stimulated ATPase activity increased by a factor of two, compared to its basal ATPase activity, while for *Ehrlichia* ClpB the stimulation was by a factor of 1.4 and for *Anaplasma* ClpB the stimulation was marginal (by a factor of 1.1). When DnaK ATPase activities were tested in the presence of casein, *E. coli* DnaK showed an increase in its ATPase activity by a factor of 1.5 while *Anaplasma* DnaK showed a significant decrease in its ATPase activity (by a factor of 0.29) with the same pseudo-substrate (figure 4.13).

4.4.6 Interaction of *Anaplasma* ClpB and DnaK with substrate mimicking peptides.

Purified *Anaplasma* ClpB, *Anaplasma* DnaK and *E. coli* DnaK were used for fluorescence anisotropy binding experiments. These experiments were performed as described for the same set of fluorescein isothiocyanate (FITC, N-terminal labelling) labelled peptides, see Chapter 2 and (Ranaweera, Glaza et al. 2018). Refer to Appendix B supplementary figure B.6 for peptide sequences. We hypothesized that *Anaplasma* ClpB and DnaK may have a different preference in binding to substrates, as compared to *E. coli* ClpB and DnaK. Our *in vitro* binding experiments results demonstrated that *Anaplasma* ClpB and DnaK showed weaker interactions (compared to *E. coli* ClpB and DnaK) with substrate mimicking peptides, except the interaction of *Anaplasma* ClpB with B2T1 peptide. Binding isotherms and apparent dissociation constants (Kd) values for

Anaplasma ClpB and DnaK are shown in table 4.1, table 4.2, figure 4.14 and Appendix B supplementary figure B.6.

4.5 Discussion.

Currently, ticks are considered as the second most common vector of human infectious diseases, after mosquitoes. Ticks are obligate hematophagous arthropods and are believed to parasitize almost every class of vertebrates across the world (Parola and Raoult 2001). In this current study we investigated protein aggregation patterns during replication of *A. phagocytophilum*, a tick-borne pathogen responsible of infections in animals and humans. In addition to studying protein aggregation patterns during *Anaplasma* replication in HL60 host cells, we performed a biochemical characterization of two essential molecular chaperones ClpB and DnaK in *A. phagocytophilum.* To the best of our knowledge, this is the first biochemical characterization of ClpB and DnaK from *A. phagocytophilum.* Besides the *in vitro* biochemical assays, we also investigated the *in vivo* function of *Anaplasma* ClpB and DnaK with a complementation assay in *E. coli*.

From the very moment when a pathogen invades a host, the pathogen is under constant host and environmental stresses such as heat shock, oxidative stress and nitrosative stress . The successful survival, replication and transmission of a pathogen depends on its ability to withstand these stresses and the pathogen molecular response and regulation to counteracts these stresses. These molecular responses also help the pathogen to express its virulence genes in a timely manner (de la Fuente, Antunes et al. 2017). In recent years, stress and stress response in host pathogen interactions have paved novel ways to study infectious disease progression mechanisms, to develop novel antimicrobials and discover novel antibiotic targets. Now it is a well-accepted and established that heat shock proteins which are involved in host pathogen stress response could

become potential antimicrobial targets (Stewart and Young 2004, Zhang, Kedzierska-Mieszkowska et al. 2013, Kuczynska-Wisnik, Cheng et al. 2017, Glaza, Ranaweera et al. 2020). In such study conducted for *Anaplasma phagocytophilum*, major surface protein 4 (MSP4) and Heat shock protein 70 (HSP70/DnaK) were shown to be localized on the bacterial membrane and to assist the bacterium during invasion of tick cells (Villar, Ayllón et al. 2010). However , to date no effective compounds have been developed that could successfully target the host pathogen stress response mechanism.

Our group (Zhang, Kedzierska-Mieszkowska et al. 2013, Kuczynska-Wisnik, Cheng et al. 2017) demonstrated that ClpB-mRNA and protein aggregate levels were increased in *Ehrlichia chaffeensis* during the first 48 hours of post infection. In order to establish a link between *A. phagocytophilum* infection , replication and protein aggregate accumulation, we propagated *A. phagocytophilum* in infected HL60 cells with or without 0.5 mM GuHCl. Host cell free bacteria were isolated, cells were broken to isolate the soluble and the aggregated protein fractions at specific post infection time points (figure 4.6 and Appendix B supplementary figure B.1).

There were distinctive protein patterns observed for soluble and aggregated protein fractions (figure 4.6) at different time points (with or without 0.5 mM GuHCl) indicating that certain fractions of proteins are affected by these different growing conditions. As an example , we observed that new protein bands after 96 hours in GuHCl soluble fraction while few protein bands were not observed within first 48 hours in GuHCl treated samples .Our data for GuHCl treated samples indicated that the amounts of aggregated proteins increased gradually within first 48 hours and decrease gradually after 48 hours (figure 4.6) while untreated samples managed to maintain protein homeostasis in balance. GuHCl is a well-established small inhibitor molecule of ClpB and its yeast ortholog Hsp104. GuHCl binds to the first nucleotide binding (NBD) domain

of ClpB (chapter 1 figure 1.3), (Grimminger, Richter et al. 2004, Zeymer, Werbeck et al. 2013). We decided to use 0.5 mM GuHCl in our experiments as it has been previously shown that up to 0.5 mM GuHCl had no impact of on growth and replication of another mammalian cell line DH82 (Kuczynska-Wisnik, Cheng et al. 2017). It is also important to note that at the sub-millimolar concentration of GuHCl, it is highly unlikely that it will cause protein denaturation, which requires significantly higher molar GuHCl concentrations (Farrell, Qi et al. 2002). Indeed, we have not observed any significant reduction of growth of uninfected HL60 cells in presence of 0.5 mM GuHCl. (figure 4.6C).

As GuHCl is a known inhibitor of ClpB, ClpB is essential for reversal of protein aggregate formation and more ClpB-mRNA is expressed within first 48 hours of infection with more protein aggregates, we set to investigate ClpB protein levels in *A. phagocytophilum* infected HL60 samples at different time points with or without GuHCl. When examining ClpB levels in soluble fractions from both treatments, GuHCl untreated samples showed undetectable levels of ClpB in soluble fractions (figure 4.7). However, in GuHCl treated soluble fractions, the ClpB band started to appear after 72 hours and amount of ClpB further increased after 96 hours. Inhibition of ClpB by GuHCl could generate relatively large amounts of partially folded proteins or ClpB aggregates as well. Cellular metabolism of *A. phagocytophilum* in DC stages is extremely slow and believed to be in a metabolically dormant state (Troese and Carlyon 2009). Perhaps ClpB is a crucial protein which guides and controls a smooth transition from metabolically active RC stage to metabolically inert DC stage .

ClpB was seen accumulated with aggregated fractions both with and without GuHCl (figure 4.7). This result confirms ClpB's preferential targeting towards aggregated proteins and its biological function as a protein disaggregase. However, we noticed that there were large amounts

of ClpB accumulated in the aggregated fractions within the first 48 hours after the GuHCl treatment. A possible explanation of that result is that as more proteins become aggregated or misfolded during the first 48 hours (RC stage), ClpB is recruited to reactivate these proteins. However, ClpB's ATPase activity and substrate translocation function are hindered by GuHCl and thus ClpB is unable to detach from its substrates, leaving ClpB trapped among aggregated proteins.

Next, we investigated the amount of DnaK found in soluble and aggregated fractions with or without 0.5 mM GuHCl. We did not see detectable levels of DnaK present in the soluble fractions regardless of GuHCl treatment. However, we noticed that there were relatively high amounts of DnaK present in non-GuHCl treated aggregated fractions. The DnaK levels in those aggregated fractions increased gradually within the first 48 hours and then started to decrease for the next 48 hours. It is well-established that DnaK activates ClpB, guides ClpB towards aggregates, or even delivers aggregates to ClpB (Zolkiewski 1999, Winkler, Seybert et al. 2010, Rosenzweig, Moradi et al. 2013, Fernández-Higuero, Aguado et al. 2018). When *A. phagocytophilum* undergo extensive replication cycles within the first 48 hours of post infection, more proteins are likely to get aggregated or misfolded, driving cells to produce more ClpB proteins. For ClpB to detect aggregates, DnaK must recognize the aggregates first. Hence it is not surprising to find large amounts of DnaK with aggregated fractions (with no GuHCl treatment) together with ClpB. As GuHCl inhibits ClpB function, and forms ClpB aggregates, perhaps DnaK also gets trapped along with ClpB. Therefore, it is fair to expect some amount of DnaK in the aggregated fractions and our results confirmed this as we observed relatively same amount of DnaK (compared to soluble fraction) in aggregated fractions as well.

To investigate biochemical properties of *Anaplasma* ClpB and DnaK , we produced the recombinant proteins using the corresponding expression plasmids (figure 4.4 and figure 4.5).

Purified recombinant *Anaplasma* ClpB and DnaK remained soluble and SDS-PAGE gel results were consistent with the predicted monomer molecular weights: *Anaplasma* ClpB (~95 kDa) and Anaplasma DnaK (~70 kDa) (data not shown). Oligomer formation of Hsp100 proteins is linked with the chaperone activity. Therefore, we tested whether purified *Anaplasma* ClpB forms oligomers under our experimental conditions. We performed preliminary glutaraldehyde crosslinking experiment as described by (Kim, Cheong et al. 2000) which revealed that *Anaplasma* ClpB forms oligomers (crosslinking experiment was performed by Dr Sunitha Shiva data not presented in this thesis).

Next, we investigated the ATPase activity of *Anaplasma* ClpB and *Anaplasma* DnaK. As shown in figure 4.13 , *Anaplasma* ClpB and *Ehrlichia* ClpB both had similar basal ATPase activity while *E. coli* ClpB showed a two-fold lower basal ATPase activity. All three ClpB variants showed an increased ATPase activity in the presence of kappa casein (κ-casein). However, *E. coli* ClpB showed a two-fold increase (2.1 times higher compared to basal activity) while *Anaplasma* ClpB only showed a slight increase (1.1 times higher) of ATPase activity in presence of kappa casein. *Ehrlichia* ClpB had 1.39 times higher ATPase activity with kappa casein. Interestingly, our results showed that kappa casein significantly decreased the ATPase activity of *Anaplasma* DnaK (3.4 times lower compared to basal activity) while it increased the ATPase of *E. coli* DnaK (1.5 times higher). Our ATPase activity results for *Anaplasma* ClpB and DnaK demonstrate that these two chaperones do not respond to a pseudo-substrate in the same manner.

To evaluate the structure function relationships for the *Anaplasma* chaperones, a sequence alignment analysis for *Anaplasma phagocytophilum* ClpB (Appendix B supplementary figures B.7 and B.8) and DnaK (figure 4.9) was performed against ClpB and DnaK from several known human pathogens. Our ClpB sequence analysis (Appendix B supplementary figures B.7 and B.8) showed

that *Anaplasma phagocytophilum* ClpB shows a high degree of amino acid conservation in all signature motifs which are characteristic for the AAA+ superfamily (Hanson and Whiteheart 2005). At the same time, *Anaplasma* ClpB contains two less-conserved regions: the N-terminal domain and the middle domain. The N-terminal domain of *Anaplasma* ClpB contains three conserved amino acids (Thr7, Asp106, Glu112), which are essential for aggregate detection and binding in *E. coli* ClpB (Liu, Tek et al. 2002, Barnett, Nagy et al. 2005). The middle domain of *Anaplasma* ClpB also contains a conserved Tyr506, which supports the functional cooperation between Hsp100 and Hsp70/40 (Haslberger, Weibezahn et al. 2007). Therefore, we hypothesize that *Anaplasma* ClpB could cooperate with *E. coli* DnaK, DnaJ and GrpE (KJE). Our *in vivo* heat shock assay with *Anaplasma* ClpB expressed in *E. coli* MC4100ΔclpB strain (figure 4.10 and Appendix B supplementary figure B.2) demonstrated that *Anaplasma* ClpB can cooperate with *E. coli* KJE chaperones. As a result of this cooperative activity between *Anaplasma* ClpB and *E. coli* KJE, MC4100ΔclpB bacterial strain managed to proliferate successfully, and its survival under heat shock became extended at least by one hour. This ClpB dependent heat shock survival assay demonstrates that *E. coli* ClpB can be functionally replaced in *E. coli* with *Anaplasma* ClpB.

In parallel, Dr Sunitha Shiva from our group investigated the reactivation of glucose-6-phosphate dehydrogenase (G6PDH) aggregates by *Anaplasma* ClpB (manuscript in preparation). This aggregate reactivation assay was performed with *Anaplasma* ClpB, *E. coli* DnaK, *E. coli* DnaJ and *E. coli* GrpE as described in (Barnett, Nagy et al. 2005). The results indicate that *Anaplasma* ClpB cooperates with *E. coli* KJE chaperones in aggregate reactivation *in vitro* and confirm that *E. coli* ClpB is functionally equivalent to *Anaplasma* ClpB.

When analyzing the *Anaplasma* DnaK sequence (figure 4.9), we focused our attention on the most crucial conserved amino acids found in *E. coli* DnaK (*E. coli* DnaK structure is shown in

Appendix B supplementary figure B.9). There are two conserved residues found in the nucleotide binding domain (NBD) of *E. coli* DnaK which are essential for subdomain movements in NBD and also for nucleotide dependent structural rearrangements in NBD (Ung, Thompson et al. 2013). These residues, G228 and G229 are known as "hinge residues" and *A. phagocytophilum* DnaK also contains these "hinge residues" as G226 and G227 in its NBD. It is well-established that the DnaK function relies on allosteric communication between the N-terminal nucleotide binding domain (NBD) and a C-terminal substrate binding domain (SBD) and this allosteric communication between two domains is maintained by a short interdomain linker of 12 residues (Rosenzweig, Moradi et al. 2013, English, Sherman et al. 2017, Wickner, Camberg et al. 2017). There is a hydrophobic, highly conserved tetrapeptide (VLLL) found in *E. coli* interdomain linker ^{388}VLLL^{392} ,(Swain, Dinler et al. 2007, Zhuravleva and Gierasch 2011, Zhuravleva, Clerico et al. 2012) which governs nucleotide dependent substrate binding. The same tetrapeptide sequence is conserved in *A. phagocytophilum* (^{383}VLLL^{386}) as well.

According to (Kityk, Kopp et al. 2012) the following residues (K70, E171, T199, G197, G198) in *E. coli* DnaK-NBD are essential for a correct orientation of Mg^{2+} ion, nucleotide binding and hydrolysis reactions. Interestingly, the same essential residues are again conserved in *Anaplasma* sequence (K72, E168, T197, G195, G196) . Based on the studies conducted by Sue Wickner and colleagues on *E .coli* DnaK, ClpB and GrpE (Doyle, Shastry et al. 2015) , L49 is a critical residue for GrpE binding to DnaK and Y145 as a critical residue for DnaJ interaction with DnaK. Our *Anaplasma* sequence alignment again revealed that these residues are also conserved (as L51, Y143) in *A. phagocytophilum* DnaK sequence. Further analysis revealed that R56, L257, R261, N282, P284 and Y285 as key residues in DnaK which are responsible for direct interaction with the middle domain of ClpB and for activation of ClpB in aggregate disaggregation.

Interestingly our sequence analysis revealed that *Anaplasma* DnaK contains an analogous group of residues (R58, M251, R255, T276, P278, F279). Thus, we hypothesized that *Anaplasma* DnaK could functionally replace *E. coli* DnaK at cellular level.

To study whether *Anaplasma* DnaK can replace *E. coli* DnaK at cellular level, we used an *E. coli* strain with defective DnaK known as DnaK103 (Spence, Cegielska et al. 1990, Winkler, Tyedmers et al. 2012), which produces a truncated inactive DnaK which is unable to bind to its substrates. Our cloned *Anaplasma* DnaK plasmid (pCBR_AphDnaK) was transformed into DnaK103 strain and bacterial viability assays were performed at 45°C. There was no improvement of the DnaK103 cell viability upon expression of *Anaplasma* DnaK (figure 4.11), which revealed that *Anaplasma* DnaK cannot complement cellular functions of *E. coli* DnaK. One possible explanation of such a result could be that *E.coli* ClpB is unable to cooperate with *Anaplasma* DnaK. To find out if *Anaplasma* DnaK can target *E. coli* ClpB to cellular aggregates, we expressed ClpB-YFP in *E. coli* DnaK103 strain that contained *Anaplasma* DnaK plasmid. The experiment was carried our as described under the confocal microscopy section in Materials and Methods. Our group and other groups have shown in the past that a fully active DnaK is required to guide ClpB towards aggregates in a cellular milieu (Winkler, Seybert et al. 2010, Glaza, Ranaweera et al. 2020). However, our confocal images revealed that *Anaplasma* DnaK is unable to guide *E. coli* ClpB toward cellular aggregates as the ClpB signal did not accumulate at the poles of *E. coli* cells (figure 4.12, Appendix B supplementary figure B.5).

An experiment performed by Dr Sunitha Shiva (data not shown in this thesis , manuscript in preparation) for the glucose-6-phosphate dehydrogenase (G6PDH) aggregate reactivation with *Anaplasma* DnaK, *E. coli* ClpB, *E. coli* DnaJ and *E. coli* GrpE (Barnett, Nagy et al. 2005) showed

that *Anaplasma* DnaK does not cooperate with *E. coli* chaperones in aggregate reactivation and that *E. coli* DnaK cannot be functionally replaced in with *Anaplasma* DnaK.

What could be the reason/s for the above-mentioned differences in the species selectivity observed for *Anaplasma* ClpB and DnaK? It has been established that Hsp70/DnaK proteins act as the central hub of the protein homeostasis network in both bacteria and eukaryotes and are believed to be among the most important and abundant cellular proteins (Balchin, Hayer-Hartl et al. 2016). Our current understanding about protein disaggregation is that DnaK functions upstream of ClpB and that DnaK delivers the substrate to ClpB for disaggregation process and simultaneously activates ClpB through a direct interaction with the ClpB middle domain (Saibil 2013). There is a strong conservation of crucial amino acids in DnaK/Hsp70 across all domains of life. However, Hsp 70 proteins could have also evolved in a species-specific way while maintaining their basic structure and functions.

This species-specific evolution could help each Hsp70 to facilitate its unique functions within the species while serving other general purposes in protein homeostasis. In fact, species specificity has been previously observed for the ClpB/DnaK system and what was observed was that both ClpB and DnaK need to be from the same species. At the same time, it was observed that the species specificity appears to reside in the middle domain of ClpB (Sielaff and Tsai 2010, Miot, Reidy et al. 2011, Doyle, Shastry et al. 2015) as well. Interestingly ,our data suggested that the middle domain of *Anaplasma* and *E. coli* ClpB are similar enough, hence *E. coli* DnaK can interact with *Anaplasma* and *E. coli* ClpB. However, *Anaplasma* DnaK may differ from *E. coli* DnaK in its ability to activate *E. coli* ClpB, indicating that there could be a selectivity motif located in the Nucleotide binding domain of DnaK which is different between *Anaplasma* and *E. coli* DnaK (Appendix B supplementary figure B.9 for *E.coli* DnaK structure).

The biological function of ClpB is to detect and resolubilize cellular aggregates. We used fluorescence anisotropy to investigate the previously described model peptide interactions with ClpB (see Chapter 2). The experimental procedure is described under Materials and Methods in our prior publication (Glaza, Ranaweera et al. 2020). Briefly, binding of a FITC-labeled peptide to hexameric ClpB (~575 kDa) significantly decelerates the rotational diffusion of the fluorescein dye and produces an increase in fluorescence anisotropy which indicates specific binding of the peptide to the protein. We used free fluorescein (without peptide) in a control experiment with ClpB and did not observe any increase in anisotropy, which demonstrated that the interaction with ClpB is mediated by the peptide core and not by the FITC tag. Refer to Appendix B supplementary figure B.6 for details about amino acid sequences for the peptides used.

We tested *in vitro* substrate detection ability of purified *Anaplasma* ClpB and *Anaplasma* DnaK. In order to compare *Anaplasma* protein interactions with the model substrates, we compared our data with the previously generated data for *E. coli* ClpB and *E .coli* DnaK (Appendix B supplementary figure B.6). *Anaplasma* ClpB interacted with all peptides in an ATP dependent manner (figure 4.14) which confirms that ClpB recognizes these peptides as pseudo-substrates. *Anaplasma* ClpB managed to discriminate between ATP and ADP during binding experiments which again indicates that the experiments with the peptides mimic the biological function of *Anaplasma* ClpB. Our binding data revealed that *Anaplasma* ClpB displays several-fold weaker affinities towards all our model substrates, except B2T1, when compared to *E. coli* ClpB. We did not observe a positive cooperativity in binding to *Anaplasma* ClpB, in contrast to *E. coli* ClpB (Glaza, Ranaweera et al. 2020). Despite a strong amino acid conservation of ClpB among bacteria, perhaps the weaker substrate binding affinity of *Anaplasma* ClpB indicates significant differences in specific substrate binding motifs, which could have arisen due to the difference in the living

environment of the different bacterial species. *Anaplasma* is an obligate intracellular organism while *E. coli* is extracellular, and these two organisms may be exposed to significantly different types of stress throughout their life cycles, which could affect the extent of protein aggregation. Hence, in both these organisms the ClpB orthologs could have developed different substrate specificities in order to minimize protein aggregation and maximize protein reactivation in their respective bacterial cells. Our *Anaplasma* DnaK binding data also revealed that *Anaplasma* DnaK has several fold weaker affinities towards all our model substrates (figure 4.14), as compared to *E. coli* DnaK. Again, perhaps *Anaplasma* DnaK, like ClpB, developed its substrate specificities based on the environment *Anaplasma* proliferates in.

In a comparison of all our binding data from *Anaplasma* and *E. coli* (figure 4.14 and Appendix B supplementary figure B.6), an interesting substrate recognition pattern by ClpB and Dnak seems to emerge. We found that whenever ClpB interacts with a certain peptide with a higher affinity in one bacterial species, then DnaK from the same species shows a weaker affinity towards the same substrate. The opposite pattern is also true, e.g., if DnaK recognizes a certain peptide with a higher affinity, then ClpB from the same species shows a weaker affinity towards same substrate.

In order to explain this substrate affinity variation between ClpB and DnaK within a single species, we postulate that it could represent a cellular mechanism to assure that all aggregated, misfolded and partially folded proteins are processed through DnaK-ClpB chaperone system. Perhaps, different substrate affinities of DnaK and ClpB towards the same substrate eliminate competition between these chaperones. If a potential protein substrate is missed by DnaK, then ClpB could become a second tier of detection. If an aggregate is detected by DnaK first, then eventually it will be delivered to ClpB for disaggregation. This imperfect overlap in substrate

specificity between ClpB and DnaK may represent nature's way of "division of labor for molecular chaperones".

The results presented in this chapter demonstrate a link between replication *of Anaplasma phagocytophilum* in HL60 cells and the activity of two molecular chaperones which are involved in protein aggregation control. To the best of our knowledge this is the first biochemical study on ClpB and DnaK from *Anaplasma phagocytophilum.* Our results highlight the roles of ClpB and DnaK in pathogen's life cycle and validate those heat shock proteins as a potential target for novel antimicrobial therapies.

4.6 Acknowledgements.

I want to thank Dr Przemyslaw Glaza, Dr Sunitha Shiva, Dr Yiva Wang , Dr Arathy Nair, Dr Chandramouli Kondethimmanahalli, Dr Huitao Liu, Dr Li Chen, Swetha Madesh and Jonathan Ferm for their valuable suggestions and support throughout this project.

I want to thank Drs Bernd Bukau and Axel Mogk for providing ClpB-YFP plasmid and DnaK103 strain.

I want to thank Joel Sanneman from the Kansas State University Confocal Microscopy and Microfluorometry Core for help with confocal microscopy imaging.

I want to thank Donald Harbidge from the Kansas State University Molecular Biology Core facility for his valuable suggestions and support in conducting RT-PCR experiments using Applied Biosciences StepOnePlus instruments.

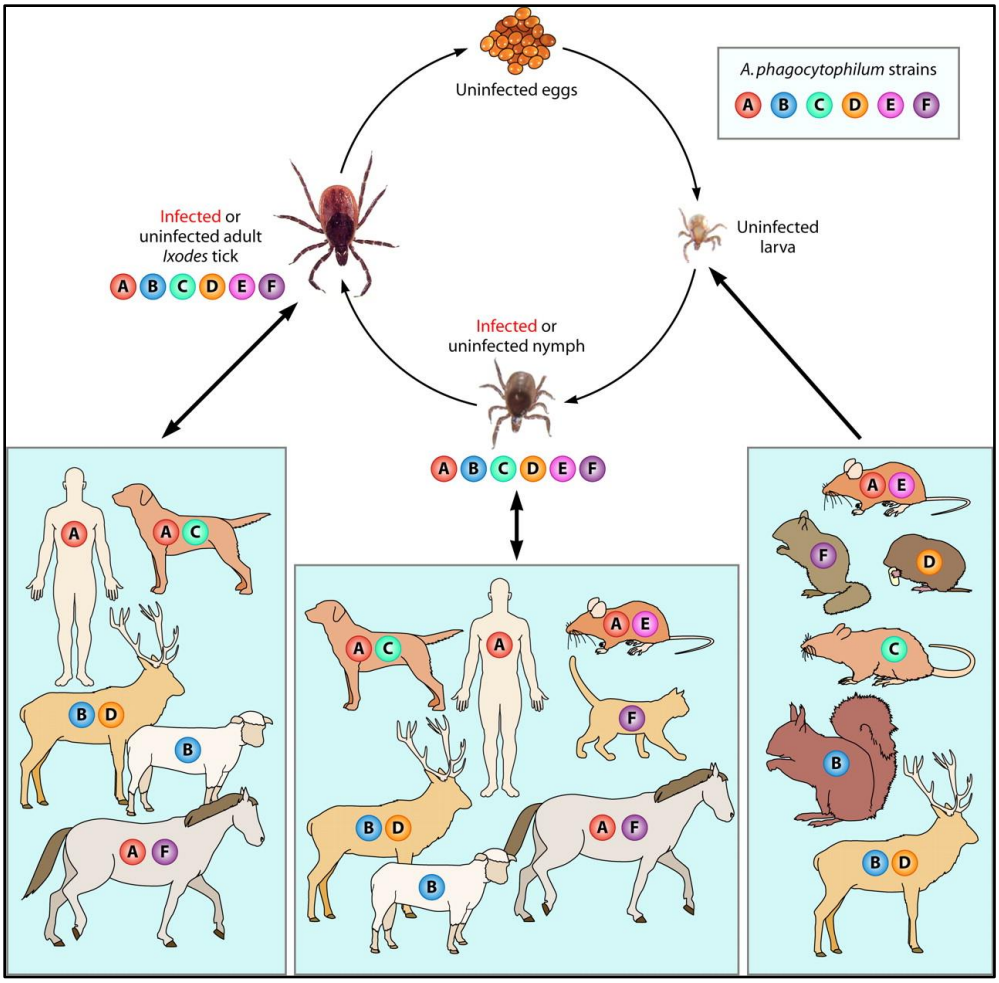


Figure 4.1. Life cycle of *Anaplasma phagocytophilum*.

There are different strains of *Anaplasma phagocytophilum* are found in the environments are shown as A,B,C,D,E and F. Susceptible mammals are also shown in this figure 4.1. *Anaplasma phagocytophilum* cannot be passed from infected adult ticks to eggs or from eggs to larva. Ticks at each larva, nymph or adult stages could acquire Anaplasma bacteria from host/s via a blood meal. This figure is taken from (Rikihisa 2011). Reprinted with permission.

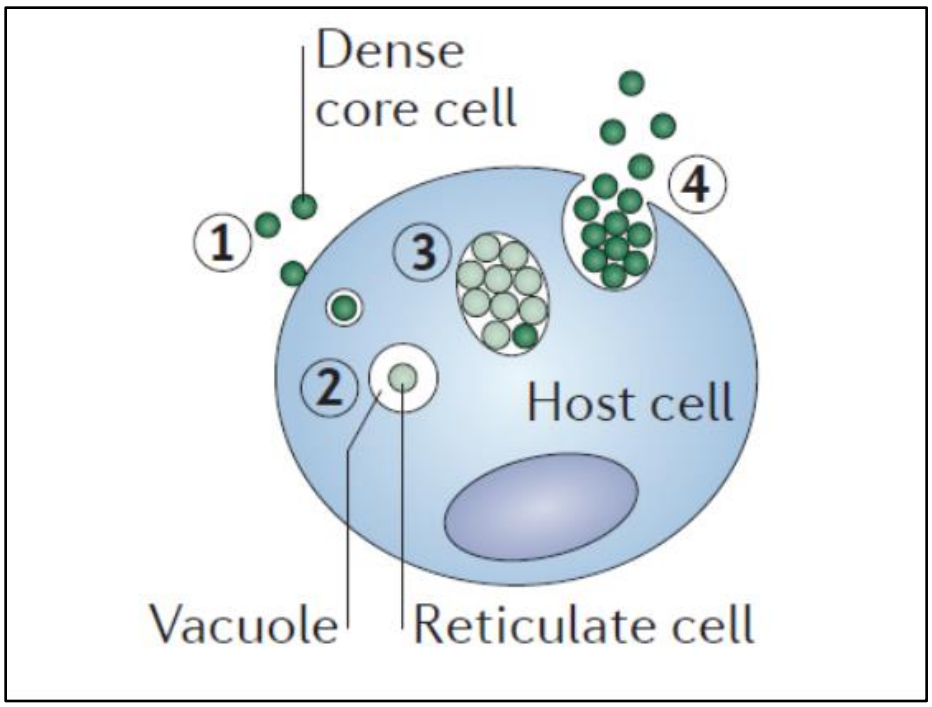
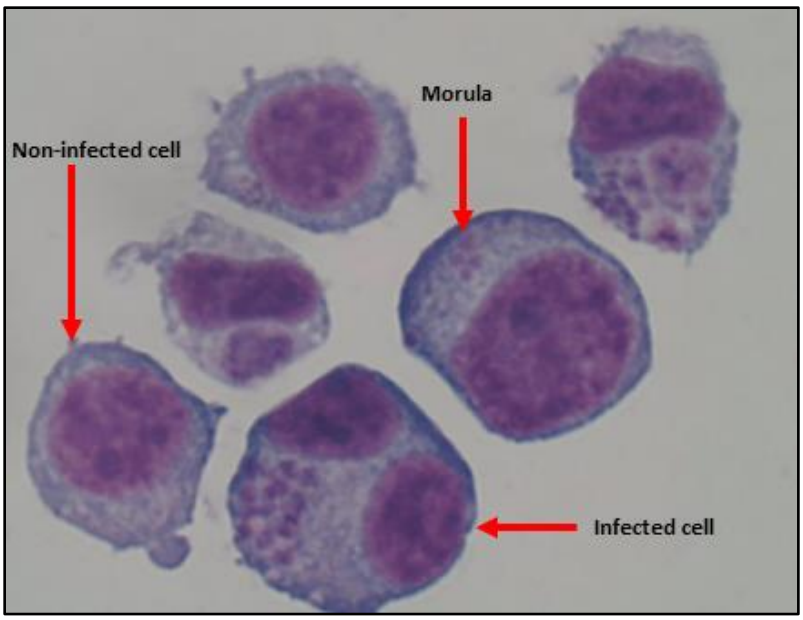


Figure 4.2. Dense Core and Reticulate Cells of *Anaplasma phagocytophilum*.

The infectious form of *Anaplasma phagocytophilum* is known as Dense Core (DC) cells and replicative form is known as Reticulate Cells (RC). When *Anaplasma phagocytophilum* encounters a suitable eukaryotic host cell, Dense Core cells get internalized (step 1). Once internalized, these DC cells differentiates into Reticulate Cells (step 2) in a vacuole. These RC cells would multiply by binary fission inside these vacuoles (step 3). Once these RC cells are multiplied, these RC cells will differentiate back to DC cells and exit the host cells by lysis (step 4). This figure is adapted from (McClure, Chávez et al. 2017). Reprinted with permission.



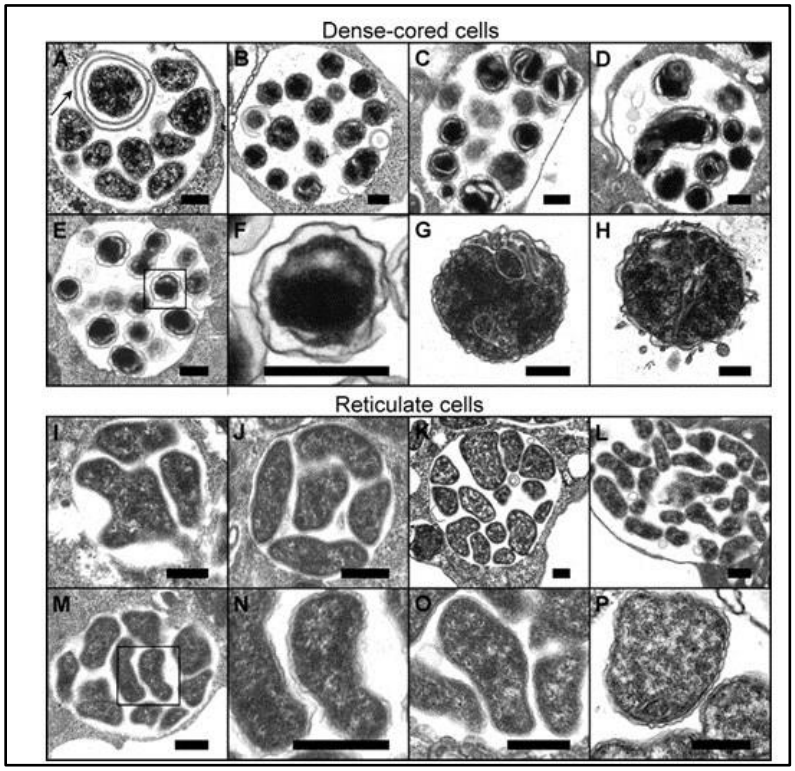


Figure 4.3. *A. phagocytophilum* morula , DCs and RCs in HL60 cells.

HL-60 cells were infected with *A.phagocytophilum* and stained with Hema 3™ Fixative stain and observed under light microscope (1000 magnification) . Both infected and uninfected

cells are shown. Images of infected HL-60 cells under transmission electron microscopy are also shown here (figure from Troese and Carlyon 2009 publication). DCs are shown from A to H and RCs are shown from I to P. Arrow in Panel A indicates an individual DC surrounded by a membranous projection. Panels F and N are magnifications of the boxed areas in panels E and M, respectively. Bars, 0.5 μm. Original figure from (Troese and Carlyon 2009). Reprinted with permission.

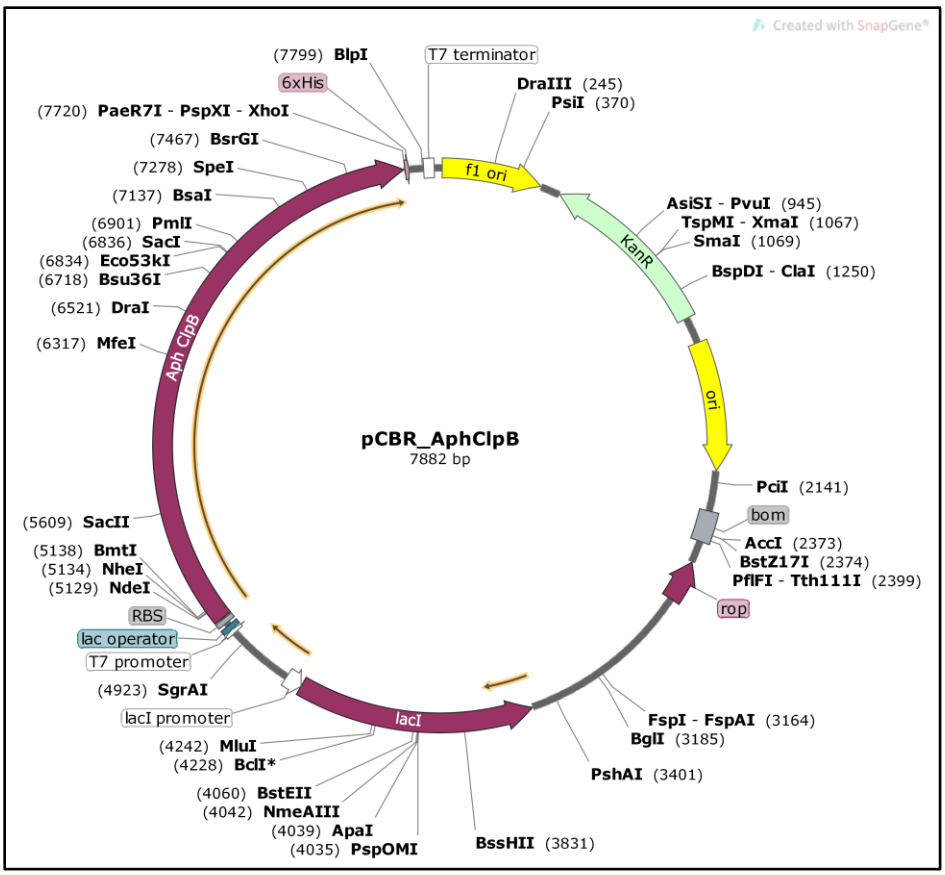


Figure 4.4. *A. Phagocytophilum* ClpB plasmid map.

Anaplasma ClpB gene (2649bp) was ligated into pET-28a(+) vector backbone after restriction digestion with Nhe1 and Xho1 restriction enzymes. This newly constructed plasmid has Kanamycin resistance. When *Anaplasma* ClpB gene gets translated from this new plasmid, six Histidine tag will be attached to the N-terminus of *Anaplasma* ClpB. There is a Thrombin cleavage site located in between six Histidine tag and very first Methionine of *Anaplasma* ClpB gene. Six Histidine tag and Thrombin cleavage site is not shown in the picture. There is a stop codon located in front of six Histidine tag at the C-terminus of *Anaplasma* ClpB gene. Hence no C-terminus Histidine tag gets attached with translated *Anaplasma* ClpB. Different restriction enzyme locations are also shown in the plasmid map. Plasmid map created using SnapGene software.

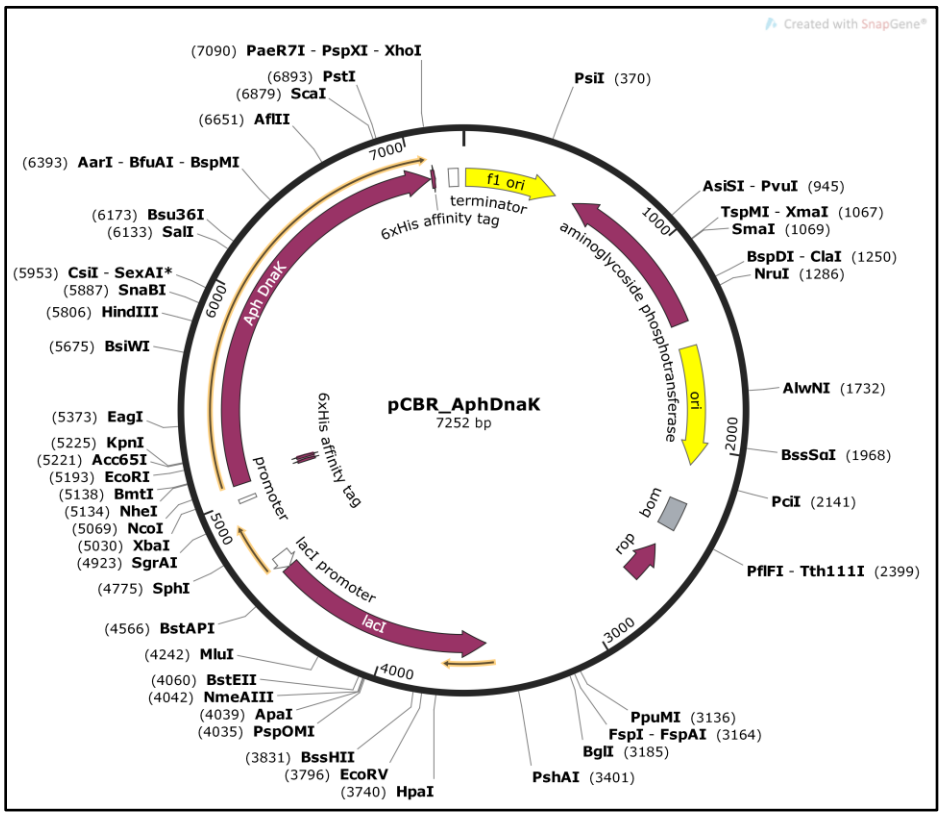
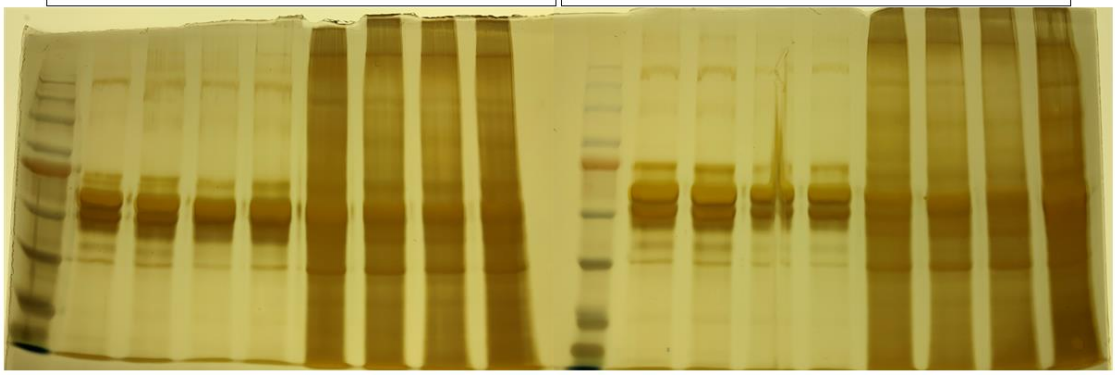


Figure 4.5. *A. Phagocytophilum* DnaK plasmid map.

Anaplasma DnaK gene (2019bp) was ligated into pET-28a(+) vector backbone after restriction digestion with Nhe1 and Xho1 restriction enzymes. This newly constructed plasmid has Kanamycin resistance. When *Anaplasma* DnaK gene gets translated from this new plasmid, six Histidine tag will be attached to the N-terminus of *Anaplasma* DnaK. There is a Thrombin cleavage site located in between six Histidine tag and very first Methionine of *Anaplasma* DnaK gene. Thrombin cleavage site is not shown in the picture. There is a stop codon located in front of six Histidine tag at the C-terminus of *Anaplasma* DnaK gene. Hence no C-terminus Histidine tag gets attached with translated *Anaplasma* DnaK. Different restriction enzyme locations are also shown in the plasmid map. Plasmid map created using SnapGene software.

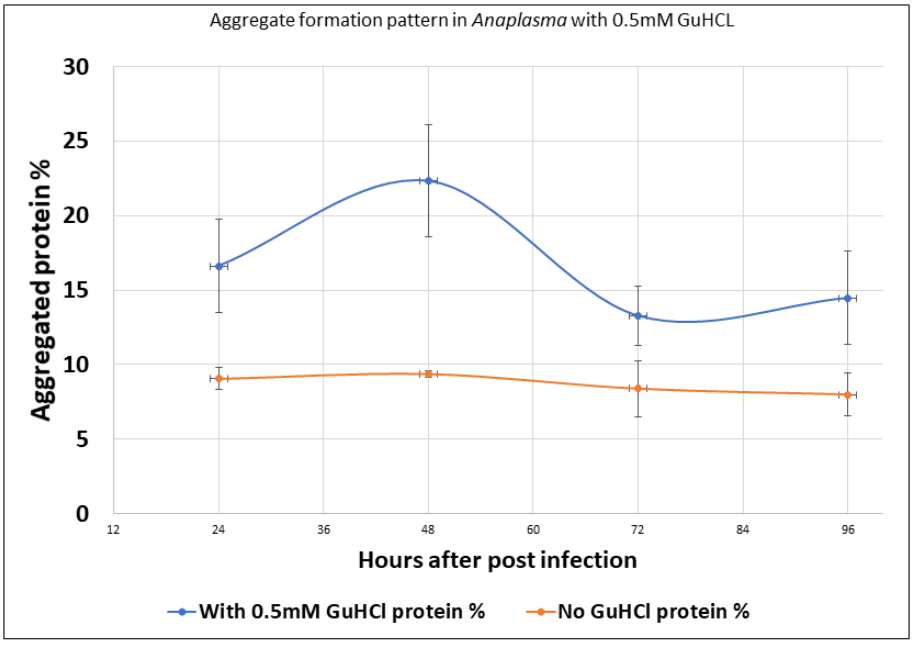
A: *A.Phagocytophilum* without GuHCl | *A.Phagocytophilum* with 0.5mM GuHCl



24S 48S 72S 96S | 24A 48A 72A 96A | 24S 48S 72S 96S | 24A 48A 72A 96A

S=Soluble fraction
A=Aggregated fraction

B:



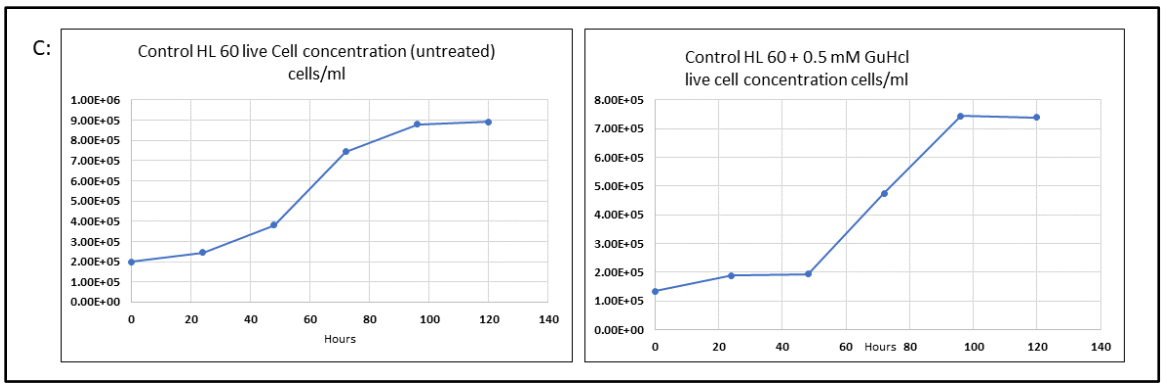


Figure 4.6. Mammalian cell environment induces protein aggregation in *A. phagocytophilum*.

A: Host cell free *A.*phagocytophilum were isolated at different post infection (hours) time points. Soluble and aggregated protein fractions were isolated as described earlier. Total protein concentration was measured by Bradford method and equal amounts of proteins were loaded and stained with silver staining method. BLUeye Prestained Protein Ladder is shown here in a 10% SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) gel. Refer to appendix B , figure B.10 for full details of the molecular weight of each lane.

B: Percentage of aggregated protein is shown here with or without GuHCl at different time points. Average value of two independent experiments were used to plot the graph and error bar indicates Standard deviation.

C: HL60 host cell growth was monitored with or without 0.5 mM GuHCl at different time points. Average values of two independent experiments were used to plot the graph. Cell counting was done using Invitrogen Countess 3 automated cell counter following supplier instructions (ThermoFisher Scientific, USA).

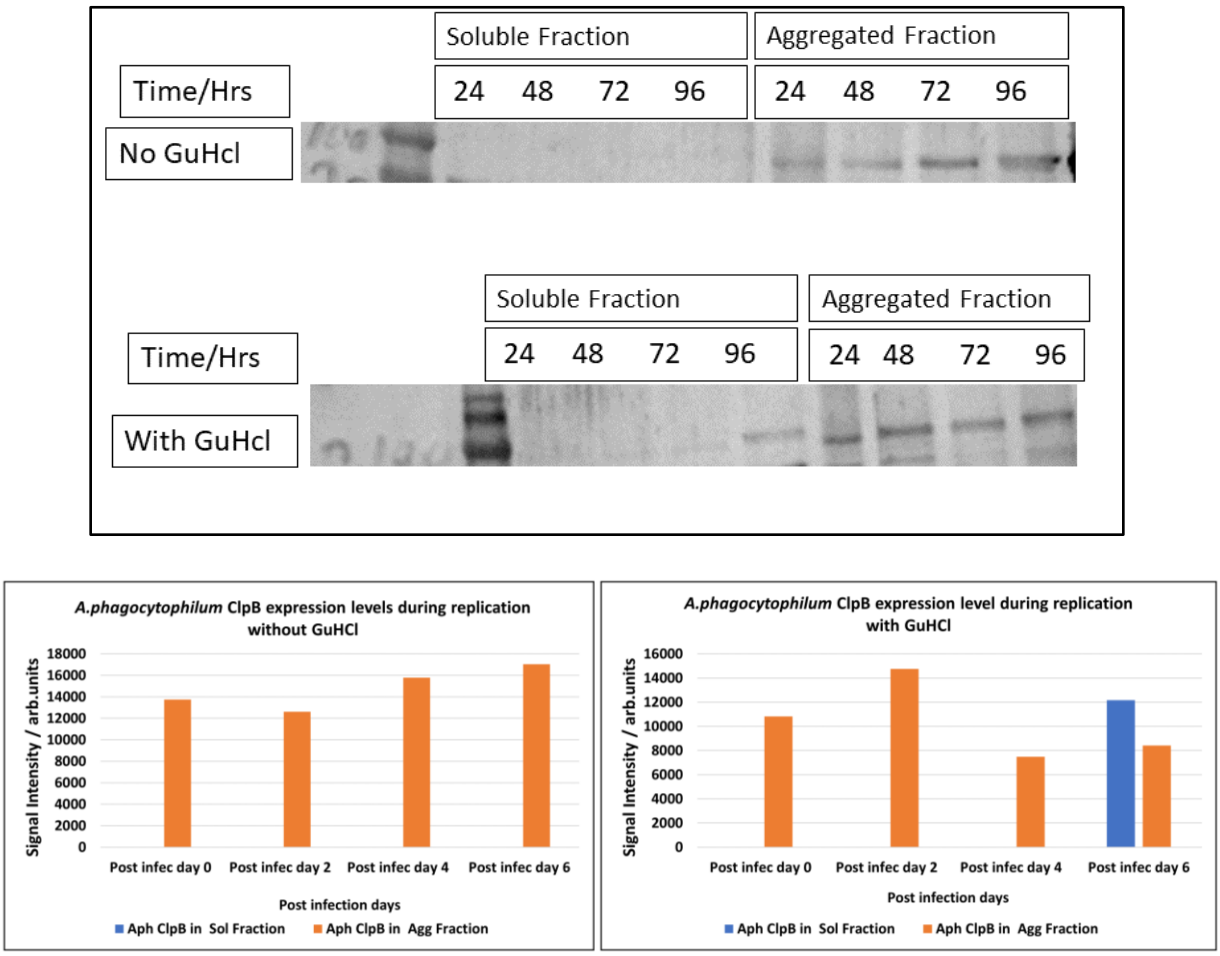
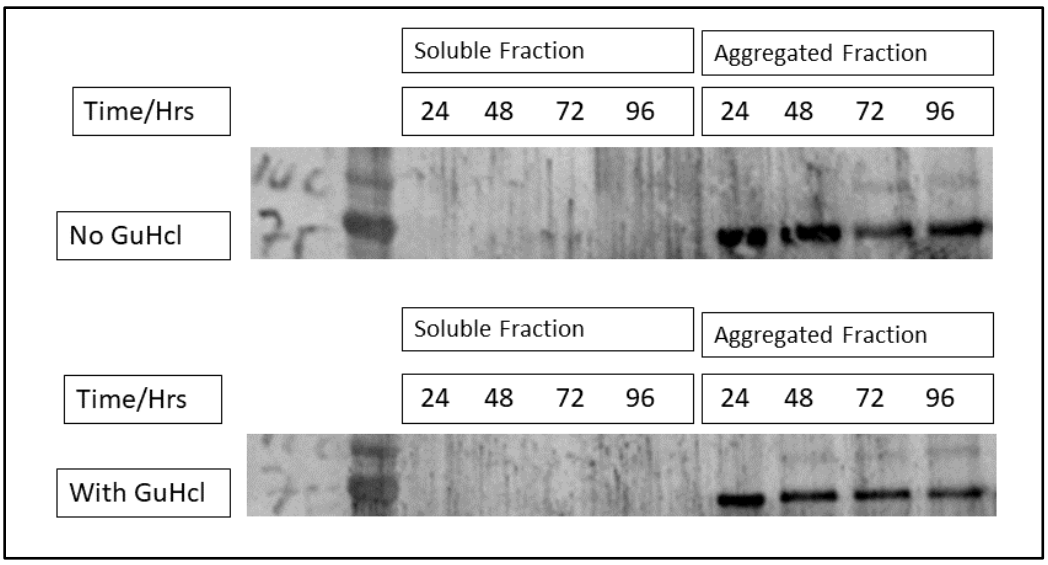
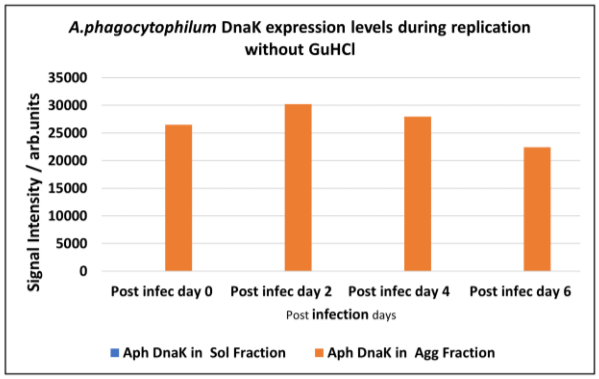


Figure 4.7. ClpB accumulation in *A. phagocytophilum* with GuHCl treatment.

Bacterial were grown without GuHCl or with 0.5 mM GuHCl in T25 flasks. Host cell free bacterial isolated and lysed as described in materials and methods. Soluble and aggregated fractions were isolated at different post infection time points. Immunodetection of ClpB (~95 kDa) carried out as mentioned in materials and methods. Part of protein ladder which shows (~100 kDa) is shown in the figure. Equal amount of proteins were loaded in each well. Image quantification (signal intensity) performed using ImageJ program from National Institutes of Health (NIH). If data is absent on any day, indicates that the signal was too low for western blot detection and quantification.





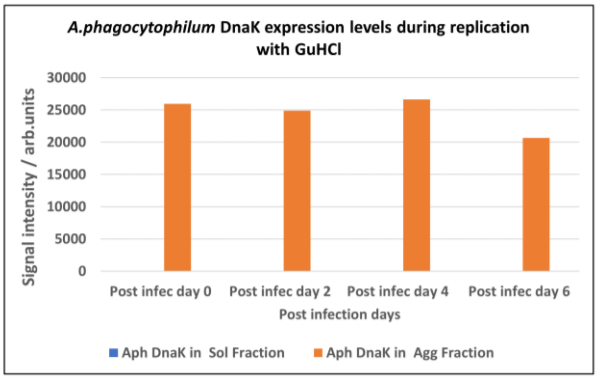
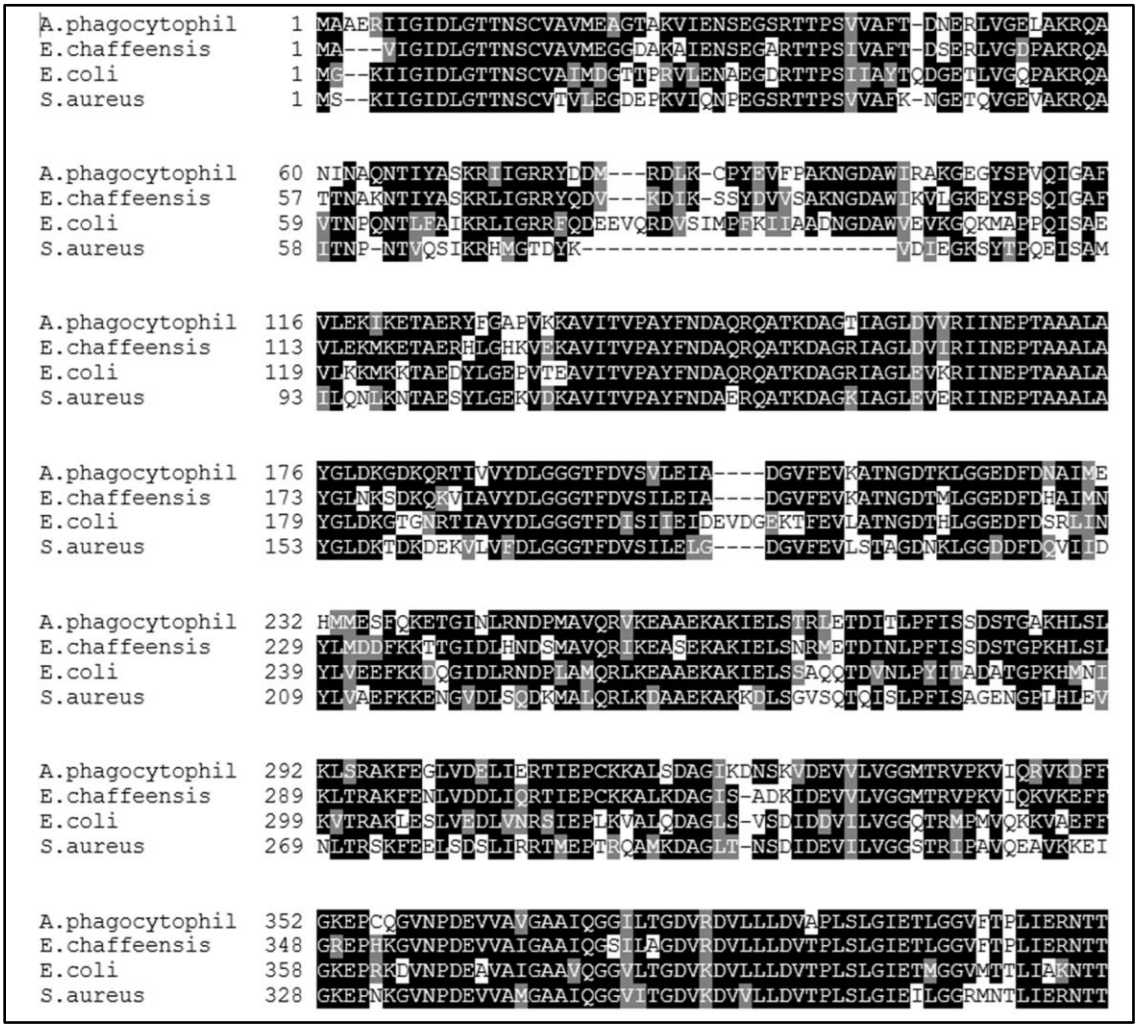


Figure 4.8. DnaK accumulation in *A. phagocytophilum* with GuHCl treatment.

The same samples (samples used in figure 4.7) were used again to detect DnaK. Bacteria were grown without GuHCl or with 0.5 mM GuHCl in T25 flasks. Host cell free bacterial isolated and lysed as described in materials and methods. Soluble and aggregated fractions were isolated at different post infection time points. Immunodetection of DnaK (~72 kDa) carried out as mentioned in materials and methods. Part of protein ladder which shows (~100 kDa, ~75 KDa) is shown in the figure. Equal amount of proteins were loaded in each well. Image quantification (signal intensity) performed using ImageJ program from National Institutes of Health (NIH). If

data is absent on any day, indicates that the signal was too low for western blot detection and quantification.



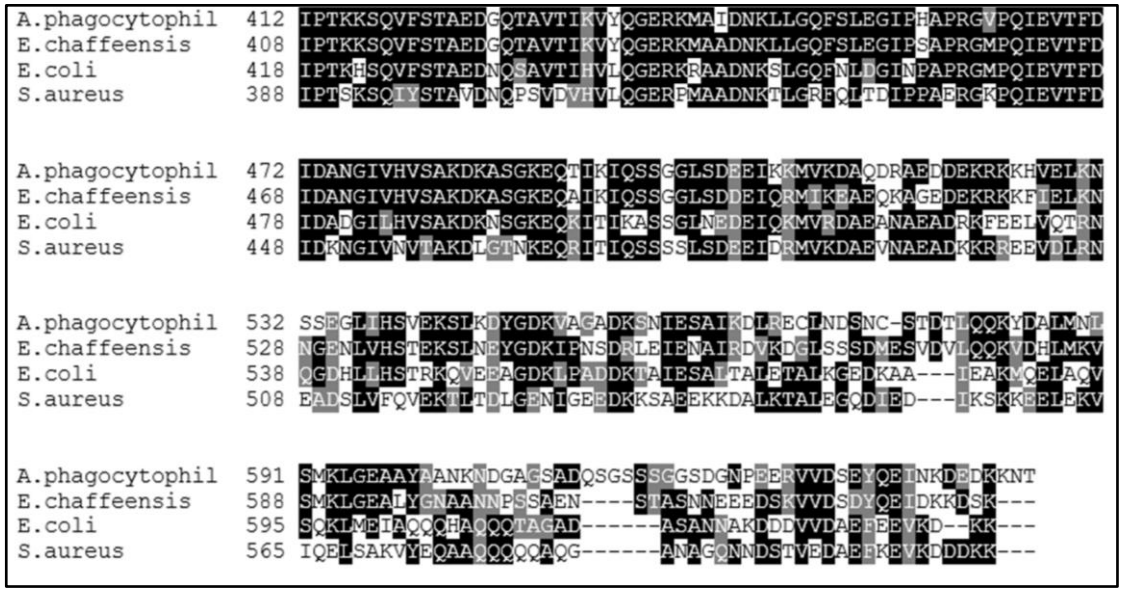


Figure 4.9. Sequence alignment of DnaK proteins from several human pathogens.

DnaK sequence from different pathogens have been compared using Clustal Omega tool provided by Bioinformatics Institute (EMBL-EBI, https://www.ebi.ac.uk/Tools/msa/clustalo/), followed by BOXSHADE 3.21 tool, from the Swiss Institute of Bioinformatics (https://embnet.vital-it.ch/software/BOX_form.html) which shaded multiple alignments. Black shaded boxes indicate conserved residues, grey indicates similar amino acids and white indicates different amino acids. (-) indicates the gaps introduce by Clustal Omega program when aligning these sequences. Full details of DnaK sequence sources are listed below. Amino acid sequences are conserved to a greater extent in both DnaK and ClpB of *Anaplasma Phagocytophilum* (refer to appendix B, figure B.7 for ClpB sequences.) All key residues which are essential for substrate recognition and binding, ATP binding and hydrolysis are conserved in both *Anaplasma C*lpB and DnaK.

A. phagocytophilum tr|Q2GKZ8|Q2GKZ8_ANAPZ Chaperone protein DnaK OS=*Anaplasma phagocytophilum* (strain HZ)

E. chaffeensis tr|Q2GGZ6|Q2GGZ6_EHRCR Chaperone protein DnaK OS=*Ehrlichia chaffeensis* (strain ATCC CRL-10679 / Arkansas)

E. coli sp|P0A6Y8|DNAK_ECOLI Chaperone protein DnaK OS=*Escherichia coli* (strain K12)

S. aureus sp|Q5HFI0|DNAK_STAAC Chaperone protein DnaK OS=*Staphylococcus aureus* (strain COL)

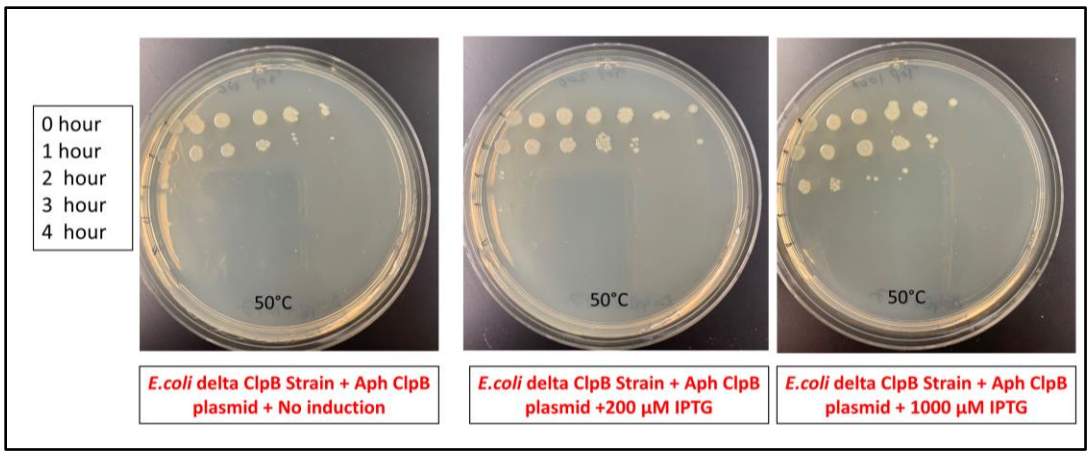


Figure 4.10. *In vivo* heat shock assay with *Anaplasma* recombinant ClpB in *E. coli* MC4100ΔclpB strain.

See appendix B figure B.2 for heat shock survival pattern of wild type *E. coli* cells (MC4100) and *E. coli* cell which lacks ClpB gene (*E. coli* MC4100ΔclpB). The *E. coli* MC4100ΔclpB strain had recombinant *Anaplasma* ClpB plasmid and plasmid pCS6 for production of T7 RNA polymerase. Transformed *E. coli* MC4100ΔclpB cells were grown under severe heat shock conditions at 50°C for the period indicated on the left (in hours) and then spotted on agar plates and incubated overnight at 37 °C. Each spot on the agar plates represents a viable culture after a 10-fold serial dilution (from left to right).

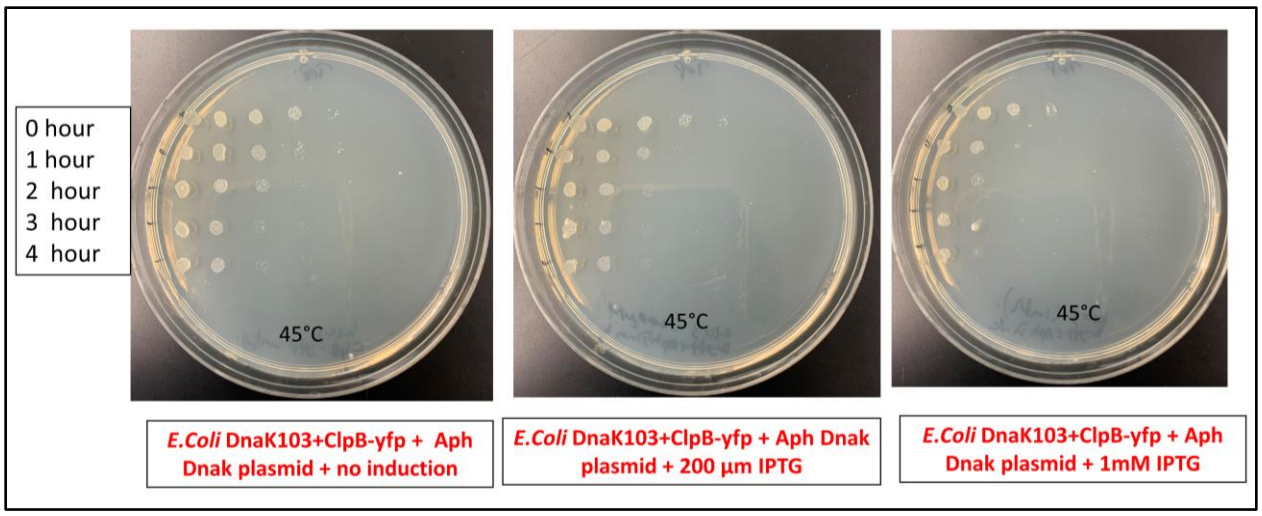


Figure 4.11. *Anaplasma* DnaK could not complement *E. coli* native DnaK for heat shock response.

The *E. coli* Dnak103 strain had recombinant *Anaplasma* DnaK plasmid ,plasmid pCS6 (for production of T7 RNA polymerase) and ClpB-YFP plasmid. Transformed *E. coli* DnaK103 cells grown under mild heat shock conditions at 45°C for the period indicated on the left (in hours) and then spotted on agar plates and incubated overnight at 37 °C. Each spot on the agar plates represents a viable culture after a 10-fold serial dilution (from left to right).

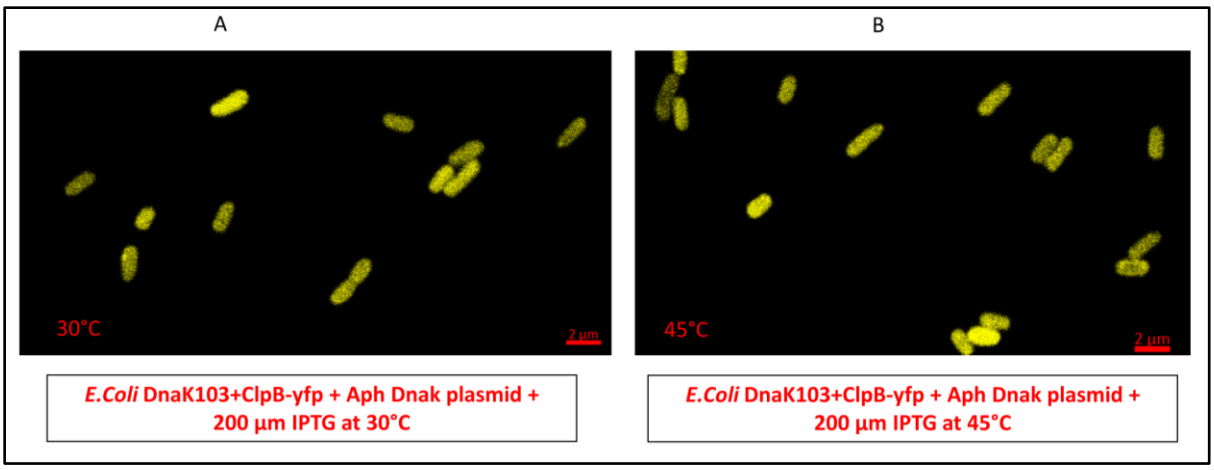
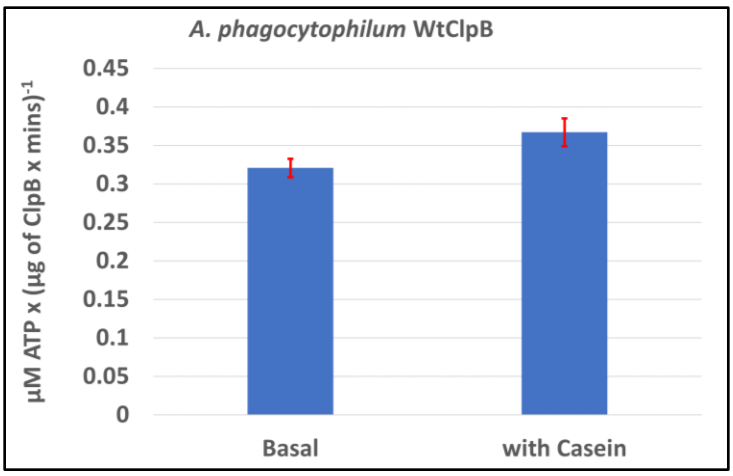
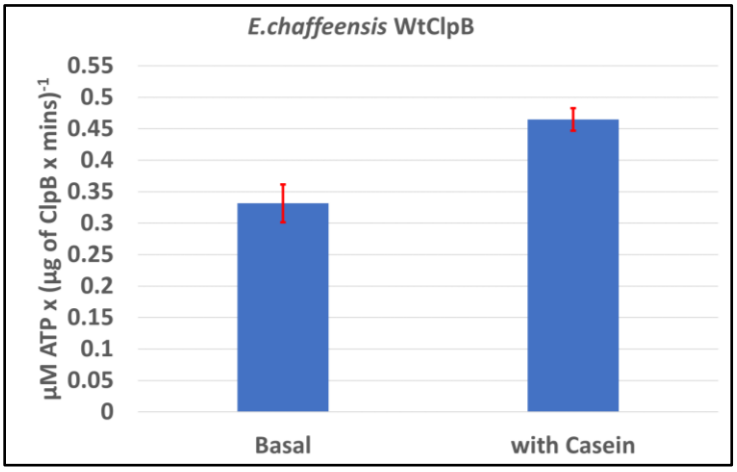
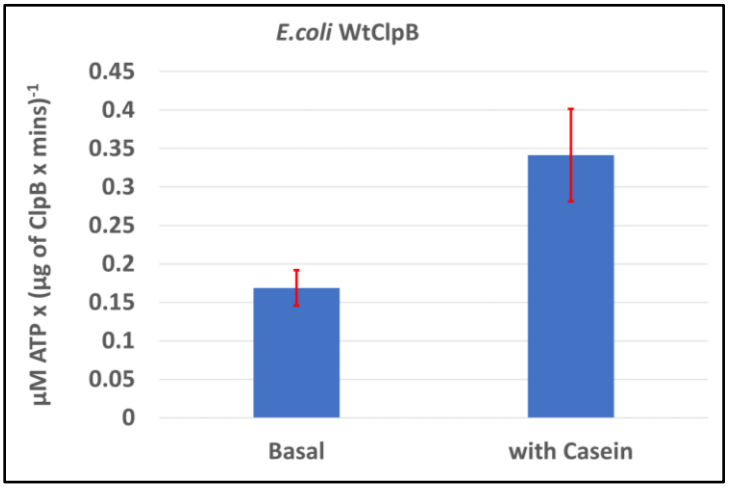


Figure 4.12. *Anaplasma* DnaK could not guide *E. coli* ClpB-YFP proteins towards aggregates.

We used same *E. coli* Dnak103 supplemented with *Anaplasma* DnaK plasmid ,plasmid pCS6 (for production of T7 RNA polymerase) and ClpB-YFP plasmid. These cells were subjected to a mild heat shock at 45°C for 30 minutes and used immediately for confocal image processing (see materials and methods). As shown in earlier publications (Winkler, Seybert et al. 2010, Glaza, Ranaweera et al. 2020)When there was no heat shock, ClpB-YFP was seen diffused throughout the cytoplasm. But when cells are under heat shock, ClpB-YFP got localized to the ends of the cell (where protein aggregates are located). Refer to Appendix B, figure B.5 for *E. coli* ClpB-YFP protein migration under heat shock. As *E. coli* Dnak103 cells were under mild heat shock, there was no migration of ClpB-YFP proteins towards the poles of these cells and ClpB-YFP was seen diffused throughout the cytoplasm even under mild heat shock conditions. The scale bar is located at bottom right-hand corner of each image and represents 2µm.


A. phagocytophilum WtClpB
µM ATP x (µg of ClpB x mins)-1
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Basal
with Casein


E.chaffeensis WtClpB
µM ATP x (µg of ClpB x mins)-1
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Basal
with Casein


E.coli WtClpB
µM ATP x (µg of ClpB x mins)-1
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Basal
with Casein

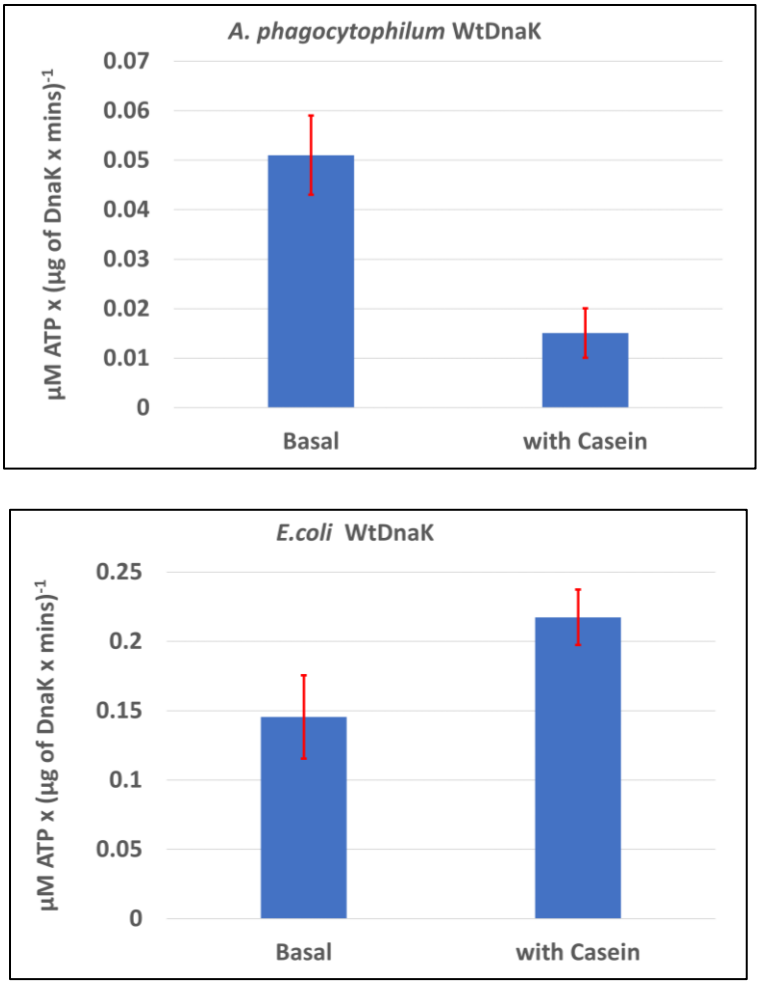


Figure 4.13. ATPase activity of purified *Anaplasma* ClpB and DnaK.

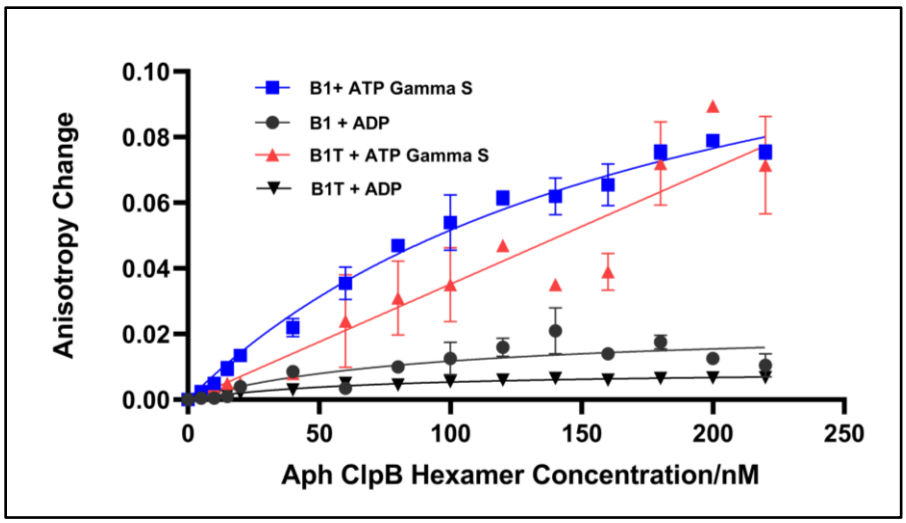
ATPase activity assay was performed as mentioned in materials and methods. For DnaK ATPase activity measurements, DnaJ and GrpE (co-chaperones) were included in the reaction mixture. Substrate stimulated ATPase activity increased (compared to basal activity) with all ClpB proteins tested. *E. coli* DnaK showed an increase in ATPase activity while *Anaplasma* DnaK showed a significant decrease in ATPase activity with the same substrate. Average values of two repeated experiments are shown here. Error bars indicates standard deviation.

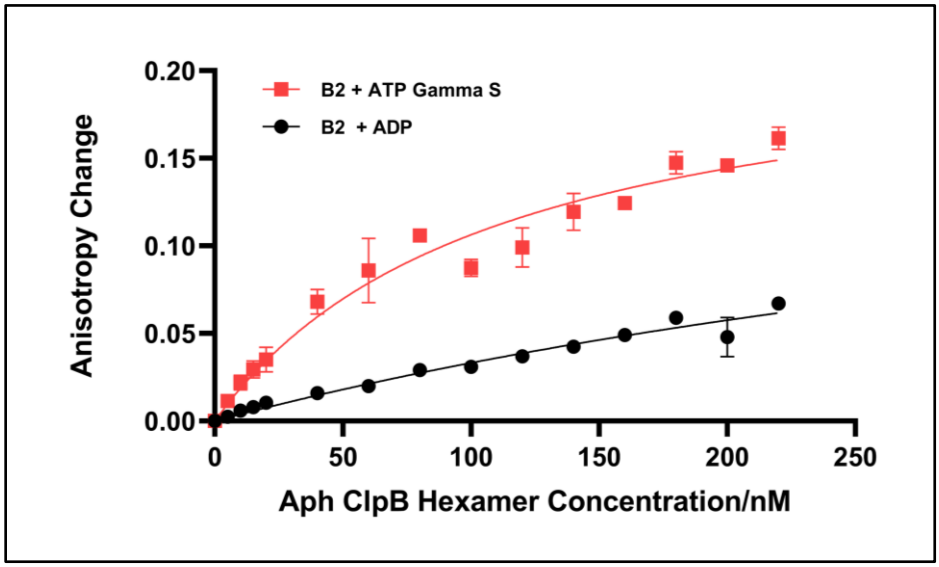
Table 4.1. Comparison of *Anaplasma* ClpB-peptide interactions with *E. coli* ClpB-peptide interactions at 25 °C.

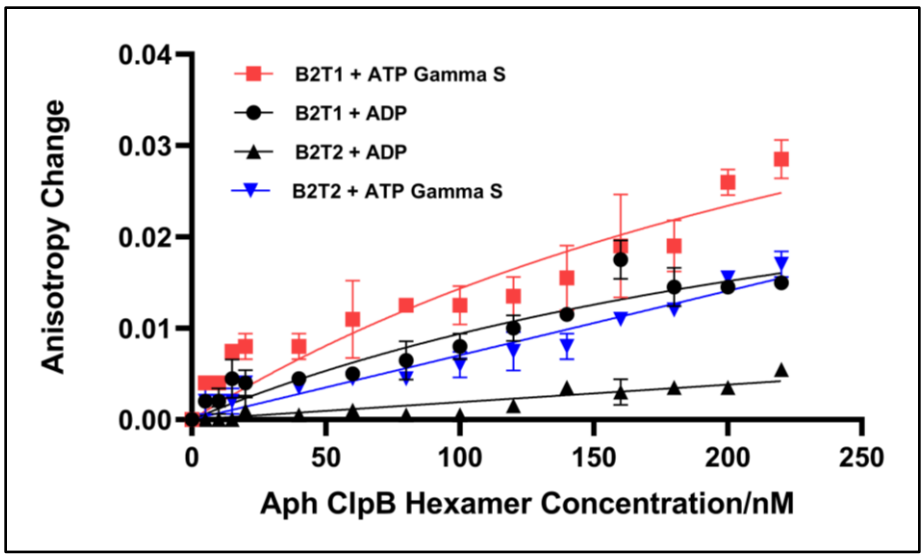
Peptides	***Anaplasma* ClpB** **Kd [nM]**	***E.coli* ClpB** **Kd [nM]**
B1	186.2	46.8±1.5
B1T	921.2	329 ±39
B1N	527.9	7.26 ±0.32
B2	110.7	13.24 ±0.42
B2T1	344.1	912 ±231
B2T2	Not determined	>1000
B2N	1706	17.54 ±0.92
B2A	425.9	10.17 ±0.80
B2B	126.4	11.51 ±0.52

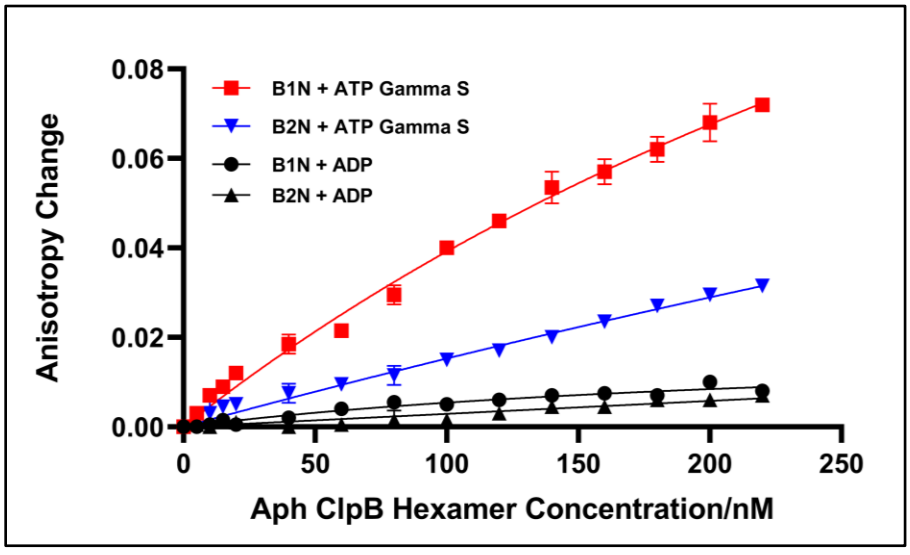
Table 4.2. Comparison of *Anaplasma* DnaK-peptide interactions with *E. coli* DnaK-peptide interactions at 25 °C.

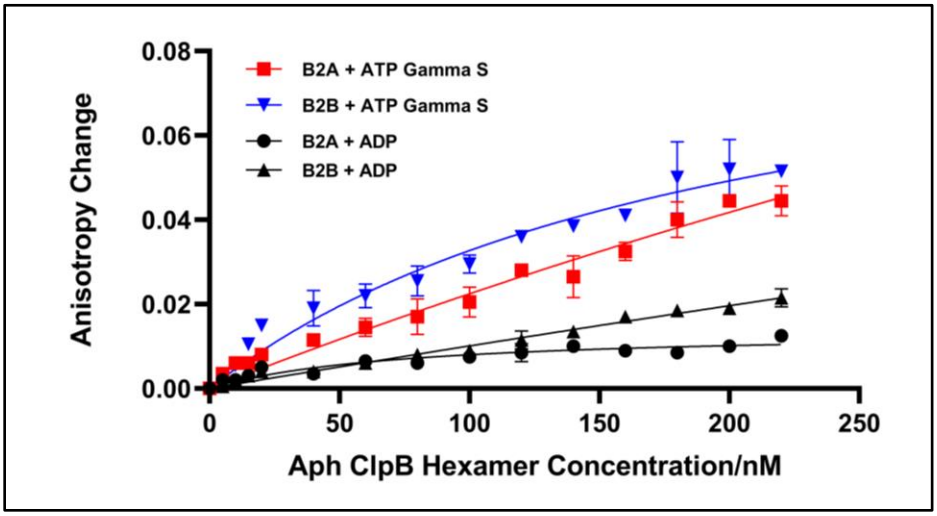
Peptides	*Anaplasma* DnaK Kd [nM]	*E.coli* DnaK Kd [nM]
B1	2542	157.6
B1T	84162	546.3
B1N	777.3	87.8
B2	606.2	126.7
B2T1	3486	498.3
B2T2	2199	513.2
B2N	1919	245.7
B2A	1138	251.6
B2B	1151	224.3

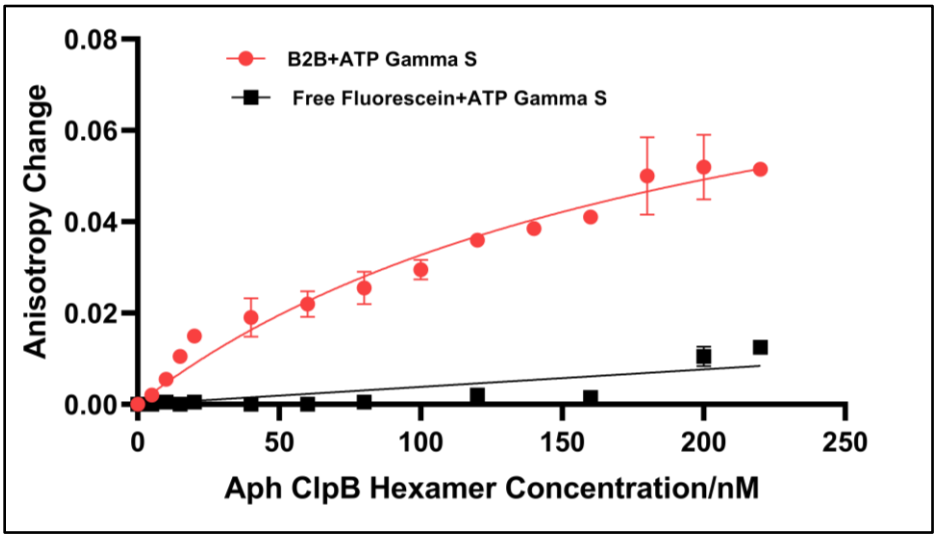

B1+ ATP Gamma S
B1 + ADP
B1T + ATP Gamma S
B1T + ADP
Anisotropy Change
0.10
0.08
0.06
0.04
0.02
0.00
0
50
100
150
200
250
Aph ClpB Hexamer Concentration/nM

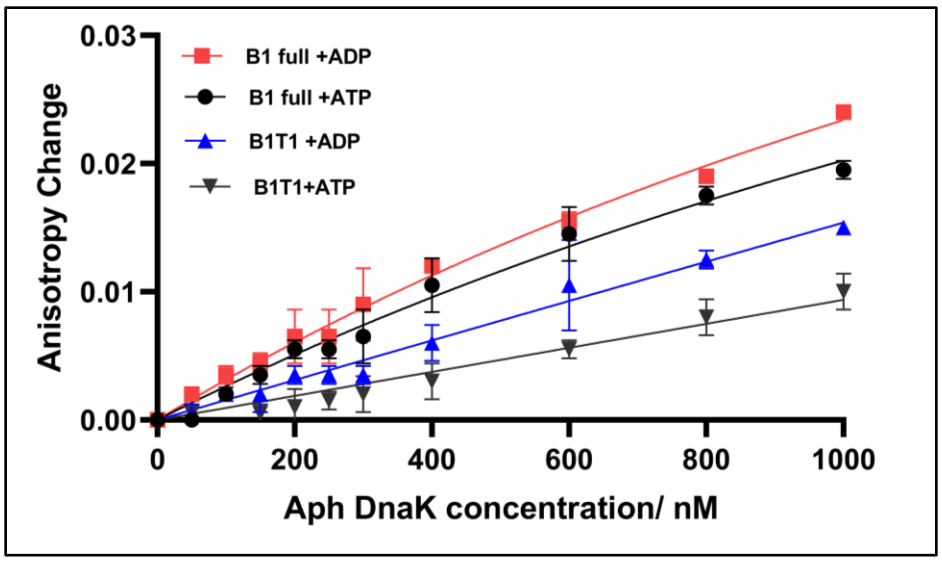

B2 + ATP Gamma S
B2 + ADP
Anisotropy Change
0.20
0.15
0.10
0.05
0.00
0
50
100
150
200
250
Aph ClpB Hexamer Concentration/nM

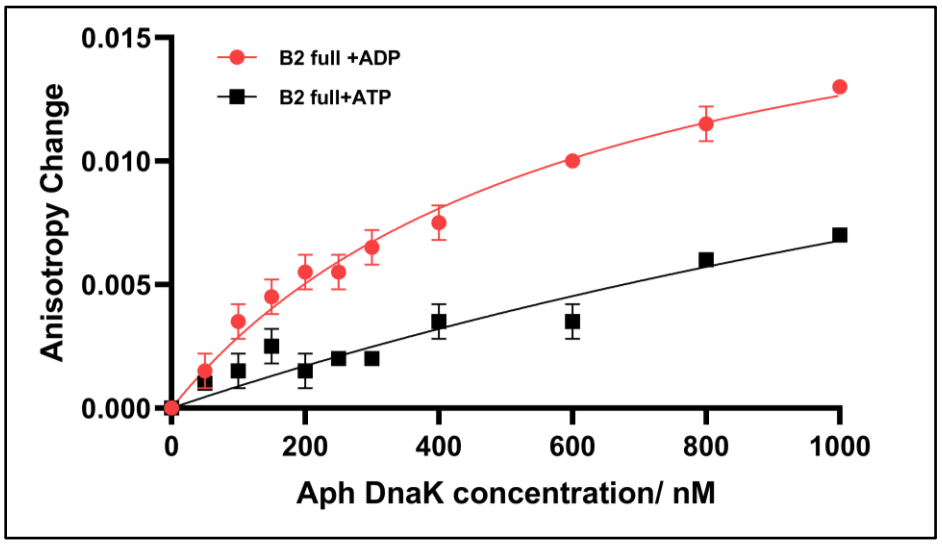

B2T1 + ATP Gamma S
B2T1 + ADP
B2T2 + ADP
B2T2 + ATP Gamma S
Anisotropy Change
0.04
0.03
0.02
0.01
0.00
0
50
100
150
200
250
Aph ClpB Hexamer Concentration/nM

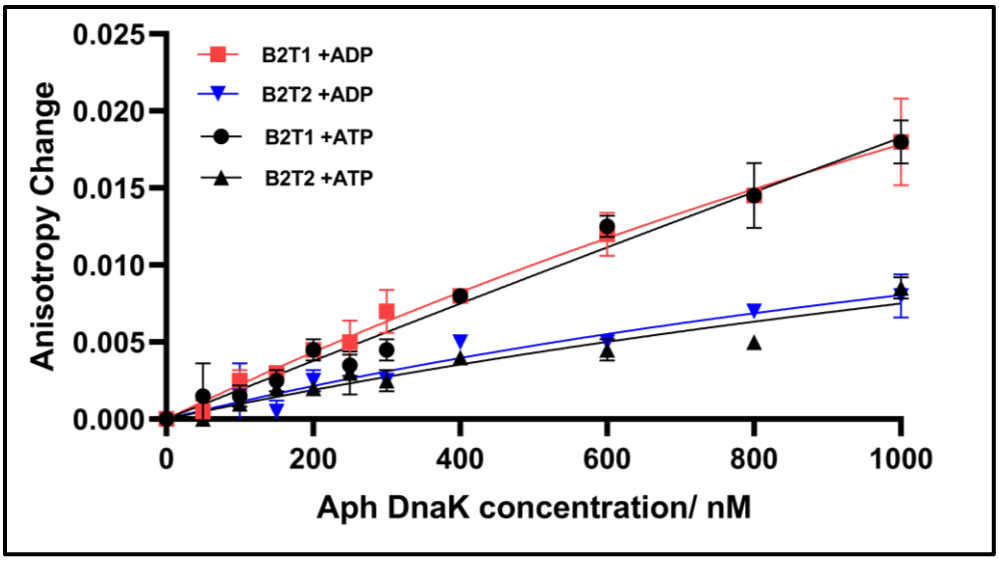

B1N + ATP Gamma S
B2N + ATP Gamma S
B1N + ADP
B2N + ADP
Anisotropy Change
0.08
0.06
0.04
0.02
0.00
0
50
100
150
200
250
Aph ClpB Hexamer Concentration/nM

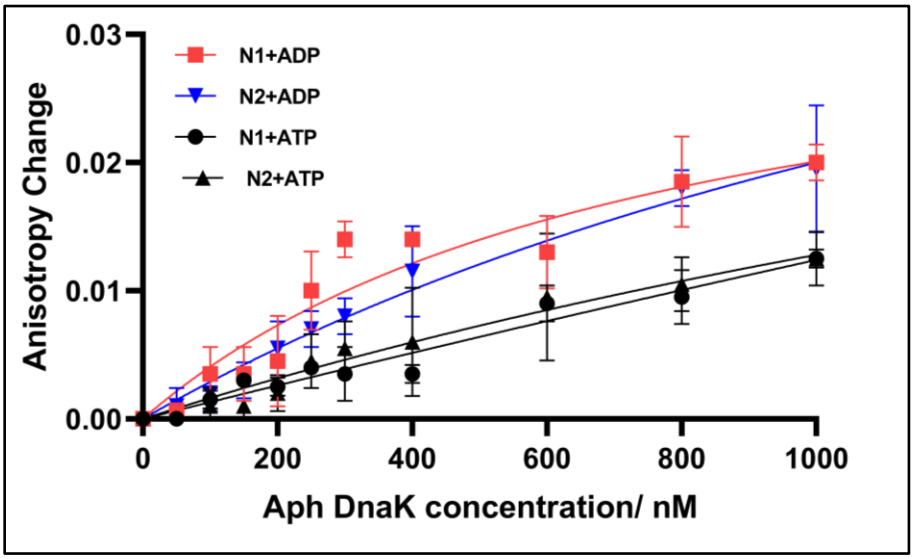

Anisotropy Change
0.08
0.06
0.04
0.02
0.00
B2A + ATP Gamma S
B2B + ATP Gamma S
B2A + ADP
B2B + ADP
0
50
100
150
200
250
Aph ClpB Hexamer Concentration/nM

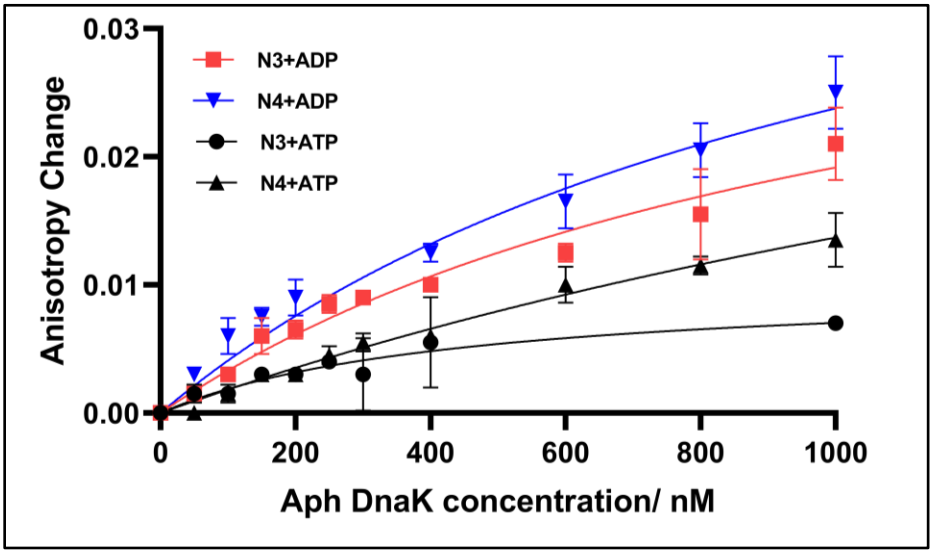

Anisotropy Change
0.08
0.06
0.04
0.02
0.00
B2B+ATP Gamma S
Free Fluorescein+ATP Gamma S
0
50
100
150
200
250
Aph ClpB Hexamer Concentration/nM


0.03
0.02
0.01
0.00
Anisotropy Change
B1 full +ADP
B1 full +ATP
B1T1 +ADP
B1T1+ATP
0
200
400
600
800
1000
Aph DnaK concentration/ nM


0.015
0.010
0.005
0.000
Anisotropy Change
B2 full +ADP
B2 full+ATP
0
200
400
600
800
1000
Aph DnaK concentration/ nM


0.025
0.020
0.015
0.010
0.005
0.000
Anisotropy Change
B2T1 +ADP
B2T2 +ADP
B2T1 +ATP
B2T2 +ATP
0
200
400
600
800
1000
Aph DnaK concentration/ nM


0.03
0.02
0.01
0.00
Anisotropy Change
N1+ADP
N2+ADP
N1+ATP
N2+ATP
0
200
400
600
800
1000
Aph DnaK concentration/ nM



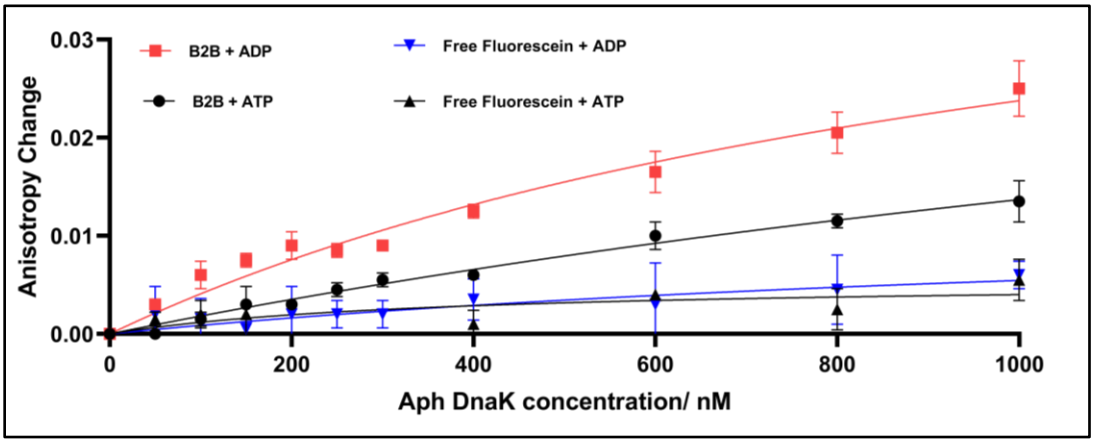


Figure 4.14. Binding isotherms for *Anaplasma* ClpB and DnaK with substrate mimicking peptides.

Purified *Anaplasma* ClpB, *Anaplasma* DnaK and *E. coli* DnaK were used for fluorescence anisotropy binding experiments. These experiments were performed as described with the same peptides mentioned in our earlier publication (Ranaweera, Glaza et al. 2018). Our data in vitro

binding experiments results demonstrated that generally *Anaplasma* ClpB and DnaK had weaker interactions with substrate mimicking peptides, except the interaction of *Anaplasma* ClpB with B2T1 peptide. Here Kd values represent apparent dissociation constants for the relevant protein (ClpB or DnaK) peptide interaction. *E. coli* ClpB-peptide Kd values (table 4.1) are from our previous publication (Ranaweera, Glaza et al. 2018) and *E. coli* DnaK binding isotherms and Kd values are shown in table 4.2 and Appendix B, figure B.6. As a control experiment free fluorescein was titrated with ClpB and ATP or with DnaK and ADP to demonstrate that the observed interaction with ClpB or DnaK was mediated by the peptide core and not by the fluorescein tag.

4.7 References.

Atif, F. A. (2015). "*Anaplasma marginale* and *Anaplasma phagocytophilum*: Rickettsiales pathogens of veterinary and public health significance." Parasitology Research **114**(11): 3941-3957.

Balchin, D., M. Hayer-Hartl and F. U. Hartl (2016). "In vivo aspects of protein folding and quality control." Science **353**(6294): aac4354.

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

Busby, A. T., N. Ayllón, K. M. Kocan, E. F. Blouin, G. de la Fuente, R. C. Galindo, M. Villar and J. de la Fuente (2012). "Expression of heat shock proteins and subolesin affects stress responses, *Anaplasma phagocytophilum* infection and questing behaviour in the tick, Ixodes scapularis." Med Vet Entomol **26**(1): 92-102.

Chen, S. M., J. S. Dumler, J. S. Bakken and D. H. Walker (1994). "Identification of a granulocytotropic *Ehrlichia* species as the etiologic agent of human disease." J Clin Microbiol **32**(3): 589-595.

Cheng, C. and R. R. Ganta (2008). "Laboratory Maintenance of *Ehrlichia chaffeensis* and *Ehrlichia canis* and Recovery of Organisms for Molecular Biology and Proteomics Studies." Current Protocols in Microbiology **9**(1): 3A.1.1-3A.1.21.

Cheng, C., A. D. S. Nair, V. V. Indukuri, S. Gong, R. F. Felsheim, D. Jaworski, U. G. Munderloh and R. R. Ganta (2013). "Targeted and random mutagenesis of *Ehrlichia chaffeensis* for the identification of genes required for in vivo infection." PLoS pathogens **9**(2): e1003171-e1003171.

Contreras, M., P. Alberdi, L. Mateos-Hernández, I. G. Fernández de Mera, A. L. García-Pérez, M. Vancová, M. Villar, N. Ayllón, A. Cabezas-Cruz, J. J. Valdés, S. Stuen, C. Gortazar and J. de la Fuente (2017). "*Anaplasma phagocytophilum* MSP4 and HSP70 Proteins Are Involved in Interactions with Host Cells during Pathogen Infection." Front Cell Infect Microbiol **7**: 307.

de la Fuente, J., S. Antunes, S. Bonnet, A. Cabezas-Cruz, A. G. Domingos, A. Estrada-Peña, N. Johnson, K. M. Kocan, K. L. Mansfield, A. M. Nijhof, A. Papa, N. Rudenko, M. Villar, P. Alberdi, A. Torina, N. Ayllón, M. Vancova, M. Golovchenko, L. Grubhoffer, S. Caracappa, A. R. Fooks, C. Gortazar and R. O. M. Rego (2017). "Tick-Pathogen Interactions and Vector Competence: Identification of Molecular Drivers for Tick-Borne Diseases." Frontiers in cellular and infection microbiology **7**: 114-114.

Doyle, S. M., S. Shastry, A. N. Kravats, Y. H. Shih, M. Miot, J. R. Hoskins, G. Stan and S. Wickner (2015). "Interplay between *E. coli* DnaK, ClpB and GrpE during protein disaggregation." J Mol Biol **427**(2): 312-327.

Dumler, J. S., K.-S. Choi, J. C. Garcia-Garcia, N. S. Barat, D. G. Scorpio, J. W. Garyu, D. J. Grab and J. S. Bakken (2005). "Human granulocytic anaplasmosis and *Anaplasma phagocytophilum*." Emerging infectious diseases **11**(12): 1828-1834.

English, C. A., W. Sherman, W. Meng and L. M. Gierasch (2017). "The Hsp70 interdomain linker is a dynamic switch that enables allosteric communication between two structured domains." The Journal of biological chemistry **292**(36): 14765-14774.

Farrell, H. M., Jr., P. X. Qi, E. M. Brown, P. H. Cooke, M. H. Tunick, E. D. Wickham and J. J. Unruh (2002). "Molten globule structures in milk proteins: implications for potential new structure-function relationships." J Dairy Sci **85**(3): 459-471.

Fernandez-Escamilla, A.-M., F. Rousseau, J. Schymkowitz and L. Serrano (2004). "Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins." Nature Biotechnology **22**(10): 1302-1306.

Fernández-Higuero, J. A., A. Aguado, J. Perales-Calvo, F. Moro and A. Muga (2018). "Activation of the DnaK-ClpB Complex is Regulated by the Properties of the Bound Substrate." Scientific Reports **8**(1): 5796.

Glaza, P., C. B. Ranaweera, S. Shiva, A. Roy, B. V. Geisbrecht, F. J. Schoenen and M. Zolkiewski (2020). "Repurposing p97 inhibitors for chemical modulation of the bacterial ClpB/DnaK bi-chaperone system." J Biol Chem.

Grimminger, V., K. Richter, A. Imhof, J. Buchner and S. Walter (2004). "The Prion Curing Agent Guanidinium Chloride Specifically Inhibits ATP Hydrolysis by Hsp104*." Journal of Biological Chemistry **279**(9): 7378-7383.

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

Haslberger, T., J. Weibezahn, R. Zahn, S. Lee, F. T. F. Tsai, B. Bukau and A. Mogk (2007). "M Domains Couple the ClpB Threading Motor with the DnaK Chaperone Activity." Molecular Cell **25**(2): 247-260.

Kim, K. I., G.-W. Cheong, S.-C. Park, J.-S. Ha, K. M. Woo, S. J. Choi and C. H. Chung (2000). "Heptameric ring structure of the heat-shock protein ClpB, a protein-activated ATPase in *Escherichia coli*1, Edited by W. Baumeister." Journal of Molecular Biology **303**(5): 655-666.

Kityk, R., J. Kopp, I. Sinning and M. P. Mayer (2012). "Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones." Mol Cell **48**(6): 863-874.

Kuczynska-Wisnik, D., C. Cheng, R. R. Ganta and M. Zolkiewski (2017). "Protein aggregation in *Ehrlichia chaffeensis* during infection of mammalian cells." FEMS microbiology letters **364**(6): fnx059.

Kuczynska-Wisnik, D., C. Cheng, R. R. Ganta and M. Zolkiewski (2017). "Protein aggregation in *Ehrlichia chaffeensis* during infection of mammalian cells." FEMS Microbiol Lett **364**(6).

Kwiatkowska, J., E. Matuszewska, D. Kuczyńska-Wiśnik and E. Laskowska (2008). "Aggregation of *Escherichia coli* proteins during stationary phase depends on glucose and oxygen availability." Res Microbiol **159**(9-10): 651-657.

Lin, M. and Y. Rikihisa (2003). "*Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* lack genes for lipid A biosynthesis and incorporate cholesterol for their survival." Infection and Immunity **71**(9): 5324.

Liu, Z., V. Tek, V. Akoev and M. Zolkiewski (2002). "Conserved Amino Acid Residues within the Amino-terminal Domain of ClpB are Essential for the Chaperone Activity." Journal of Molecular Biology **321**(1): 111-120.

McClure, E. E., A. S. O. Chávez, D. K. Shaw, J. A. Carlyon, R. R. Ganta, S. M. Noh, D. O. Wood, P. M. Bavoil, K. A. Brayton, J. J. Martinez, J. W. McBride, R. H. Valdivia, U. G. Munderloh and J. H. F. Pedra (2017). "Engineering of obligate intracellular bacteria: progress, challenges and paradigms." Nat Rev Microbiol **15**(9): 544-558.

Naimi, W. A., J. J. Gumpf, R. S. Green, J. R. Izac, M. P. Zellner, D. H. Conrad, R. T. Marconi, R. K. Martin and J. A. Carlyon (2020). " Immunization against *Anaplasma phagocytophilum* Adhesin Binding Domains Confers Protection against Infection in the Mouse Model." Infection and Immunity **88**(10): e00106-00120.

Ojogun, N., A. Kahlon, S. A. Ragland, M. J. Troese, J. E. Mastronunzio, N. J. Walker, L. VieBrock, R. J. Thomas, D. L. Borjesson, E. Fikrig and J. A. Carlyon (2012). " *Anaplasma phagocytophilum* outer membrane protein A interacts with sialylated glycoproteins to promote infection of mammalian host cells." Infection and Immunity **80**(11): 3748-3760.

Parola, P. and D. Raoult (2001). "Ticks and Tickborne Bacterial Diseases in Humans: An Emerging Infectious Threat." Clinical Infectious Diseases **32**(6): 897-928.

Ranaweera, C. B., P. Glaza, T. Yang and M. Zolkiewski (2018). "Interaction of substrate-mimicking peptides with the AAA+ ATPase ClpB from *Escherichia coli*." Archives of Biochemistry and Biophysics **655**: 12-17.

Rikihisa, Y. (2011). " Mechanisms of obligatory intracellular infection with *Anaplasma phagocytophilum* " Clinical Microbiology Reviews **24**(3): 469.

Rosenzweig, R., S. Moradi, A. Zarrine-Afsar, J. R. Glover and L. E. Kay (2013). "Unraveling the Mechanism of Protein Disaggregation Through a ClpB-DnaK Interaction." Science **339**(6123): 1080.

Rousseau, F., J. Schymkowitz and L. Serrano (2006). "Protein aggregation and amyloidosis: confusion of the kinds?" Current Opinion in Structural Biology **16**(1): 118-126.

Saibil, H. R. (2013). "Machinery to Reverse Irreversible Aggregates." Science **339**(6123): 1040.

Schmidt, C. M., D. L. Shis, T. D. Nguyen-Huu and M. R. Bennett (2012). "Stable maintenance of multiple plasmids in *E. coli* using a single selective marker." ACS Synth Biol **1**(10): 445-450.

Sirigireddy, K. R. and R. R. Ganta (2005). "Multiplex Detection of *Ehrlichia* and *Anaplasma* Species Pathogens in Peripheral Blood by Real-Time Reverse Transcriptase-Polymerase Chain Reaction." The Journal of Molecular Diagnostics **7**(2): 308-316.

Spence, J., A. Cegielska and C. Georgopoulos (1990). "Role of *Escherichia coli* heat shock proteins DnaK and HtpG (C62.5) in response to nutritional deprivation." Journal of Bacteriology **172**(12): 7157.

Squires, C. L., S. Pedersen, B. M. Ross and C. Squires (1991). "ClpB is the *Escherichia coli* heat shock protein F84.1." Journal of Bacteriology **173**(14): 4254.

Stewart, G. R. and D. B. Young (2004). "Heat-shock proteins and the host-pathogen interaction during bacterial infection." Curr Opin Immunol **16**(4): 506-510.

Swain, J. F., G. Dinler, R. Sivendran, D. L. Montgomery, M. Stotz and L. M. Gierasch (2007). "Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker." Mol Cell **26**(1): 27-39.

Troese, M. J. and J. A. Carlyon (2009). "*Anaplasma phagocytophilum* dense-cored organisms mediate cellular adherence through recognition of human P-selectin glycoprotein ligand 1." Infection and immunity **77**(9): 4018-4027.

Troese, M. J. and J. A. Carlyon (2009). " *Anaplasma phagocytophilum* dense-cored organisms mediate cellular adherence through recognition of human P-selectin glycoprotein ligand 1." Infection and Immunity **77**(9): 4018-4027.

Ung, P. M.-U., A. D. Thompson, L. Chang, J. E. Gestwicki and H. A. Carlson (2013). "Identification of Key Hinge Residues Important for Nucleotide-Dependent Allostery in *E. coli* Hsp70/DnaK." PLOS Computational Biology **9**(11): e1003279.

Villar, M., N. Ayllón, A. T. Busby, R. C. Galindo, E. F. Blouin, K. M. Kocan, E. Bonzón-Kulichenko, Z. Zivkovic, C. Almazán, A. Torina, J. Vázquez and J. de la Fuente (2010). " Expression of Heat Shock and Other Stress Response Proteins in Ticks and Cultured Tick Cells in Response to *Anaplasma* spp. Infection and Heat Shock." International Journal of Proteomics **2010**: 657261.

Wickner, S., J. L. Camberg, S. M. Doyle and D. M. Johnston (2017). Molecular Chaperones. Reference Module in Life Sciences, Elsevier.

Winkler, J., A. Seybert, L. König, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A. S. Frangakis, A. Mogk and B. Bukau (2010). "Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing." The EMBO journal **29**(5): 910-923.

Winkler, J., J. Tyedmers, B. Bukau and A. Mogk (2012). "Hsp70 targets Hsp100 chaperones to substrates for protein disaggregation and prion fragmentation." Journal of Cell Biology **198**(3): 387-404.

Zeymer, C., N. D. Werbeck, I. Schlichting and J. Reinstein (2013). "The Molecular Mechanism of Hsp100 Chaperone Inhibition by the Prion Curing Agent Guanidinium Chloride." Journal of Biological Chemistry **288**(10): 7065-7076.

Zhang, T., S. Kedzierska-Mieszkowska, H. Liu, C. Cheng, R. R. Ganta and M. Zolkiewski (2013). "Aggregate-Reactivation Activity of the Molecular Chaperone ClpB from *Ehrlichia chaffeensis*." PLOS ONE **8**(5): e62454.

Zhang, Y., L. Chen, C. Kondethimmanahalli, H. Liu and R. R. Ganta (2021). " Protein and DNA Biosynthesis Demonstrated in Host Cell-Free Phagosomes Containing *Anaplasma phagocytophilum* or *Ehrlichia chaffeensis* in Axenic Media." Infection and Immunity: IAI.00638-00620.

Zhuravleva, A., E. M. Clerico and L. M. Gierasch (2012). "An interdomain energetic tug-of-war creates the allosterically active state in Hsp70 molecular chaperones." Cell **151**(6): 1296-1307.

Zhuravleva, A. and L. M. Gierasch (2011). "Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones." Proceedings of the National Academy of Sciences **108**(17): 6987.

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

Chapter 5 - Role of ClpB in proliferation of obligate intra cellular pathogens , *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum*.

5.1 Introduction.

Ehrlichia chaffeensis and *Anaplasma phagocytophilum* are Gram negative, obligate intracellular pathogens. Both organisms lack the genes for peptidoglycan and lipopolysaccharides synthesis (Rikihisa 2010). *A. phagocytophilum* and *E. chaffeensis*, both have relatively small genomes (1.2-1.5 Mb in size). These organisms have lost many redundant genes via reductive evolutionary process and maintain only the most essential genes. Both *A. phagocytophilum* and *E. chaffeensis* maintain all the genes necessary for aerobic respiration, genes for biosynthesis of most vitamins and cofactors and genes for the nucleotides production etc. (Rikihisa and Lin 2010). *E. chaffeensis* causes Human Monocytic Ehrlichiosis (HME) and infects monocytes and macrophages. HME is a vector borne disease transmitted to humans by *Amblyomma Americanum* tick bites. Humans, deer and dogs act as host for *E. chaffeensis* whereas *Amblyomma Americanum* is considered as the reservoir (Rikihisa and Lin 2010, Wang, Nair et al. 2020).

A. phagocytophilum causes Human Granulocytic Anaplasmosis (HGA) and affects granulocytes and endothelial cells. The infection of a host occurs by tick bites of *Ixodes* species ticks (*Ixodes scapularis, Ixodes pacificus, Ixodes ricinus*). HGA is initiated through bites of infected *Ixodes* ticks during a blood meal (Rikihisa 2010, Zhang, Chen et al. 2021). The severity of both HGE and HGA varies and can range from asymptomatic to severe morbidity or death

(Ismail, Bloch et al. 2010). The life cycle of *Anaplasma* species and *Ehrlichia* species is shown in figure (5.5)(McClure, Chávez et al. 2017).

Once suitable eukaryotic host cell is encountered by *Anaplasma* species and *Ehrlichia* species, bacteria are engulfed by endocytosis (step 1 in figure 5..5). Then, the bacteria undergo cyclical rounds of infectious and replicative forms inside the host cell. The infectious form is known as Dense Core (DC) cells and the replicative form (noninfectious form) is known as Reticulate Cells (RC). Once internalized, the DC cells differentiate into Reticulate Cells (step 2 in figure 5.5) in a vacuole and the bacteria in these inclusions are known as morula. The RC cells multiply by a binary fission inside these vacuoles (step 3 in figure 5.5). Once the RC cells complete proliferation, they differentiate back to DC cells and exit the host cells by lysis (step 4 in figure 5.5). In the presence of a suitable eukaryotic host cell, the Dense Core cells become again internalized (step 1 in figure 5.5) and the cycle repeats.

The general life cycle of *Ixodes* species ticks is explained in chapter 4, hence the life cycle of *Amblyomma Americanum* tick is explained below. *Amblyomma* ticks have three feeding stages (larval, nymph, and adult); each developmental stage feeds only once (Ismail, Bloch et al. 2010). Generally, infected ticks would transmit *E. chaffeensis* to mammals during a blood meal (figure 5.6). If these mammals are already infected with *Ehrlichia*, then an uninfected tick could acquire *Ehrlichia* during a blood meal. *Ehrlichia is* found during the later larval, nymph or adult stage of tick life cycle. So far there are no evidence to indicate that bacteria are transmitted from infected adults to tick eggs or from eggs to the larval stage. As adults will lay eggs in the environment and as the eggs hatch, larvae will emerge. In order to survive, the larvae must get a blood meal from a suitable host for their further development into a nymph. Once they acquire a blood meal from a host, these larvae will drop off from the host and will prepare themselves to become nymphs. The

nymphs will grow into adult ticks that will lay eggs and this cycle repeats. The transmission of *Ehrlichia* happens during blood meal (either from infected tick to an uninfected host or from infected host to an uninfected tick). Humans are considered incidental hosts for *Ehrlichia* and *Anaplasma* species (Ismail, Bloch et al. 2010).

The number of HGA and HGE cases have increased rapidly since the diseases became reportable. According to the latest figures from Centers for Disease Control and Prevention (CDC, accessed on 03/07/2021), HGE fatality rate is around 1% while HGA fatality rate is less than 1% in the USA.

(www.cdc.gov/ehrlichiosis/stats/index.html www.cdc.gov/anaplasmosis/stats/index.html).

Despite rapid increase in occurrence of HGE and HGA cases, doxycycline remains the only antibiotic treatment against these infections. Delayed diagnosis in older patients and in immunocompromised individuals could expose them to adverse outcomes, including death (Bakken and Dumler 2006).

Genetic tools available for genome manipulations in obligate intracellular pathogens are limited and genetic manipulations become even harder as most of the genes these organisms harbor seem to be critical for their survival. As a result, relatively little is known and understood about the physiology, host pathogen interaction, and virulence factor expression of obligate intracellular pathogens. Moreover, there are no vaccines available for HGA or HGE. (McClure, Chávez et al. 2017, Wang, Nair et al. 2020).

Previously, we have shown that a compound known as DBeQ (N^2,N^4-dibenzylquinazoline-2,4-diamine, figure 5.7) was active against wild type *E. coli* (Glaza, Ranaweera et al. 2021). We proposed that the main target of DBeQ is the ClpB–DnaK bichaperone system in *E. coli*. We observed that DBeQ inhibited aggregate reactivation by the ClpB–DnaK

bichaperone system with an IC_{50}~5 μM and suppressed the growth of cultured *Escherichia coli* (figure 5.8 and Chapter 3). As antibacterial activity of DBeQ was quite potent against an extracellular organism, we pursued the study of DBeQ effects on the growth of obligate intracellular pathogens *A. phagocytophilum* and *E. chaffeensis*.

As DBeQ function in *E. coli* is linked to the ClpB expression and the ClpB function is closely related to aggregate reactivation process and maintaining cellular homeostasis, we decided to study aggregate formation patterns in obligate intracellular bacteria first and then investigate the DBeQ effects. Hence, we first characterized aggregate formation pattern in *A. phagocytophilum*, which was described in detail in Chapter 4. Analogous aggregate formation studies for wild type *E. chaffeensis* and its strain with ClpB deleted (*E. chaffeensis* ΔClpB, or TR91 strain) are described in the first part of this chapter, followed by the studies on DBeQ effects in *A. phagocytophilum* and *E. chaffeensis* later.

5.2 Materials and Methods.

5.2.1 Cultivation and isolation of *Anaplasma phagocytophilum* in HL60 cells.

The human promyelocytic cell lines HL-60 (ATCC CCL-240, Manassas, VA) and *Anaplasma phagocytophilum* strain HZ were used for the experiments described in this chapter. The cells were cultured in complete RPMI 1640 (Gibco/ThermoFisher Scientific) where RPMI media had been supplemented with 10% fetal bovine serum (Invitrogen/ThermoFisher Scientific, Waltham, MA) and 2 mM L-glutamine (Mediatech, Manassas, VA) . Media preparation and cells were handled under sterile conditions all the time. We followed previously described protocols from following groups (Chen, Dumler et al. 1994, Zhang, Chen et al. 2021). The cells (infected and uninfected) were maintained in T25 tissue culture flasks or 12 well tissue culture plates (TPP Techno Plastic Products AG). For certain experiments culture flasks were treated either with or

without 0.5 µM N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ). At specific time intervals, cultures were used for *A. phagocytophilum* isolation and analysis of protein expression levels and protein aggregation where necessary. Infection status of each culture was checked by hematoxylin staining as described by (Cheng and Ganta 2008). HL60 live cell counting was done using, Invitrogen Countess II Automated Cell Counter (ThermoFisher Scientific, USA).

To isolate host cell free *Anaplasma phagocytophilum* , when HL-60 cells were infected about 80-90%, cells were collected and harvested by centrifugation at 400 × g for 10 min at 4°C. The pellet was resuspended in 5 ml of RPMI media , and the cells were disrupted by sonication (Fisher Scientific, 60 Sonic Dismembrator). Sonicated two times for 30 seconds, each time, at a setting of 6.5 to release *Anaplasma* organisms from host HL-60 cells. The lysate was Centrifuged for 5 min at 100 × g, 4°C, to pellet host nucleus and other host cell debris. The supernatant was filtered by passing it through 2.7µm sterile isopore membrane filters (Whatman, Pittsburgh, PA). This filtrate then centrifuged for 15 minutes at 15,560 × g, 4°C to isolate host cell free *Anaplasma phagocytophilum*.

5.2.2 Cultivation and isolation of *Ehrlichia chaffeensis* in DH82 cells.

Canine macrophage like cell line DH82 (ATCC® CRL-10389™, Manassas, VA) and *Ehrlichia chaffeensis* Arkansas isolate were used for the experiments described in this chapter. The cells were cultured in complete MEM media (Gibco/ThermoFisher Scientific) where MEM media had been supplemented with 10% fetal bovine serum (Invitrogen/ThermoFisher Scientific, Waltham, MA) and 2 mM L-glutamine (Mediatech, Manassas, VA) . Media preparation and cells were handled under sterile conditions all the time. We followed previously described protocols from following groups (Chen, Dumler et al. 1994, Zhang, Chen et al. 2021). The cells (infected and uninfected) were maintained in T25 tissue culture flasks or 12 well tissue culture plates (TPP

Techno Plastic Products AG). For certain experiments culture flasks were treated either with or without 0.1 µM N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ). At specific time intervals, cultures were used for *E. chaffeensis* isolation and analysis of protein expression levels and protein aggregation where necessary. Infection status of each culture was checked by hematoxylin staining as described by (Cheng and Ganta 2008). DH82 live cell counting was done using, Invitrogen Countess II Automated Cell Counter (ThermoFisher Scientific, USA).

To isolate host cell free *E. chaffeensis*, when DH82 cells were infected about 80-90%, cells were collected and harvested by centrifugation at 400 × g for 10 min at 4°C. The pellet was resuspended in 5 ml of MEM media , and the cells were disrupted by sonication (Fisher Scientific, 60 Sonic Dismembrator). Sonicated two times for 30 seconds, each time, at a setting of 6.5 to release *Ehrlichia* organisms from host DH82 cells. The lysate was Centrifuged for 5 min at 100 × g, 4°C, to pellet host nucleus and other host cell debris. The supernatant was filtered by passing it through 2.7µm sterile isopore membrane filters (Whatman, Pittsburgh, PA). This filtrate then centrifuged for 15 minutes at 15,560 × g, 4°C to isolate host cell free *Ehrlichia chaffeensis*.

5.2.3 Isolation of *Ehrlichia* protein aggregates.

Once host cell free *Ehrlichia* were isolated, cells were lysed by sonication as mentioned in (Kwiatkowska, Matuszewska et al. 2008). We used equal number of *Ehrlichia* cells from all samples for protein isolation experiments. Equal number of Ehrlichia were taken as described by (Sirigireddy and Ganta 2005) .Unbroken bacterial cells were removed by centrifugation (2000 × g, at 4°C for 20 minutes). The supernatant from earlier step was incubated with 2% Triton X-100 for 30 minutes at room temperature. Next , Triton treated sample was centrifuged at 21000 × g, at 4°C for 30 minutes to separate soluble protein and insoluble aggregates from *Ehrlichia*. The

Protein concentration of the aggregated and soluble fractions were measured by Bradford method (Bio-Rad ,USA).

5.2.4 Bacterial strains.

Anaplasma phagocytophilum (strain HZ), *Ehrlichia chaffeensis* Arkansas isolate, *Ehrlichia chaffeensis* ClpB deleted Arkansas isolate (named as TR91 strain) and following strains of *E. coli* were used: MC4100, MC4100ΔclpB::kan,(Squires, Pedersen et al. 1991).

Ehrlichia chaffeensis ClpB deleted strain was generated at Dr Roman Ganta's lab. This strain was generated by inserting a mCherry and gentamicin cassette within ClpB coding sequence by homologous recombination. In TR91 strain, ClpB gene is disrupted and hence no ClpB protein is being produced (Appendix C supplementary figure C.3 and C.4).

5.2.5 Bacterial viability studies.

E. coli MC4100, and MC4100ΔclpB, strains were used to determine bacterial growth and viability. Bacteria were maintained in the LB media, and in the case of the MC4100ΔclpB strain, LB was supplemented with 30 μg/ml kanamycin. All experiments were initiated by preparing overnight cultures inoculated from single colonies and grown at 37 °C. On the following day, the cultures were diluted 100-fold in 10 ml of LB without antibiotics and incubated at 37 °C. The culture optical density was monitored at 600 nm. When absorbance at 600 nm reached 0.4, samples were immediately transferred into an incubator/shaker at temperature 30, or 45°C. Bacteria were cultured for up to 4 hours with shaking (200 RPM). At specific time points, 100 μl of each culture was withdrawn and serially diluted in sterile 0.9% NaCl up to 10^6 -fold dilution. After a complete adsorption of liquid on the LB agar surface, the plates were incubated overnight at 37 °C.

5.2.6 Examination of host cells and infected cells by polychromatic staining.

Aliquots of 200 µl of uninfected or infected host cells were withdrawn at specific time points. Withdrawn aliquots were transferred onto glass slides using a Cytospin centrifuge and stained using Hema-3 staining kit (Fisher Diagnostics) according to supplier's instructions. The slides were air dried and examined using a light microscope with an oil immersion 100× objective. Image was captured and generated using ACCU-SCOPE Excelis camera and CaptaVision+™ Software.

5.2.7 Western blotting.

The ClpB , DnaK and ClpP protein levels in *E. chaffeensis* cells were analyzed by Western blotting. Host cell free *E. chaffeensis* cells were collected at different time post infection time points as described in materials and methods. The cells were suspended in 160 µl of 2x Laemmli sample buffer. Sixty microliters of each sample was resolved in duplicates in 10% SDS-PAGE gel. One gel was used for blotting onto a nitrocellulose membrane and the second part of gel was stained with silver staining. The membrane was blocked with 3% Bovine serum albumin (BSA) dissolved in TBST buffer (19 mM Tris-HCl, pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.2% Tween-20) overnight at 4 °C. Next, the membrane was incubated with rabbit polyclonal anti-ClpB IgG or anti-DnaK IgG or ClpP Antibody (PA5-52722, ThermoFisher Scientific USA) (1:50,000 in BSA) for 1 hour at room temperature. After incubation with the primary antibody, the membrane was washed in TBST buffer (4 times, 10 min). Next, the membrane was incubated with anti-rabbit horseradish peroxidase conjugated secondary antibodies for 1 hour at room temperature and washed in TBST (4 times, 10 min) and then TBST without Tween-20 (10 min). Signal detection was performed using SuperSignal West Pico Chemiluminescent Substrate (Pierce) and image

acquiring was done using an Invitrogen iBright FL1500 Imaging System (ThermoFisher Scientific, USA).

5.2.8 Molecular detection of *Anaplasma* and *Ehrlichia* by RT-PCR assay.

Real-time quantitative PCR (RT-PCR) was performed using TaqMan probes ,SuperScript™ III Platinum™ One-Step qRT-PCR Kit (Invitrogen, USA) and Applied Biosciences StepOnePlus thermocyclers. Followed the exact protocol as described by (Sirigireddy and Ganta 2005) using species specific primers and common TaqMan probes. Bacterial numbers were estimated by real-time quantitative PCR targeting the16S rRNA gene segment and preparing a standard curve as described by (Sirigireddy and Ganta 2005).

5.3 Results.

5.3.1 Wild type *E. chaffeensis* and the ClpB-deleted strain produce different protein aggregation patterns.

DH82 cells were infected with wild type *E. chaffeensis* and the *E. chaffeensis* ClpB-deleted strain (TR91). Cells were grown in 12-well plates and harvested on the indicated post-infection days as in figure 5.1. Host cell free *Ehrlichia* was isolated as described in materials and methods. Soluble and aggregated fractions were isolated as described earlier. Total protein concentration was measured by Bradford method and equal amounts of proteins were loaded and stained with the silver staining method. Both soluble and aggregated protein patterns were distinctive between wild type and ClpB deleted *Ehrlichia* strains.

5.3.2 DnaK and ClpP proteins are expressed in high amounts in the ClpB-deleted strain of *E.chaffeensis*.

As we observed distinctive protein aggregation patterns in wild type *E. chaffeensis* and the ClpB-deleted strain (TR91), we investigated whether the ClpB deletion had any impact on the expression of other chaperones. We decided to study DnaK and ClpP (ATP-dependent Clp protease proteolytic subunit) protein expression levels at different post infection time points. First, deletion of the ClpB gene in *E. chaffeensis* was confirmed by Western blotting using anti-ClpB antibodies (figure 5.2). Second, our data revealed that DnaK and ClpP, were expressed at a higher level in the ClpB-deleted strain than in wild type *Ehrlichia* (figure 5.2) and they mostly accumulated with the aggregated proteins.

5.3.3 Protein aggregation does not apparently increase upon deletion of ClpB in *E. chaffeensis*.

DH82 cells were infected with wild type *E. chaffeensis* and the ClpB-deleted strains (TR91). Cells were grown in 12-well plates and harvested on the indicated post infection days as in figure 5.3. Soluble and aggregated fractions were isolated as described earlier. Total protein concentration was measured by Bradford method. Our results revealed that the amount of aggregated proteins in *E. chaffeensis* ClpB-deleted strain (TR91) did not show a significant change when compared with the wild type strain.

5.3.4 Replication of *Ehrlichia chaffeensis* is affected by ClpB deletion.

We investigated the impact of ClpB deletion in *E. chaffeensis* by studying its replication in DH82 cells. Equal number of the infected DH82 cells was added to each well. Cells were grown in 12-well plates and harvested on the indicated post infection days as in figure 5.4. The number of *E. chaffeensis* cells were determined by quantitative real time PCR and normalized based on the initial infectious inoculum. We found that the ClpB-deleted strain (TR91) had a significantly slower replication rate compared to wild type *E. chaffeensis*.

5.3.5 Proliferation of DH82 and HL60 cells is not significantly affected by DBeQ.

Uninfected DH82 and HL60 cells were grown with 0.5 µM DBeQ or with DMSO (figure 5.9). We decided to use DBeQ at 0.5 µM concentration since the previously reported IC_{50} for p97 ATPase was 1.5 µM (Chou, Brown et al. 2011). We monitored cell proliferation over time and both DH82 and HL60 flasks became saturated with cells within 5 days. Our results indicated that growth and replication of uninfected DH82 and HL60 cells were not significantly affected by 0.5 µM DBeQ in the media.

5.3.6 DBeQ inhibits replication of *Anaplasma phagocytophilum* in HL60 cells.

Uninfected HL60 cells were maintained in T25 flasks. 0.5 µM DBeQ was added to one flask and DMSO was added to the other flask. Flasks were infected with the same number of *A. phagocytophilum*. 200 µl of sample was withdrawn every day and used for detection of 16S DNA (16S rRNA gene) using species specific primers and TaqMan probes (refer to materials and methods). Our quantitative Polymerase Chain Reaction (qPCR) data revealed that the infection was cleared in the flasks treated with 0.5 µM DBeQ within first 48 hours and the amount of *A. phagocytophilum* 16S DNA reached almost undetectable level between 48-96 hours. In the control flasks treated with DMSO, the infection was propagated and cells were lysed completely after 96 hours (figure 5.10 and appendix C, figure C.1).

5.3.7 DBeQ inhibits replication of *Ehrlichia chaffeensis* in DH82 cells.

Uninfected DH82 cells were maintained in T25 flasks. 0.1 µM DBeQ was added to one flask and DMSO was added to the other flask. Flasks were infected with the same number of wild type *E. chaffeensis* or the ClpB-deleted strain (TR91). 200 µl of sample was withdrawn every day and used for detection of 16S DNA (16S rRNA gene) using species specific primers and TaqMan probes (refer to materials and methods). Our quantitative Polymerase Chain Reaction (qPCR) data revealed that the infection was cleared in the flasks treated with 0.1 µM DBeQ within first 48 hours and the amount of *E. chaffeensis* 16S DNA reached almost undetectable level between 48-96 hours. In the flasks with wild type *E. chaffeensis* treated with DMSO, the infection was propagated and cells were lysed completely after 96 hours (figure 5.11 and appendix C, figure C.2). However, the infection with the *E. chaffeensis* ClpB-deleted strain (TR91) was also cleared in the presence of 0.1 µM DBeQ by the sixth day (96 hours) post infection (figure 5.11).

5.4 Discussion.

Our earlier work demonstrated that the potential target of DBeQ in *E. coli* could be ClpB (Glaza, Ranaweera et al. 2021). In fact, our Surface Plasmon Resonance (SPR) data revealed that DBeQ binds to *E. coli* ClpB with an apparent K_d ~ 65 µM. Since ClpB is involved in maintaining protein homeostasis, we speculate that the DBeQ effect on *E. coli* was linked to dysregulation of protein homeostasis in *E. coli* cells. We further hypothesized that ClpB could be the main target for DBeQ also in other bacteria, including the obligate intracellular pathogens. Therefore, we wanted to study protein aggregation in obligate intracellular pathogens first and gain some insight into this process prior to conducting studies with DBeQ. Our studies on protein aggregation in *A. phagocytophilum* are explained in detail in chapter 4 and we are presenting our results on *E. chaffeensis* in this chapter.

We studied protein aggregation in wild type *E. chaffeensis* and in the ClpB-deficient *E. chaffeensis* strain (TR91, appendix C, figure C.4). As shown in figure 5.1, wild type *E. chaffeensis* and the ClpB-deleted strain produce different protein aggregation patterns indicating that both soluble and aggregated protein patterns were distinctive between wild type and ClpB deleted *Ehrlichia* strains. This result indicated that protein aggregation affects a subset of *E.chaffeensis* proteins and. As expected, ClpB was only detected in wild type *E. chaffeensis* (figure 5.2) and the amount of ClpB found in the soluble fraction was undetectable. However, our results revealed that the amount of aggregated proteins in *E. chaffeensis* ClpB-deleted strain (TR91) did not show a significant change when compared with the wild type strain,(figure 5.2). Our results showed that protein aggregation did not increase upon deletion of ClpB in *E. chaffeensis*.

Amount of aggregated protein amount increased gradually up to 48 hours (figure 5.3) and declined thereafter in both strains. Our current observation corroborate the previously reported

data on wild type *E. chaffeensis* (Rikihisa 2010, Kuczynska-Wisnik, Cheng et al. 2017). The observed increase in the amounts of aggregated proteins (first 48 hours) occurred during intense pathogen replication (RC stage) within host cells as reported earlier (Zhang, Kedzierska-Mieszkowska et al. 2013) and declined slowly after 48 hours as pathogen transformed itself into non-replicating, infectious form (DC stage). The highest amount of ClpB was found after 48 hours in the aggregated fraction (figure 5.2) and declined gradually. It is not surprising to find ClpB with aggregated proteins as its biological role is to reactivate and disaggregate the aggregated proteins within a cellular milieu. Accumulation of ClpB within the first 48 hours also agrees with the previously reported accumulation of ClpB mRNA within 48 hour post infection window of *E. chaffeensis* (Zhang, Kedzierska-Mieszkowska et al. 2013).

Both wild type and TR91 *E. chaffeensis* had detectable levels of DnaK in the soluble fraction from 48 hours onwards. Based on the signal intensity, the highest amount of DnaK in the soluble fraction was observed after 48 hours in both strains (figure 5.2) and decreased gradually thereafter. It is well-established that DnaK can prevent misfolding of nascent peptides, solubilize small aggregates and guide ClpB towards larger aggregates (Diamant, Ben-Zvi et al. 2000, Fay and Glickman 2014, Mattoo and Goloubinoff 2014). Hence it is not surprising to find the highest amount of DnaK (in the soluble fractions) within the 48 hour post-infection window. Interestingly, the amount of DnaK found in the aggregated fraction increased rapidly from 0^{th} day to 6th post infection days. In Wt *E. chaffeensis* DnaK showed ≈ 3-fold increase in the amount in the aggregated fraction, whereas the same ratio was ≈ 7-fold higher in TR91 (figure 5.2). This large increase in DnaK accumulation observed in TR91 could be due to lack of ClpB in the TR91 strain. Hence, one can speculate that observation of increase in DnaK could be due to the loss of ClpB function in TR91 strain.

If a cell has lost a protein which is responsible for resolubilization and recycling of misfolded and aggregated proteins , to maintain protein homeostasis, the cell must start degrading the accumulating misfolded and aggregated proteins. Hence, we speculated that perhaps cellular proteases might have been also preferentially expressed in the TR91 strain. Therefore, we investigated expression levels of an ATP-dependent Clp protease proteolytic subunit (ClpP) in both *Ehrlichia* strains. ClpP is a serine protease that is highly conserved throughout bacteria and eukaryotes (Katayama-Fujimura, Gottesman et al. 1987, Corydon, Bross et al. 1998, Yu and Houry 2007, Li, Chung et al. 2010, Moreno-Cinos, Goossens et al. 2019). ClpP proteases typically interact with ATP dependent AAA+ chaperones (ClpA, ClpX etc, but not ClpB) that bind and unfold target substrates and then translocate them into ClpP for degradation. ClpP was present in the soluble fractions in both *E. chaffensis* strains, but the amount of ClpP trapped with aggregated fractions increased rapidly up to 96 hours (post infection day 4) and remained relatively same thereafter until post infection day 6 (figure 5.2). As the cells are undergoing intense replication within the first 48 hours of infection, more proteins are likely to be misfolded and become aggregated. ClpB helps maintain protein homeostasis in the cells by reactivating and resolubilizing the aggregated proteins, but our results suggest an additional role of ClpB in maintaining protein homeostasis in *E. chaffensis* during infection of mammalian cells.

Next, we investigated the effect of DBeQ on HL60 and DH82 cell proliferation (figure 5.9). Our results demonstrated that HL60 and DH82 cell proliferation was not significantly affected by 0.5 µM DBeQ. We decided to test cell proliferation with 0.5 µM DBeQ as the previously reported IC_{50} value for p97 ATPase with DBeQ was 1.5 µM (Chou, Brown et al. 2011). Both HL60 and DH82 contain p97 and we used a three-times lower concentration of DBeQ in our experiment

than its IC_{50} for p97. Hence, we assume that at 0.5 µM DBeQ, a majority of HL60 and DH82 cell proliferation and metabolic functions are not significantly affected by DBeQ.

Finally, we tested the effect of DBeQ on two obligate intracellular pathogens, *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. Since ClpB appeared to be the main target of DBeQ in *E. coli* (Glaza, Ranaweera et al. 2021), we proposed that DBeQ could perform a similar function in *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* , i.e. targeting ClpB . Most strikingly, we observed a rapid clearance of infection with either pathogen by very low concentrations of DBeQ. Specifically, 0.5 µM DBeQ managed to clear the *Anaplasma phagocytophilum* infection by 48 hours after initial infection (figure 5.10). When DBeQ was used at 0.1 µM concentration with wildtype *E. chaffeensis*, the infection also cleared within the first 48 hours (figure 5.11). However, when DBeQ was used at 0.1 µM concentration with *E. chaffeensis* TR91 strain, it took 144 hours (6 days, figure 5.11) to clear the infection. These results demonstrated that ClpB is not the main target of DBeQ in *Ehrlichia* despite its extraordinary antibacterial activity against *Ehrlichia* infection in DH82 cells. However, it is important note that a slower clearance of TR91 infection by DBeQ could be a result of loss of ClpB in TR91 strain and might indicate that ClpB is among the DBeQ targets in *E. chaffeensis*, but is not the primary one.

In summary, we demonstrated that DBeQ manifests a potent antibacterial activity against *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* at sub-micromolar concentrations. This result shows that DBeQ can be used as a lead compound to develop novel antibacterial compounds against Human Granulocytic Anaplasmosis (HGA) and Human Granulocytic Ehrlichiosis (HGE). Antibacterial activity of DBeQ has been reported before (DeGraw, Brown et al. 1974, Van Horn, Burda et al. 2014, Glaza, Ranaweera et al. 2021) and we demonstrated recently (Glaza, Ranaweera et al. 2021) that the main target of DBeQ in *E. coli* is ClpB. However, our current experimental

results with *Ehrlichia chaffeensis* indicated that ClpB is not the main target in that pathogen. Perhaps DBeQ has other cellular targets, including the ClpAP, ClpXP protease systems (Singh, Guo et al. 1999, Baker and Sauer 2012) in the obligate intracellular organisms. Hence, it is important to conduct *in vitro* DBeQ experiments with ClpAP, ClpXP and ClpP proteins to get an insight on potential DBeQ effects on the biological activity of these protease systems. At the same time, studies with DBeQ on global gene expression patterns using RNA deep-sequencing technology may help us to narrow down the set of possible cellular targets of DBeQ in the obligate intracellular pathogens.

5.8 Acknowledgements.

I want to thank Dr Sunitha Shiva, , Dr Arathy Nair, Dr Paidashe Hove, Dr Chandramouli Kondethimmanahalli, Dr Huitao Liu, Jonathan Ferm, Swetha Madesh and Guanpeng Wang for their valuable suggestions and support throughout this project.

I want to thank Donald Harbidge from the Kansas State University Molecular Biology Core facility for his valuable suggestions and support in conducting RT-PCR experiments using Applied Biosciences StepOnePlus instruments.

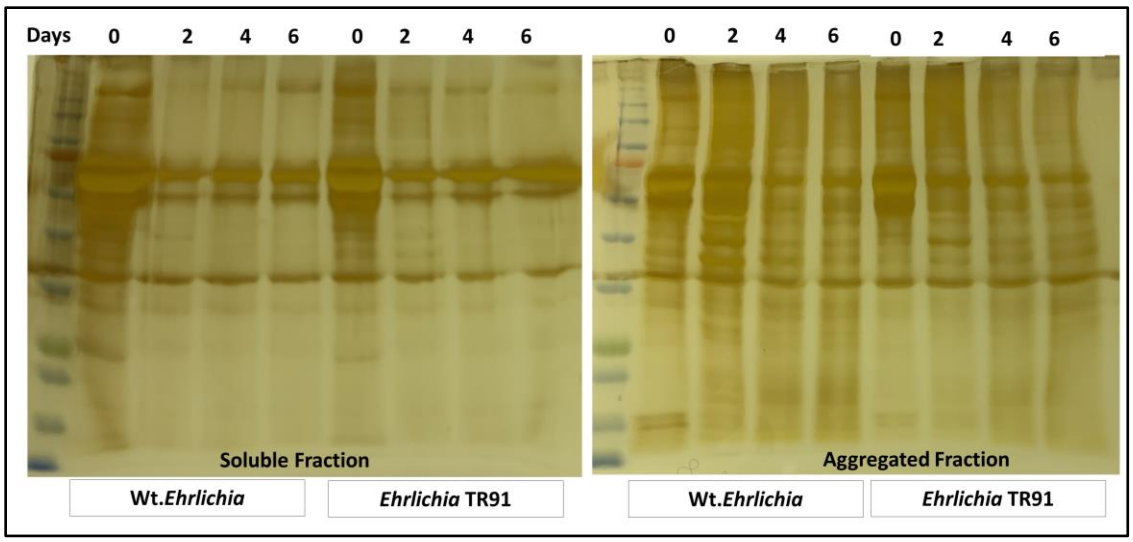
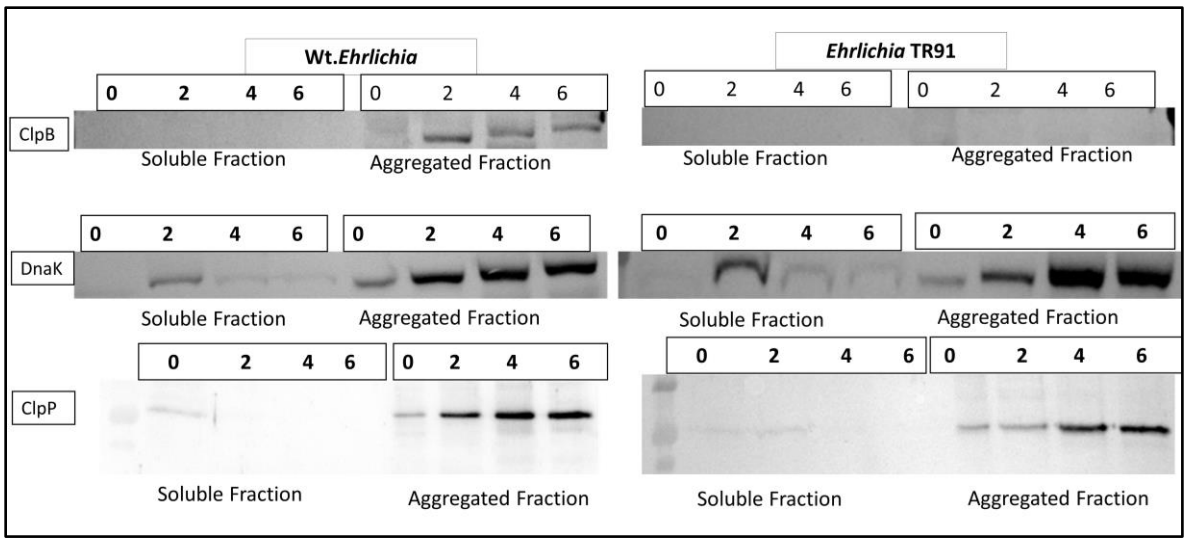
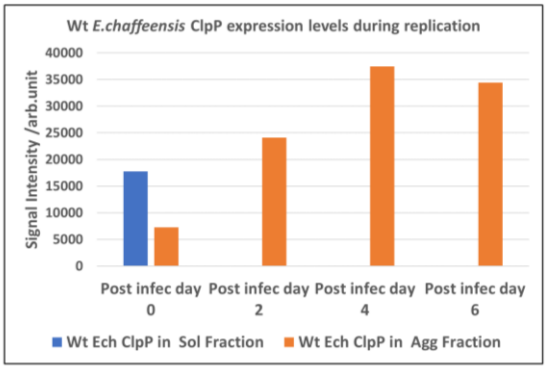
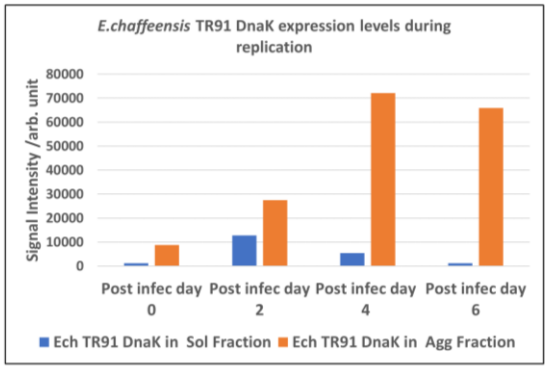


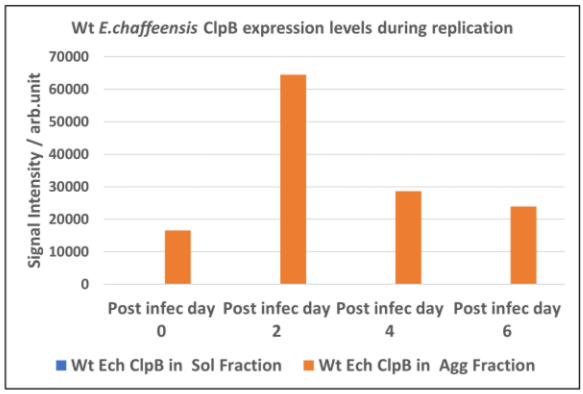
Figure 5.1. Protein aggregate patterns in wild type *E. chaffeensis* and *E.chaffeensis* ΔClpB.

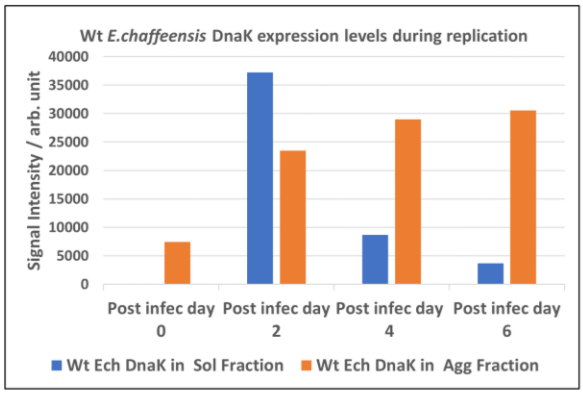
DH82 cells were infected by wild type *E. chaffeensis* and *E. chaffeensis* ClpB deleted (TR91) strains. Cells were grown in 12 well plates and harvested on the indicated post infection days as shown. Total protein concentration was measured by Bradford method and equal amounts of proteins were loaded and stained with silver staining method. BLUeye Prestained Protein Ladder (Sigma, USA) was used as the molecular weight indicator .Experiment was done in duplicates.


Wt.Ehrlichia
Ehrlichia TR91
ClpB
DnaK
ClpP
0 2 4 6
Soluble Fraction
Aggregated Fraction


Wt E.chaffeensis ClpP expression levels during replication
Signal Intensity /arb.unit
Post infec day 0
Post infec day 2
Post infec day 4
Post infec day 6
Wt Ech ClpP in Sol Fraction
Wt Ech ClpP in Agg Fraction


E.chaffeensis TR91 DnaK expression levels during replication
Signal Intensity /arb. unit
Post infec day 0
Post infec day 2
Post infec day 4
Post infec day 6
Ech TR91 DnaK in Sol Fraction
Ech TR91 DnaK in Agg Fraction


Wt E.chaffeensis ClpB expression levels during replication
Signal Intensity / arb.unit
Post infec day 0
Post infec day 2
Post infec day 4
Post infec day 6
Wt Ech ClpB in Sol Fraction
Wt Ech ClpB in Agg Fraction


Wt E.chaffeensis DnaK expression levels during replication
Signal Intensity / arb. unit
Post infec day 0
Post infec day 2
Post infec day 4
Post infec day 6
Wt Ech DnaK in Sol Fraction
Wt Ech DnaK in Agg Fraction

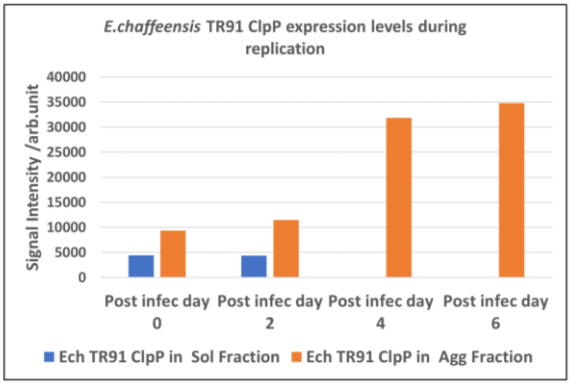
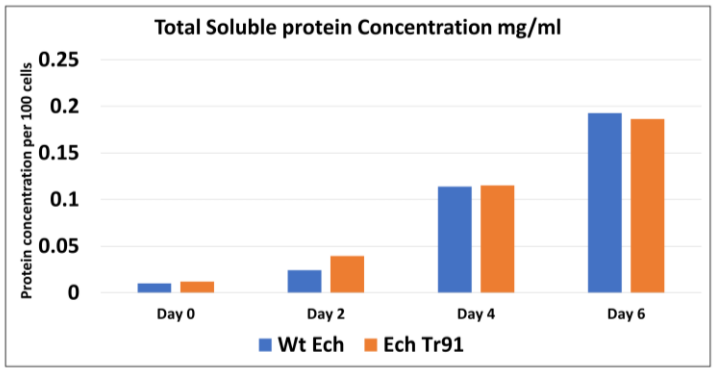


Figure 5.2. ClpB, DnaK and ClpP protein expression in wild type *E. chaffeensis* and *E .chaffeensis* ClpB deleted strains.

DH82 cells were infected by wild type *E. chaffeensis* and *E. chaffeensis* ClpB deleted (TR91) strains. Cells were grown in 12 well plates and harvested on the indicated post infection days as shown. Immunodetection of ClpB, DnaK and ClpP was done as explained in materials and methods. Equal amounts of proteins were loaded. Experiment was done in duplicates and shown here representation of one experiment. Image quantification (signal intensity) performed using ImageJ program from National Institutes of Health (NIH). If data is absent on any day, indicates that the signal was too low for western blot detection and quantification.



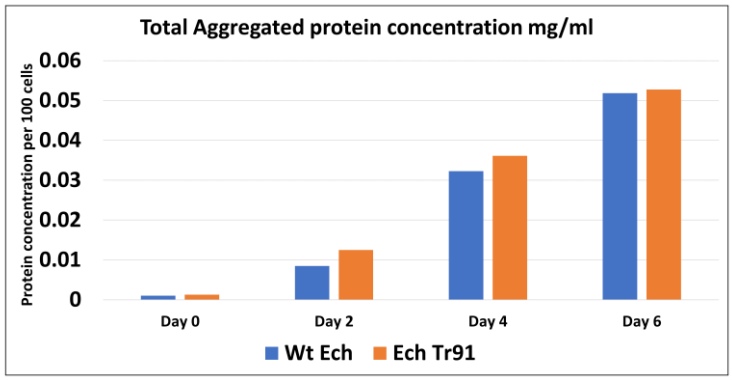


Figure 5.3. Accumulation of aggregated proteins in *E. chaffeensis* ClpB deleted strain.

DH82 cells were infected by wild type *E. chaffeensis* and *E. chaffeensis* ClpB deleted (TR91) strains. Cells were grown in 12 well plates and harvested on the indicated post infection days as shown. Number of *E. chaffeensis* cells were calculated as explained in materials and methods. Total protein concentration was measured by Bradford method. Experiment was done in duplicates and average values are shown here.

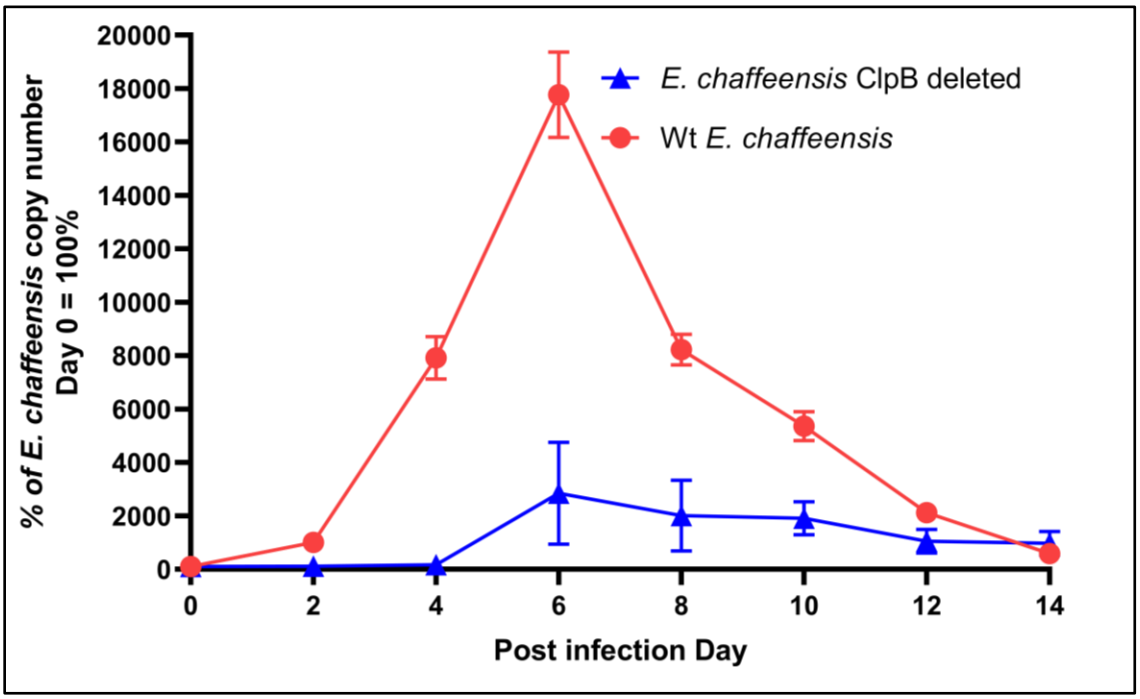


Figure 5.4. *E. chaffeensis* ClpB deleted strain (TR91) replication pattern in DH82 cells.

Cells were grown in 12 well plates and harvested on the indicated post infection days as shown here. Number of *E. chaffeensis* cells were calculated as explained in materials and methods and number of *E. chaffeensis* were adjusted based on initial infectious inoculum . Experiments was done in triplicates. Data represent the average values and error bars represent standard deviation.

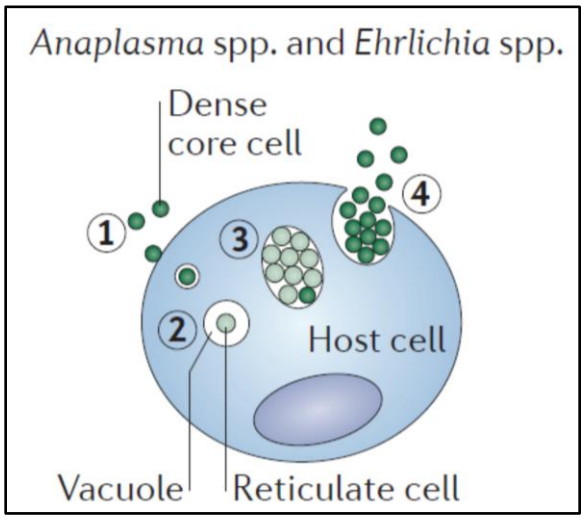


Figure 5.5. Life cycle of *A. phagocytophilum* and *E. chaffeensis*.

The infectious form of *A. phagocytophilum* and *E. chaffeensis* is known as Dense Core (DC) cells and replicative form is known as Reticulate Cells (RC). When these bacteria encounters a suitable eukaryotic host cell, Dense Core cells get internalized (step 1). Once internalized, these DC cells differentiates into Reticulate Cells (step 2) in a vacuole. These RC cells would multiply by binary fission inside these vacuoles (step 3). Once these RC cells are multiplied, these RC cells will differentiate back to DC cells and exit the host cells by lysis (step 4). This figure is adapted from (McClure, Chávez et al. 2017). Reprinted with permission.

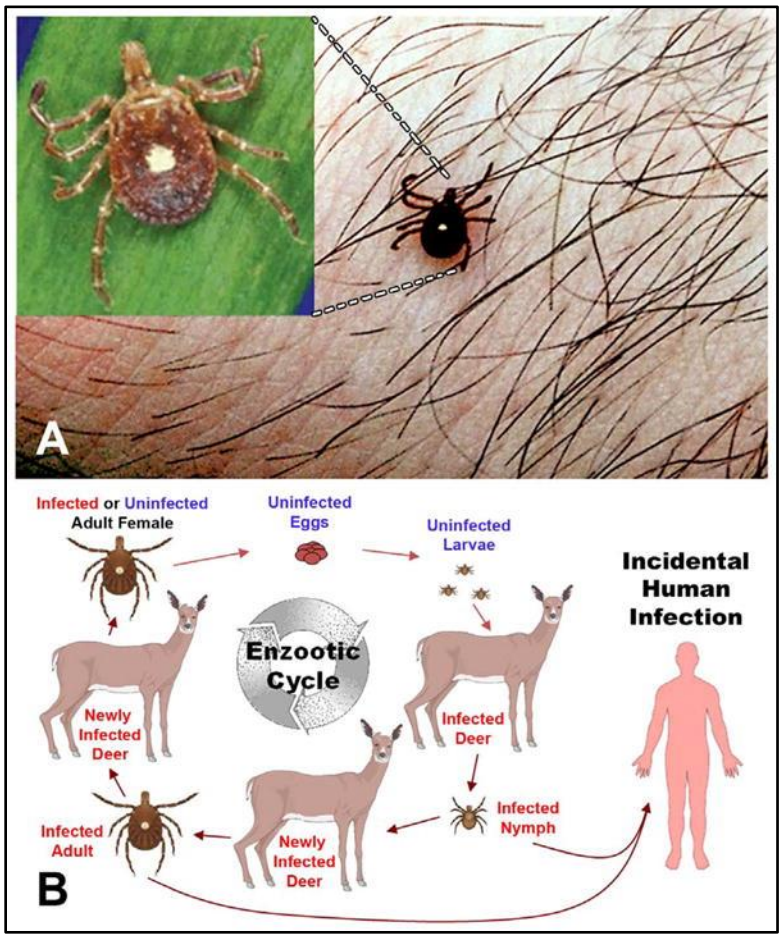


Figure 5.6. Life cycle of *Amblyomma Americanum*.

Amblyomma ticks have three feeding stages (B, larval, nymph, and adult); each developmental stage feeds only once. The transmission of *Ehrlichia* happens during blood meal (either from infected tick to an uninfected host or from infected host to an uninfected tick). Picture of *Amblyomma Americanum* (A) is shown here . Original image is taken from (Ismail, Bloch et al. 2010). Reprinted with permission.

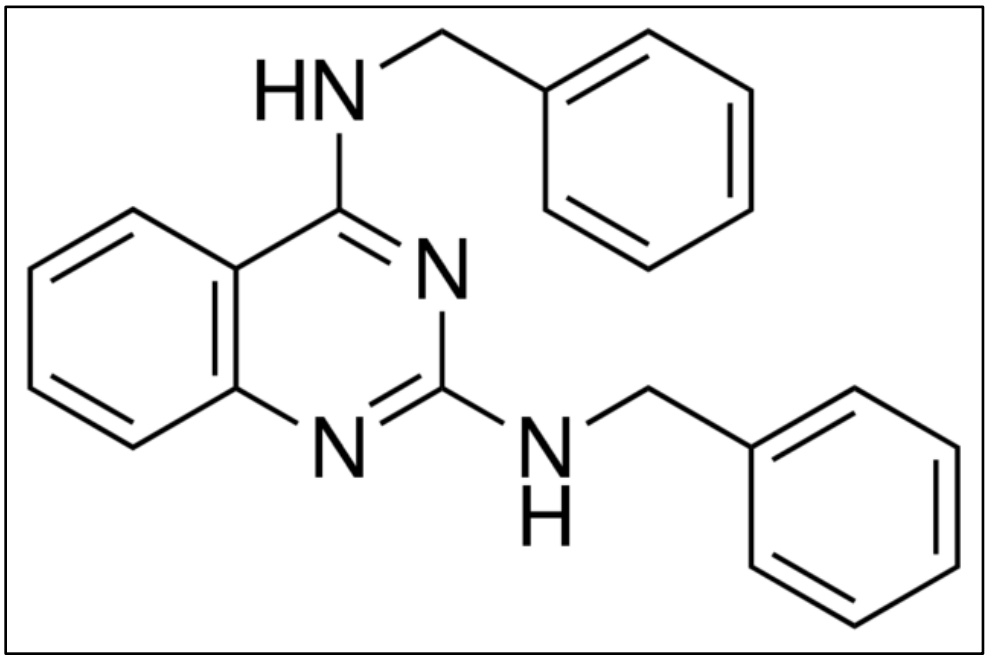


Figure 5.7. Structure of DBeQ.

N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ) structure is shown here. Based on the structures supplied by the manufacturer and redrawn using ChemDraw 12. DBeQ was initially described as inhibitor (an antitumor target) for the human AAA+ ATPase p97. DBeQ is a potent, reversible, inhibitor of p97 with IC_{50} of 1.5 μM (Chou, Brown et al. 2011). p97 (also known as Valosin-containing protein (VCP), Cdc48 in yeast and plants, CDC-48 in worms, Ter94 in flies) is a hexameric AAA+ protein. p97 is mainly involve in cellular protein degradation machinery and is not found in bacteria (Meyer and Weihl 2014, Glaza, Ranaweera et al. 2021).

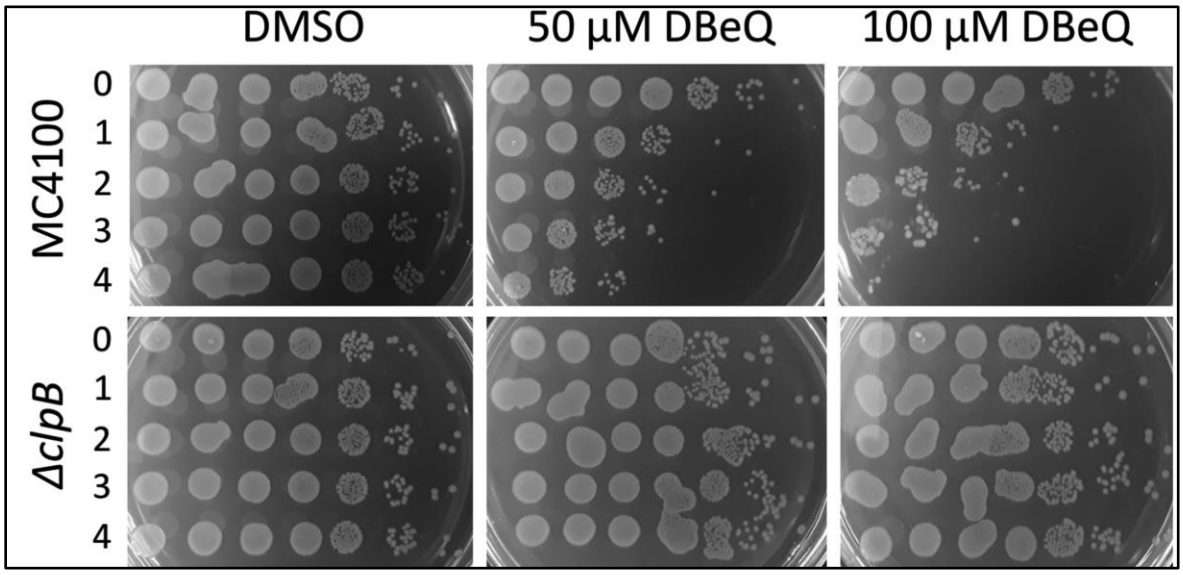


Figure 5.8. Viability of *E. coli* cells with DBEQ at 45°C.

Wild type *E. coli* (MC4100) and *E. coli* strain with ClpB gene deleted (MC4100 ΔClpB) were incubated at 45°C in the presence of DBeQ or with only DMSO for the period of time indicated on the left (in hours) and then spotted on agar plates and incubated overnight at 37°C. Each spot on the agar plates represents a viable culture after a 10-fold serial dilution (from left to right). Shown are representative results from 3 repeated experiments. DMSO, dimethyl sulfoxide. Original image is taken from (Glaza, Ranaweera et al. 2021). Reprinted with permission.

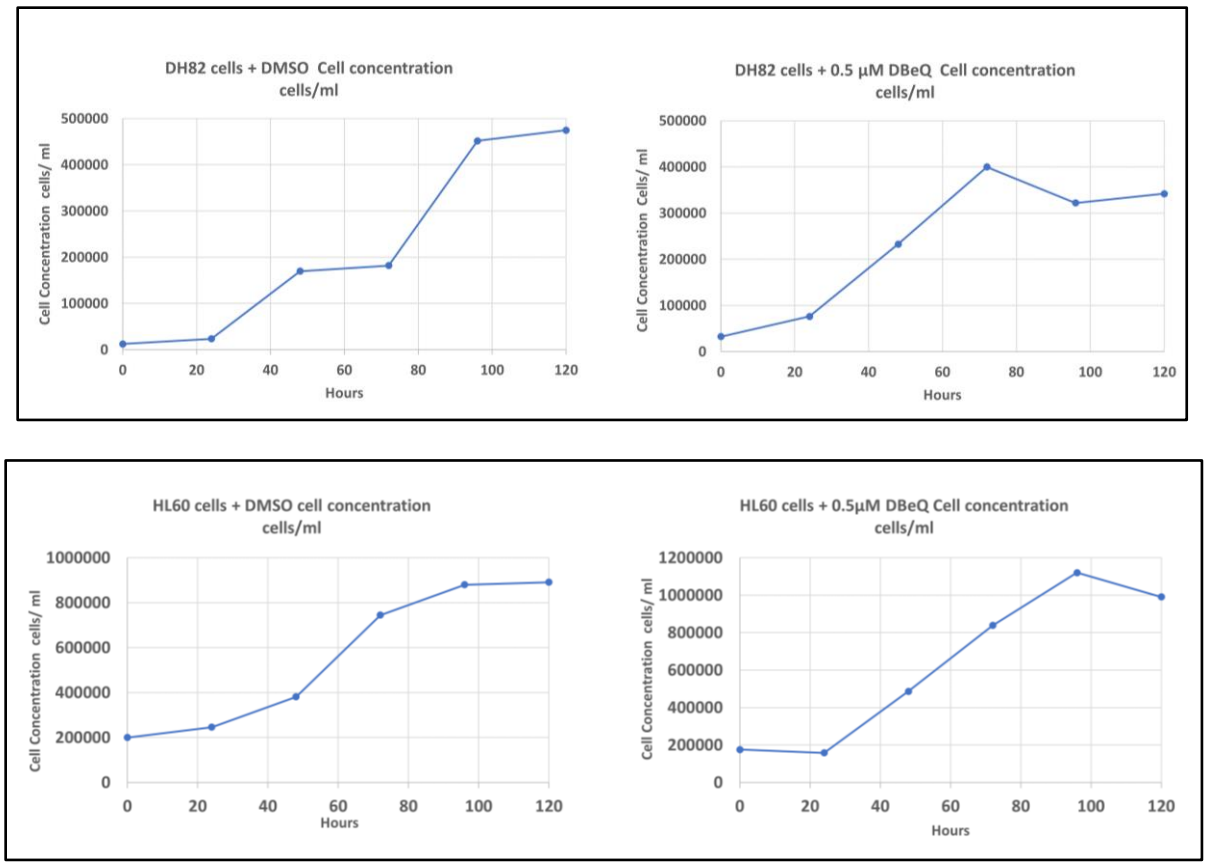


Figure 5.9. DH82 cells and HL60 cells proliferation is not significantly affected by DBeQ.

Uninfected DH82 and HL60 cells were grown with 0.5 µM DBeQ. DBeQ was dissolve in Dimethyl Sulfoxide (DMSO). Same number of host cells were used at the beginning of the experiment and cells were maintained for a week. Cell counting was done using automated Countess Cell Counter (ThermoFisher Scientific, USA). Average values of two independent experiments are shown.

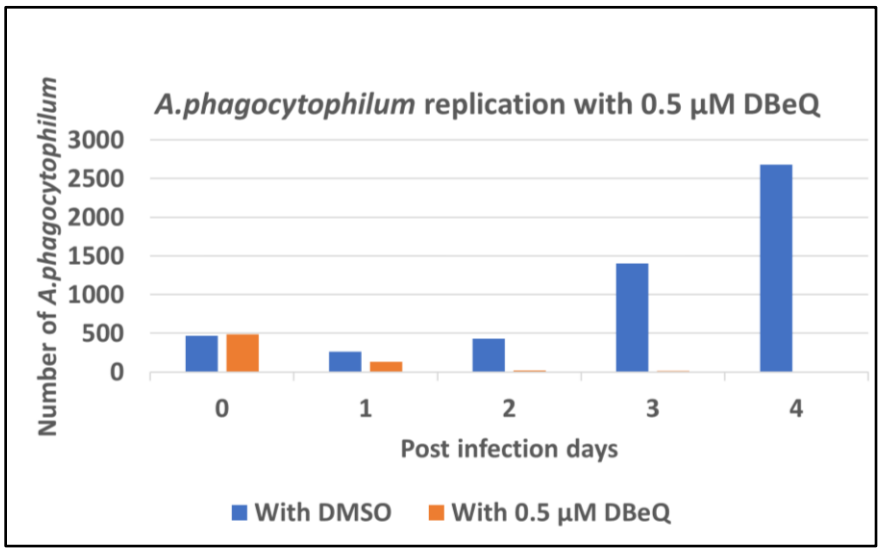


Figure 5.10. DBeQ inhibits Anaplasma Phagocytophilum in HL60 cells.

Uninfected HL60 cells were maintained in T25 flasks. 0.5 µM DBeQ was added to one flask and DMSO was added to the other flask. Experiment was done in duplicates. Flasks were infected with the same number of host cell free *A. phagocytophilum* at the same time. 200 µl of sample was withdrawn every day and used for detection of 16S rRNA gene DNA using species specific primers and TaqMan probes. Infection propagation was monitored every day by hematoxylin staining. Infection was cleared in the flasks treated with 0.5 µM DBeQ within first 48 hours. Flasks treated with DMSO , infection was propagated, and cells were lysed completely after 96 hours. Average values of two independent experiments are shown.

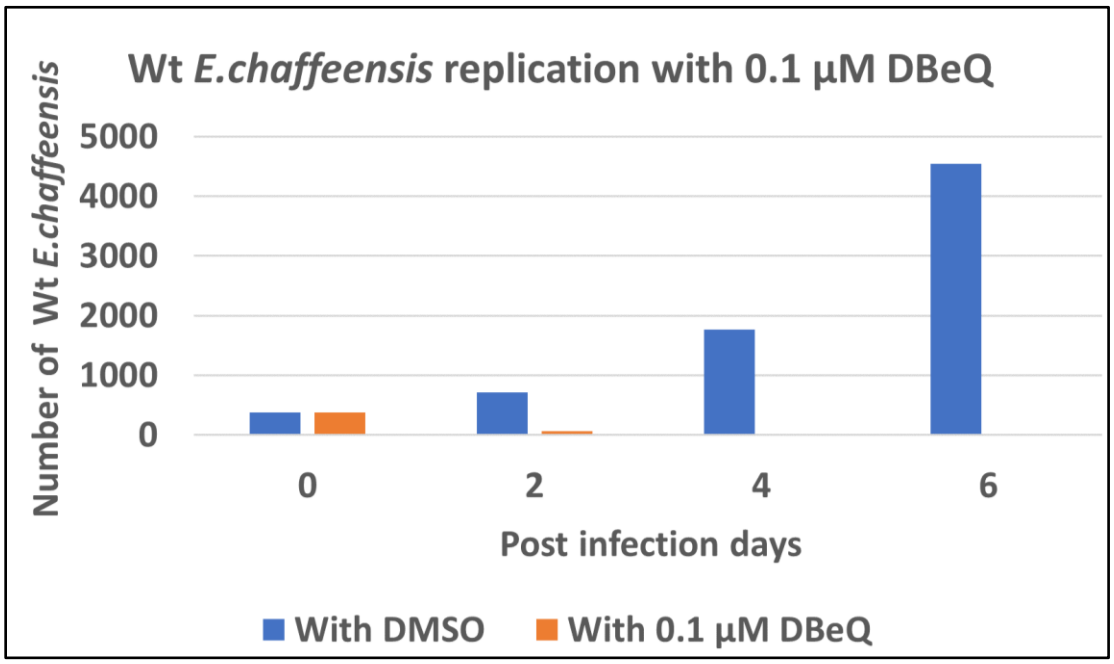


Figure 5.11. DBeQ inhibits replication of wild type *E. chaffeensis* in DH82 cells.

Uninfected DH82 cells were maintained in T25 flasks. 0.1 µM DBeQ was added to one flask and DMSO was added to the other flask. Experiment was done in duplicates. Flasks were infected with the same number of host cell free *E. chaffeensis* at the same time. 200 µl of sample was withdrawn every day and used for detection of 16S rRNA gene DNA using species specific primers and TaqMan probes. Infection propagation was monitored every day by hematoxylin staining. Infection was cleared in the flasks treated with 0.1 µM DBeQ within first 48 hours in wild type *E. chaffeensis*. Flasks treated with DMSO , infection was propagated, and cells were lysed completely after 96 hours.

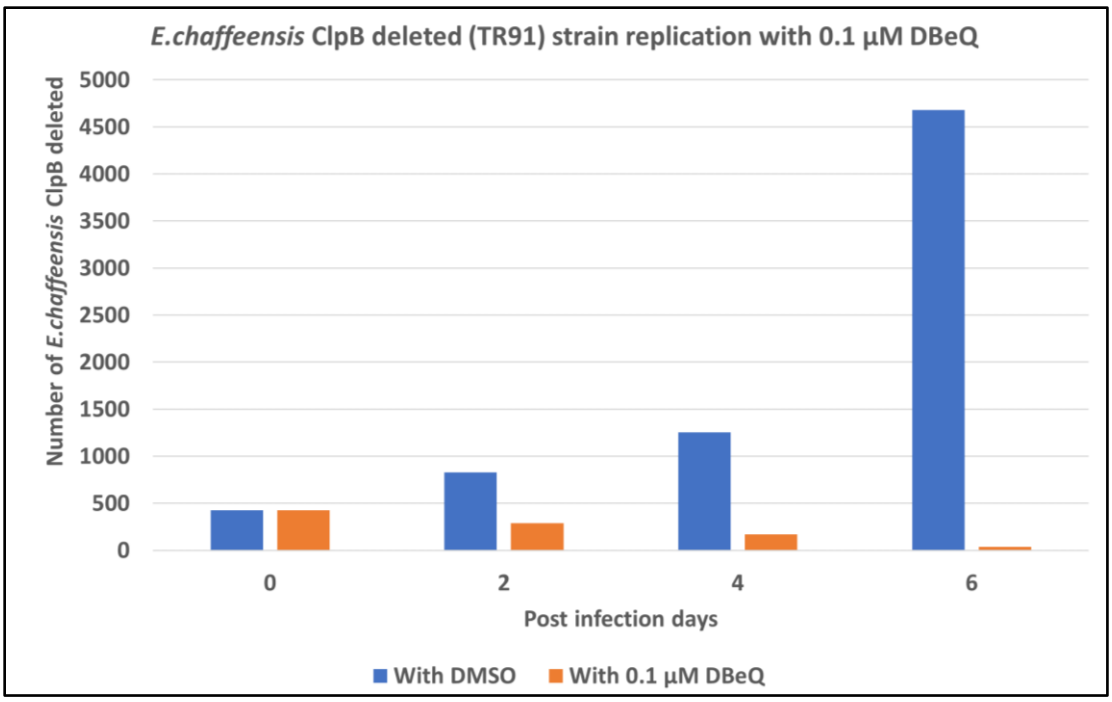


Figure 5.12. DBeQ inhibits replication of ClpB deleted *E. chaffeensis* (TR91) in DH82 cells.

Uninfected DH82 cells were maintained in T25 flasks. 0.1 µM DBeQ was added to one flask and DMSO was added to the other flask. Experiment was done in duplicates. Flasks were infected with the same number of host cell free *E. chaffeensis* at the same time. 200 µl of sample was withdrawn every day and used for detection of 16S rRNA gene DNA using species specific primers and TaqMan probes. Infection propagation was monitored every day by hematoxylin staining. Infection by *E. chaffeensis* ClpB deleted strain also got cleared by 0.1 µM DBeQ by sixth day while DMSO treated flask had infection propagated and cells were lysed completely after 96 hours. Average values of two independent experiments are shown.

5.9 References.

Baker, T. A. and R. T. Sauer (2012). "ClpXP, an ATP-powered unfolding and protein-degradation machine." Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1823**(1): 15-28.

Bakken, J. S. and J. S. Dumler (2006). "Clinical diagnosis and treatment of human granulocytotropic anaplasmosis." Ann N Y Acad Sci **1078**: 236-247.

Chen, S. M., J. S. Dumler, J. S. Bakken and D. H. Walker (1994). "Identification of a granulocytotropic *Ehrlichia* species as the etiologic agent of human disease." J Clin Microbiol **32**(3): 589-595.

Cheng, C. and R. R. Ganta (2008). "Laboratory Maintenance of *Ehrlichia chaffeensis* and *Ehrlichia canis* and Recovery of Organisms for Molecular Biology and Proteomics Studies." Current Protocols in Microbiology **9**(1): 3A.1.1-3A.1.21.

Chou, T. F., S. J. Brown, D. Minond, B. E. Nordin, K. Li, A. C. Jones, P. Chase, P. R. Porubsky, B. M. Stoltz, F. J. Schoenen, M. P. Patricelli, P. Hodder, H. Rosen and R. J. Deshaies (2011). "Reversible inhibitor of p97, DBeQ, impairs both ubiquitin-dependent and autophagic protein clearance pathways." Proc Natl Acad Sci U S A **108**(12): 4834-4839.

Corydon, T. J., P. Bross, H. U. Holst, S. Neve, K. Kristiansen, N. Gregersen and L. Bolund (1998). "A human homologue of *Escherichia coli* ClpP caseinolytic protease: recombinant expression, intracellular processing and subcellular localization." Biochem J **331 (Pt 1)**(Pt 1): 309-316.

DeGraw, J., V. Brown and W. Colwell (1974). "Potential Antileprotic Agents. 3. Inhibition of Mycobacterial Dihydrofolic Reductase by 2, 4-diamino-5-methyl-6-alkylquinazolines." Journal of medicinal chemistry **17**(7): 762-764.

Diamant, S., A. P. Ben-Zvi, B. Bukau and P. Goloubinoff (2000). "Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery." J Biol Chem **275**(28): 21107-21113.

Fay, A. and M. S. Glickman (2014). "An essential nonredundant role for mycobacterial DnaK in native protein folding." PLoS Genet **10**(7): e1004516.

Glaza, P., C. B. Ranaweera, S. Shiva, A. Roy, B. V. Geisbrecht, F. J. Schoenen and M. Zolkiewski (2021). "Repurposing p97 inhibitors for chemical modulation of the bacterial ClpB–DnaK bichaperone system." Journal of Biological Chemistry **296**: 100079.

Ismail, N., K. C. Bloch and J. W. McBride (2010). "Human ehrlichiosis and anaplasmosis." Clinics in laboratory medicine **30**(1): 261-292.

Katayama-Fujimura, Y., S. Gottesman and M. R. Maurizi (1987). "A multiple-component, ATP-dependent protease from *Escherichia coli*." Journal of Biological Chemistry **262**(10): 4477-4485.

Kuczynska-Wisnik, D., C. Cheng, R. R. Ganta and M. Zolkiewski (2017). "Protein aggregation in *Ehrlichia chaffeensis* during infection of mammalian cells." FEMS microbiology letters **364**(6): fnx059.

Kwiatkowska, J., E. Matuszewska, D. Kuczynska-Wisnik and E. Laskowska (2008). "Aggregation of *Escherichia coli* proteins during stationary phase depends on glucose and oxygen availability." Res Microbiol **159**(9-10): 651-657.

Li, D. H. S., Y. S. Chung, M. Gloyd, E. Joseph, R. Ghirlando, G. D. Wright, Y.-Q. Cheng, M. R. Maurizi, A. Guarné and J. Ortega (2010). "Acyldepsipeptide antibiotics induce the formation of a structured axial channel in ClpP: A model for the ClpX/ClpA-bound state of ClpP." Chemistry & biology **17**(9): 959-969.

Mattoo, R. U. H. and P. Goloubinoff (2014). "Molecular chaperones are nanomachines that catalytically unfold misfolded and alternatively folded proteins." Cellular and molecular life sciences : CMLS **71**(17): 3311-3325.

McClure, E. E., A. S. O. Chávez, D. K. Shaw, J. A. Carlyon, R. R. Ganta, S. M. Noh, D. O. Wood, P. M. Bavoil, K. A. Brayton, J. J. Martinez, J. W. McBride, R. H. Valdivia, U. G. Munderloh and J. H. F. Pedra (2017). "Engineering of obligate intracellular bacteria: progress, challenges and paradigms." Nat Rev Microbiol **15**(9): 544-558.

Meyer, H. and C. C. Weihl (2014). "The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis." Journal of cell science **127**(Pt 18): 3877-3883.

Moreno-Cinos, C., K. Goossens, I. G. Salado, P. Van Der Veken, H. De Winter and K. Augustyns (2019). "ClpP Protease, a Promising Antimicrobial Target." International journal of molecular sciences **20**(9): 2232.

Rikihisa, Y. (2010). "*Anaplasma phagocytophilum* and *Ehrlichia chaffeensis:* subversive manipulators of host cells." Nature Reviews Microbiology **8**(5): 328-339.

Rikihisa, Y. and M. Lin (2010). "*Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins." Current opinion in microbiology **13**(1): 59-66.

Singh, S. K., F. Guo and M. R. Maurizi (1999). "ClpA and ClpP remain associated during multiple rounds of ATP-dependent protein degradation by ClpAP protease." Biochemistry **38**(45): 14906-14915.

Sirigireddy, K. R. and R. R. Ganta (2005). "Multiplex detection of Ehrlichia and Anaplasma species pathogens in peripheral blood by real-time reverse transcriptase-polymerase chain reaction." J Mol Diagn **7**(2): 308-316.

Squires, C. L., S. Pedersen, B. M. Ross and C. Squires (1991). "ClpB is the *Escherichia coli* heat shock protein F84.1." Journal of Bacteriology **173**(14): 4254.

Van Horn, K. S., W. N. Burda, R. Fleeman, L. N. Shaw and R. Manetsch (2014). "Antibacterial Activity of a Series of N2,N4-Disubstituted Quinazoline-2,4-diamines." Journal of Medicinal Chemistry **57**(7): 3075-3093.

Wang, Y., A. D. S. Nair, A. Alhassan, D. C. Jaworski, H. Liu, K. Trinkl, P. Hove, C. K. Ganta, N. Burkhardt, U. G. Munderloh and R. R. Ganta (2020). "Multiple *Ehrlichia chaffeensis* Genes Critical for Its Persistent Infection in a Vertebrate Host Are Identified by Random Mutagenesis Coupled with In Vivo Infection Assessment." Infection and Immunity **88**(10): e00316-00320.

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

Zhang, T., S. Kedzierska-Mieszkowska, H. Liu, C. Cheng, R. R. Ganta and M. Zolkiewski (2013). "Aggregate-Reactivation Activity of the Molecular Chaperone ClpB from *Ehrlichia chaffeensis*." PLoS ONE **8**(5): e62454.

Zhang, Y., L. Chen, C. Kondethimmanahalli, H. Liu and R. R. Ganta (2021). "Protein and DNA Biosynthesis Demonstrated in Host Cell-Free Phagosomes Containing *Anaplasma phagocytophilum* or *Ehrlichia chaffeensis* in Axenic Media." Infection and Immunity: IAI.00638-00620.

Chapter 6 - Conclusions.

Molecular chaperones are essential proteins that bind to unfolded and partially folded polypeptide chains to prevent polypeptide aggregation and precipitation. Many molecular chaperones were first described as heat shock proteins (Hsp) because their expression is increased at elevated temperatures. Bacterial ClpB is a molecular chaperone, which is involved in reactivation of aggregated proteins. Yeast Hsp104 and plant Hsp100 are the eukaryotic orthologs of the bacterial ClpB. ClpB is the only known bacterial chaperone capable of cooperating with other heat-shock proteins (DnaK, DnaJ, GrpE) in suppressing and reversing protein aggregation.

In this dissertation we have studied substrate recognition mechanism of bacterial ClpB, we proposed that ClpB can be used as a novel antimicrobial target and studied interaction of potential inhibitor molecules.

In a cellular milieu, ClpB must accurately discriminate between properly folded, misfolded and aggregated. Only aggregated proteins will be selected by ClpB for processing. However, the molecular mechanism of substrate recognition and binding to ClpB is poorly understood. Therefore, to address this knowledge gap, we hypothesized that ClpB recognizes the aggregation-prone polypeptide segments that promote protein aggregation by recognizing specific physical and/or chemical surface properties and discrete sequences of these polypeptides. We proposed that such specific physical and/or chemical surface properties are likely present at the surface of growing aggregates. In Chapter 2 of this dissertation, I described *in vitro* experimental approaches to study substrate recognition mechanism of ClpB. We designed peptides containing discrete sequence segments that could promote beta-sheet driven aggregation and peptides without these discrete sequence segments. Our results confirmed that ClpB recognizes these peptides as pseudo-substrates, as postulated before. Unexpectedly our peptide interaction assay with ClpB revealed

that ClpB does recognize certain peptides even without the discrete aggregation-prone sequences. Therefore, we conclude that substrate recognition mechanism of ClpB may build upon global surface properties of aggregated proteins rather than on local sequence motifs.

As no apparent ClpB orthologs are found in metazoans and ClpB is essential for survival of bacteria, fungi, protozoa, and plants under stress, we proposed that ClpB is a promising target for the development of antimicrobials. The third chapter of this dissertation describes a group of small inhibitor molecules we had selected for initial screening. In summary, we investigated ClpB activity in the presence of several compounds that were previously described as inhibitors of the human AAA+ ATPase p97 (an antitumor target). Our initial screening revealed that N^2,N^4-dibenzylquinazoline-2,4-diamine (DBeQ), the least potent among the tested p97 inhibitors, binds to ClpB and inhibits the casein-activated ATPase activity of ClpB, while not affecting the basal ATPase activity ClpB.

To establish the effects of DBeQ on bacterial survival, we have used four different variants of *E. coli*, including the ClpB and DnaK deficient strains. Bacterial survival assays showed that DBeQ has a significant effect on *E. coli* proliferation under heat shock conditions, but the ClpB deficient *E. coli* strain managed to survive in the presence of DBeQ. We also used confocal fluorescence microscopy to track the localization of ClpB in *E. coli* cells under heat shock with and without DBeQ. We found that DBeQ enhances the targeting of ClpB to aggregated proteins in *E. coli*. These results demonstrated a significant selectivity of DBeQ toward ClpB in bacterial cells. Overall, our results were significant as it provided the chemical validation of ClpB as a target for developing novel antimicrobials. At the same time, we have identified DBeQ as a promising lead compound for future structural optimization aimed at selective targeting of ClpB and/or DnaK.

In the fourth chapter, I described protein aggregation patterns during infection of mammalian cells with *Anaplasma phagocytophilum* and the first biochemical characterization of ClpB and DnaK from *Anaplasma phagocytophilum*. *Anaplasma phagocytophilum* is an obligate intracellular Gram-negative bacterium which causes a disease known as Anaplasmosis. We cloned *Anaplasma* ClpB and DnaK genes and produced the recombinant proteins in *E.coli*. In order to establish a link between *A. phagocytophilum* infection, replication, and protein aggregate accumulation during infection, we propagated *A. phagocytophilum* in HL60 cells with or without 0.5 mM GuHCl, a known ClpB inhibitor. Our data for GuHCl treated samples indicated that the amounts of aggregated proteins increased gradually within first 48 hours and decrease gradually after 48 hours (Chapter 4), while untreated samples managed to maintain protein homeostasis and low protein aggregation levels. ClpB accumulated with aggregated fractions in both GuHCl-treated and untreated samples (Chapter 4 figure 4.7A). We did not observe any detectable levels of DnaK present in the soluble fractions, regardless of GuHCl treatment. However, we noticed that there were high amounts of DnaK present in non-GuHCl treated aggregated fractions.

Our *in vitro* biochemical data on ClpB and DnaK indicated that *Anaplasma* ClpB can complement certain functions of native *E. coli* ClpB. Our heat shock and confocal experiments with *Anaplasma* DnaK demonstrated that *Anaplasma* DnaK cannot complement functions of *E. coli* DnaK. Our ATPase activity for *Anaplasma* ClpB and DnaK demonstrate that these two chaperones have different substrate preferences. We tested *in vitro* substrate detection ability of purified *Anaplasma* ClpB and *Anaplasma* DnaK. Our binding data for *Anaplasma* ClpB revealed that *Anaplasma* ClpB has several-fold weaker affinities (as compared to *E. coli* ClpB) towards most of our model substrates. We did not observe a positive cooperativity in binding of the selected peptides to *Anaplasma* ClpB, contrary to what we observed with *E. coli* ClpB. Our *Anaplasma*

DnaK binding data also revealed that *Anaplasma* DnaK has several-fold weaker affinities (as compared to *E. coli* DnaK) towards all the model substrates. Despite large amino acid conservation of ClpB and DnaK among bacteria, perhaps the weaker substrate affinity of *Anaplasma* ClpB and DnaK arises from the difference in the living environment (extracellular vs intracellular).

While comparing all the binding data for *E. coli* and *Anaplasma phagocytophilum*, we observed that there is a pattern in substrate affinities of ClpB and DnaK. Namely, we observed that when ClpB recognizes a certain peptide with a higher affinity, then DnaK from the same species shows a relatively weaker affinity towards the same substrate. The opposite recognition pattern is also true, when DnaK recognizes a certain peptide with a higher affinity, then ClpB from the same species shows a relatively weaker affinity towards the same substrate.

In order to explain this substrate affinity variation between ClpB and DnaK, we proposed the following hypothesis. Namely, this chaperone property allows cells to make sure that all aggregated, misfolded and partially folded proteins are recycled through the DnaK-ClpB chaperone system. Having different substrate affinities for DnaK and ClpB towards the same substrate eliminates the competition between these chaperones for substrate attachment. At the same time, this mechanism ensures that if an aggregated protein is not recognized or missed by DnaK , then there is a second line of detection by ClpB. If an aggregate is detected by DnaK early, then eventually it will be delivered to ClpB for disaggregation. Perhaps this is nature's way of "division of labor of molecular chaperones".

The fifth chapter of this dissertation describes the results of protein aggregation studies conducted with wild type *Ehrlichia chaffeensis* and its ClpB-deficient strain. We studied replication patterns of wild type *E.chaffeensis* and its ClpB-deficient strain. Our replication data

revealed that the ClpB-deleted strain (TR91) had a significantly slower replication rate compared to the wild type *E.chaffeensis* strain. In the last part of the fifth chapter, we investigated the effect of DBeQ on *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* replication inside mammalian HL60 and DH82 cells.

First, we established that host cell replication was not significantly affected by DBeQ at the concentrations we intended to investigate. Our *in vivo* data revealed that 0.5 µM DBeQ inhibits replication of *Anaplasma phagocytophilum* in HL60 cells within 48 hours of infection. Wild type *E.chaffeensis* infection in DH82 cells was also cleared rapidly when treated with 0.1µM DBeQ (i.e. within 48 hours). These results confirmed that DBeQ possesses a potent antimicrobial activity. However, when DBeQ was used at 0.1 µM concentration with the *E.chaffeensis* ClpB-deficient strain (TR91), the clearing of infection was delayed up to 6 days post infection . These results suggest that ClpB is not the main target of DBeQ in *Ehrlichia chaffeensis.* However, our results demonstrated that DBeQ can be used as a lead to develop novel antibacterial compounds against Human Granulocytic Anaplasmosis (HGA) and Human Granulocytic Ehrlichiosis (HGE). It is important to determine the actual cellular target(s) of DBeQ in these intracellular pathogens. Perhaps, investigation of the DBeQ effects on global gene expression patterns in the host and pathogen cells using mRNA deep-sequencing technology may help us uncover the cellular targets of DBeQ in these models of infection.

Appendix A - Chapter 3 Supplementary Data.

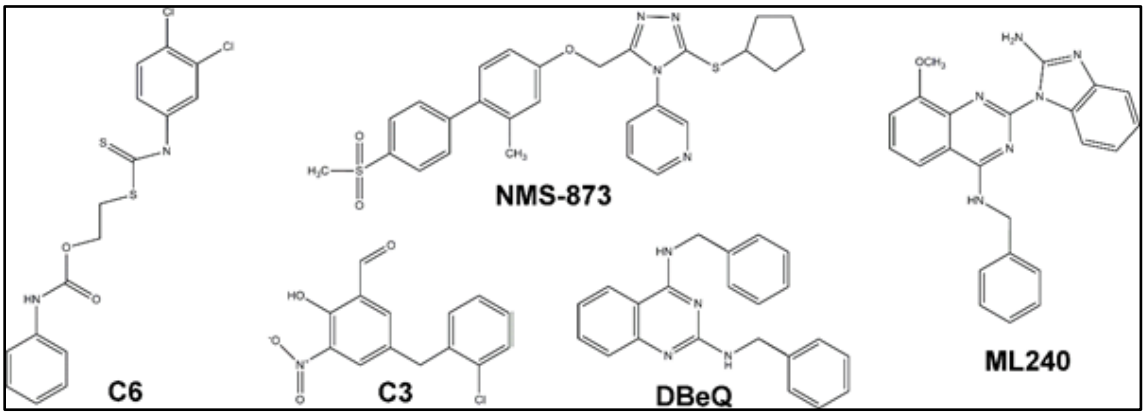
Supplementary Table A.1. Thermodynamic parameters for the interaction between DBeQ and ClpB or DnaK.

The values were obtained by fitting a binding isotherm equation with positive cooperativity to the SPR data for ClpB and a non-cooperative binding isotherm equation to the SPR data for DnaK. B_{max} indicates the SPR signal at saturation, r is the correlation coefficient, r^2 represents the goodness of the fit.

Protein	K_d [μM]	Hill coefficient	B_{max}	r^2
ClpB	64.4 ± 4.7	2.3 ± 0.3	1.09 ± 0.02	0.999
ClpB W462F	61.4 ± 4.7	2.4 ± 0.4	1.08 ± 0.03	0.998
ClpB W543F	59.1 ± 4.1	2.4 ± 0.4	1.07 ± 0.03	0.998
DnaK	105 ± 17	-	2.07 ± 0.20	0.996
DnaK NBD	95 ± 16	-	1.98 ± 0.20	0.995

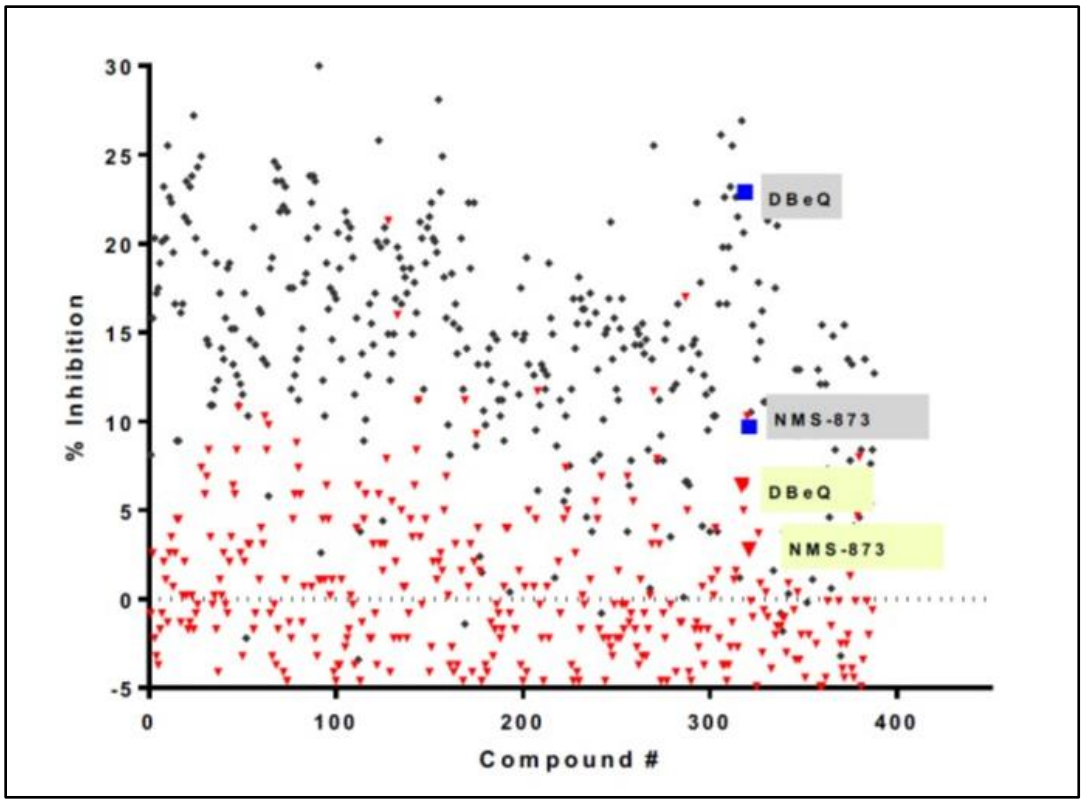
Supplementary Figure A.1. Chemical structure of the compounds investigated in this study.

(based on the structures supplied by the manufacturer and redrawn using ChemDraw 12).



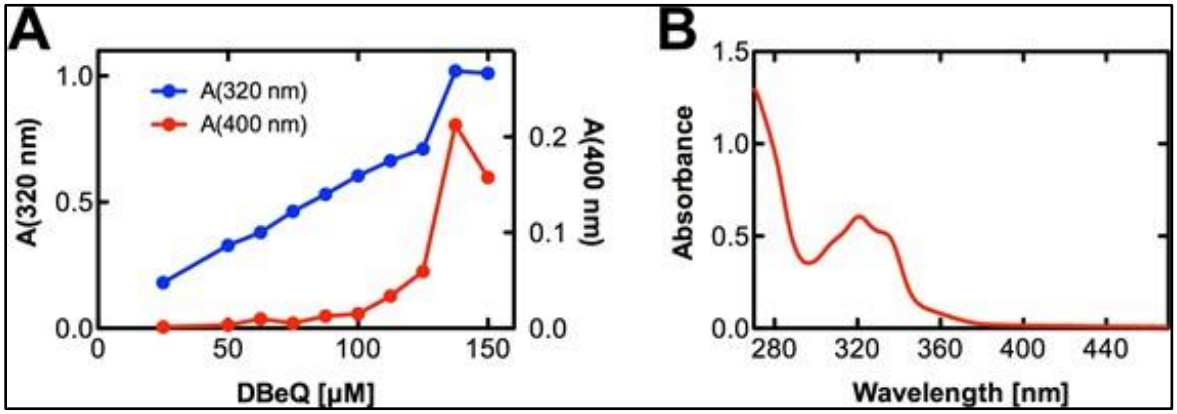
Supplementary Figure A.2. Pilot screen for inhibitors of the ClpB ATPase activity.

Shown is a scatter-graph of the ClpB ATPase inhibition across 388 compounds from Selleck Bioactive library screened in the presence (black) or absence (red) of 25 μM κ-casein. This experiment was performed by Dr Anuradha Roy.



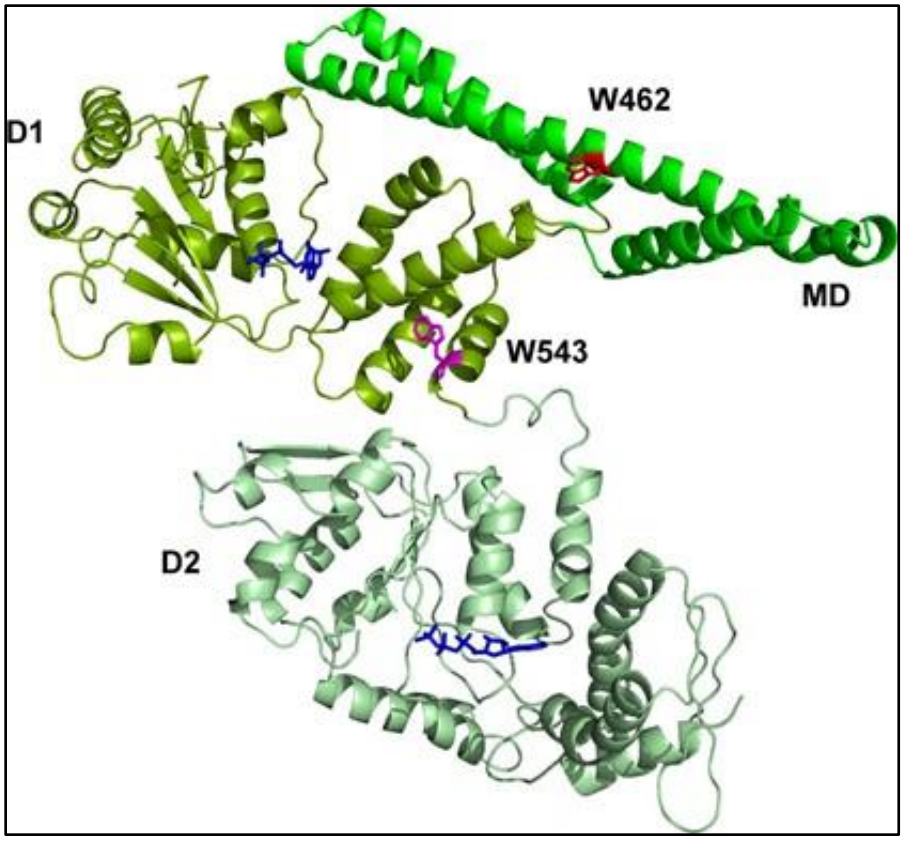
Supplementary Figure A.3. Solubility of DBeQ in aqueous solution.

Solutions of DBeQ were prepared in the ATPase assay buffer containing 1% DMSO. (A) DBeQ absorbance at 320 nm (blue) and solution turbidity at 400 nm (red) for increasing DBeQ concentrations. (B) UV absorption spectrum of 100 μM DBeQ. This experiment was performed by Dr Przemyslaw Glaza.



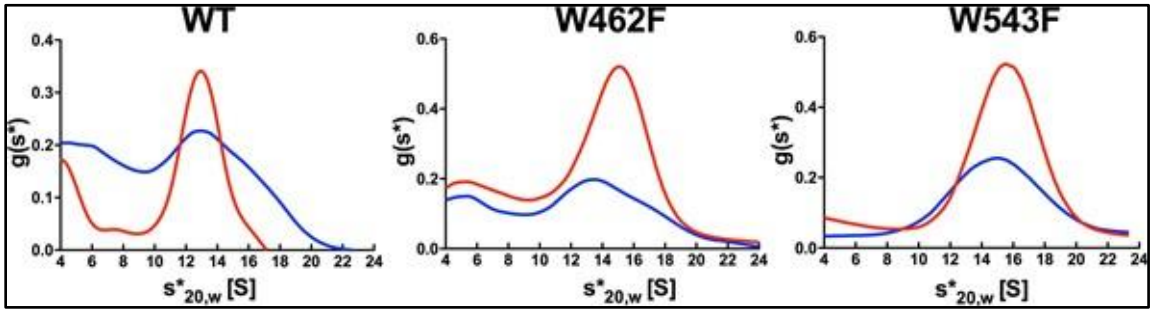
Supplementary Figure A.4. The ATP-binding modules (D1 and D2) separated with the middle domain (MD) in E. coli ClpB (PDB entry: 5OG1).

Trp462 (red), Trp543 (magenta), and the ATP analog, ATPγS (dark blue) bound in D1 and D2 are shown with a "stick" representation. Image generated by Dr Przemyslaw Glaza, using PyMOL 1.3 (Schrödinger LLC, www.pymol.com).



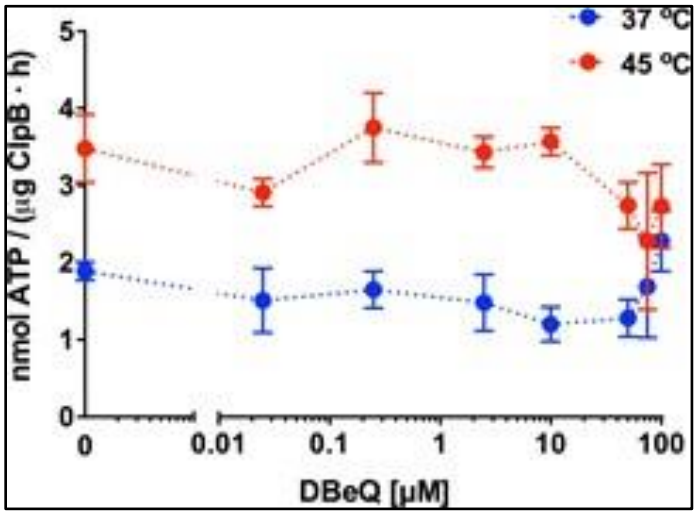
Supplementary Figure A.5. Sedimentation velocity analysis of wt ClpB, W462F, and W543F in the absence (blue) and presence (red) of 50 µM DBeQ.

Analytical ultracentrifugation was performed with 1 mg/ml ClpB in the presence of 1 mM ATPγS. Shown are the apparent distributions of sedimentation coefficient expressed in Svedbergs (S). This experiment was performed by Dr Michal Zolkiewski and Dr Przemyslaw Glaza.



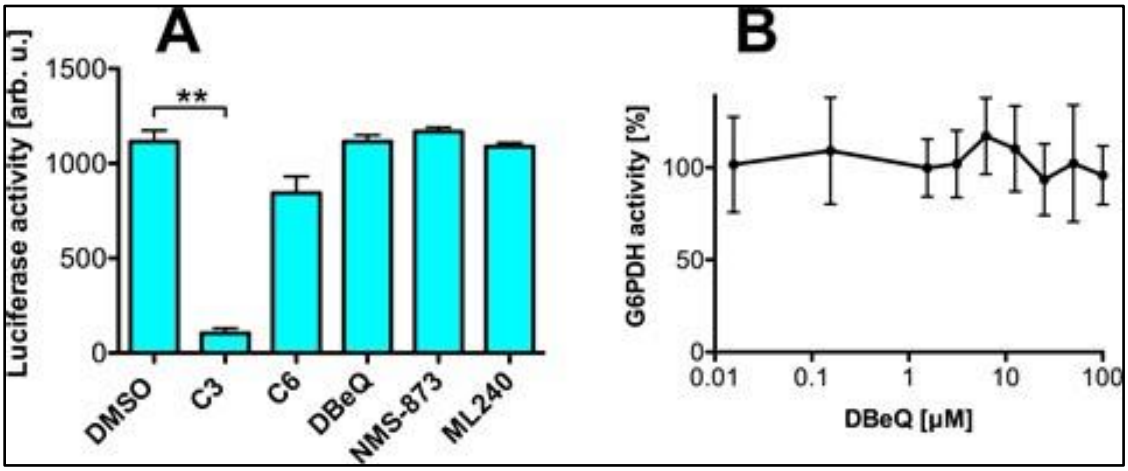
Supplementary figure A.6. DBeQ does not inhibit the ClpB ATPase under heat-shock conditions.

The ATPase activity of wt ClpB was determined at 37 and 45 °C as the function of DBeQ concentration. Average values from 3 measurements are shown with standard deviations. This experiment was performed by Dr Sunitha Shiva.



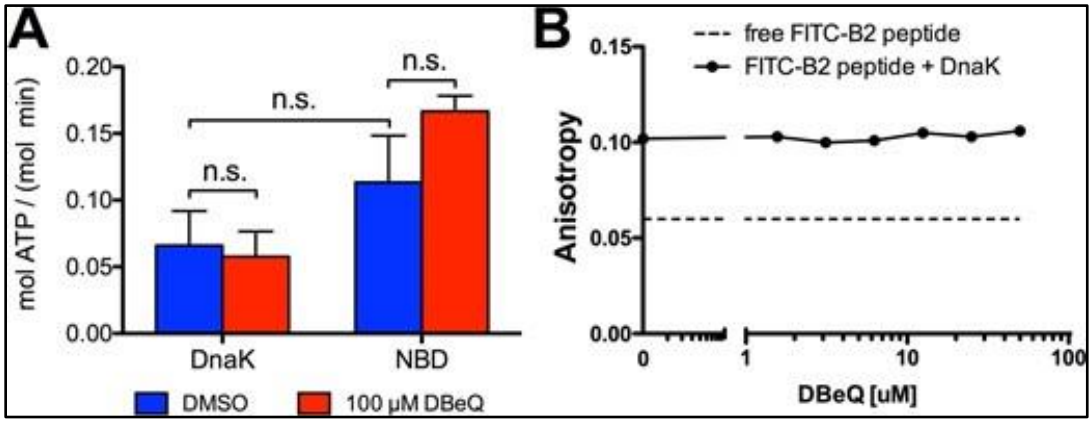
Supplementary Figure A.7. Effects of the inhibitors on the activity of luciferase and glucose-6-phosphate dehydrogenase (G6PDH).

(A) Firefly luciferase activity was determined in the absence (DMSO) or in the presence 100 µM ClpB inhibitor candidates. The only statistically significant difference is shown with ** ($p < 0.01$, $n = 2$). (B) G6PDH activity was determined in the presence of an increasing concentration of DBeQ and normalized to the DMSO control. Average values from two measurements are shown with standard deviations. This experiment was performed by Dr Przemyslaw Glaza.



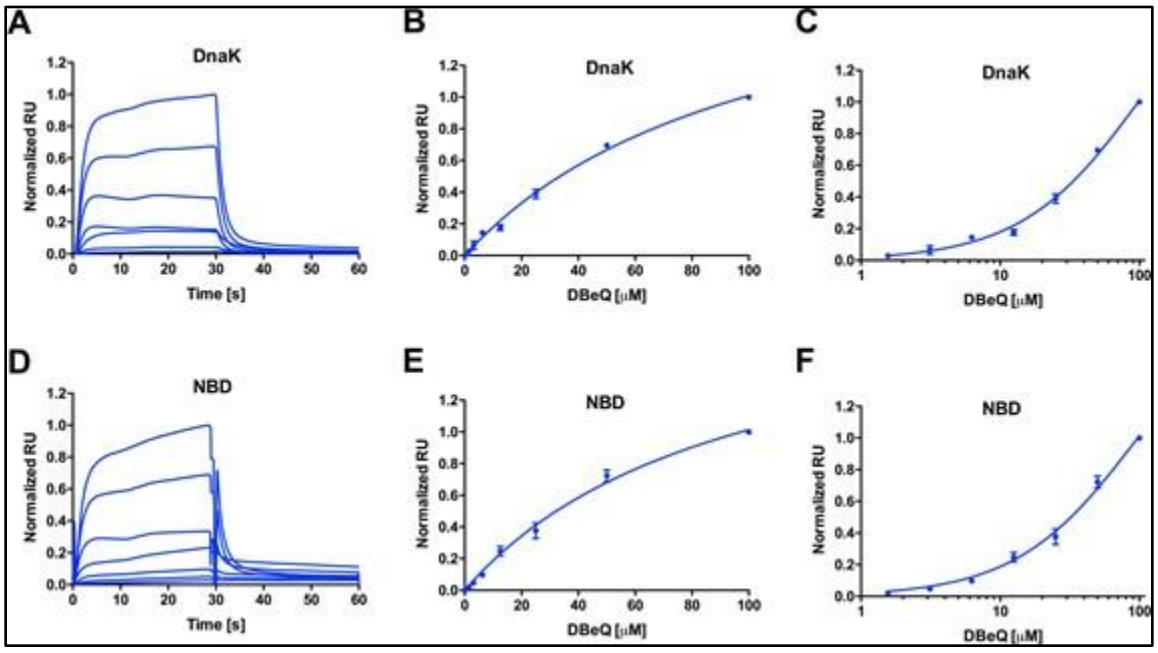
Supplementary Figure A.8. Effects of DBeQ on DnaK.

(A) The ATPase activity of DnaK and its nucleotide binding domain (NBD) in the absence of DBeQ (DMSO control, blue) and with 100 µM DBeQ (red). Shown are the values from 3 independent measurements with standard deviations (n.s. indicates that the difference between the data sets is not statistically significant, $p > 0.05$). (B) Fluorescence anisotropy of FITC-B2 (20 nM) in the absence (broken line) and presence of DnaK (solid line) at increasing DBeQ concentrations. This experiment (Supplementary figure 8A) was performed by Dr Sunitha Shiva.



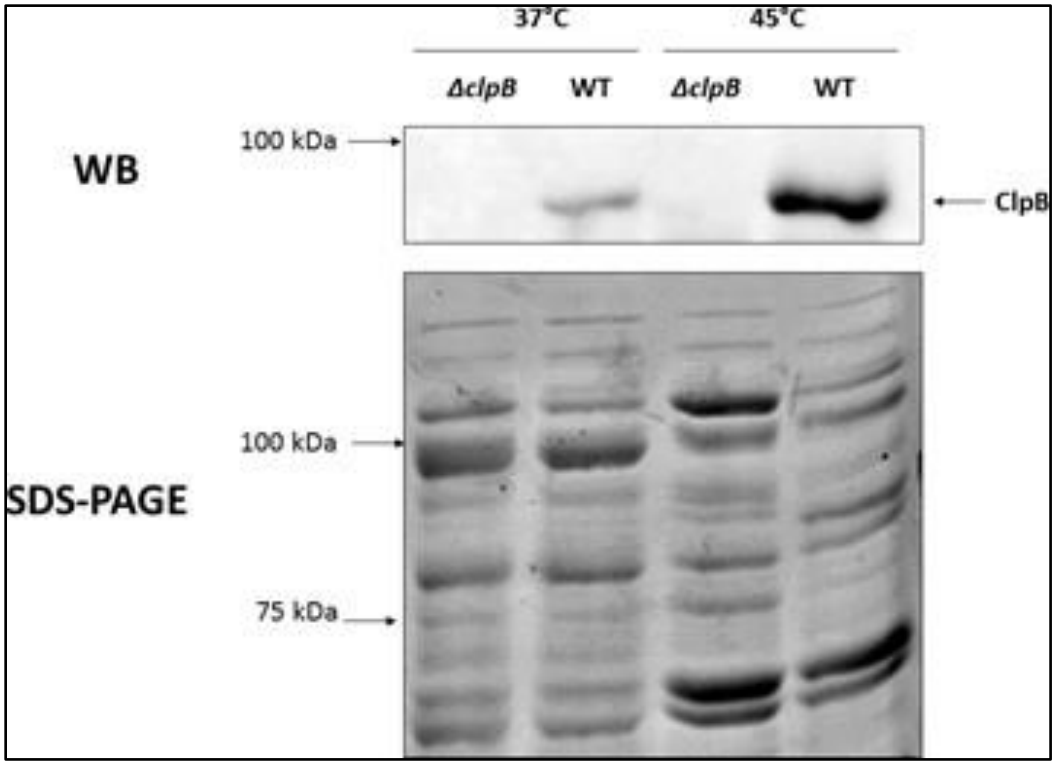
Supplementary Figure A.9. Surface Plasmon Resonance (SPR) analysis of the binding of DBeQ to DnaK (A-C) and its nucleotide binding domain (NBD) (D-F).

(A, D) Representative SPR sensograms for the DnaK variants. Each solid line represents a sensogram obtained with a given DBeQ concentration; a higher DBeQ concentration produces a higher SPR response. DBeQ binding isotherms are shown with the linear ligand concentration scale (B, E) or the logarithmic scale (C, F). Shown are the averages with standard deviations from three repeated experiments. Solid lines in the panels B, C, E, and F represent the fits of the non-cooperative binding model with the parameters listed in Supplementary Table 1. This experiment was performed with Dr Brian V. Geisbrecht, , Dr. Przemyslaw Glaza.



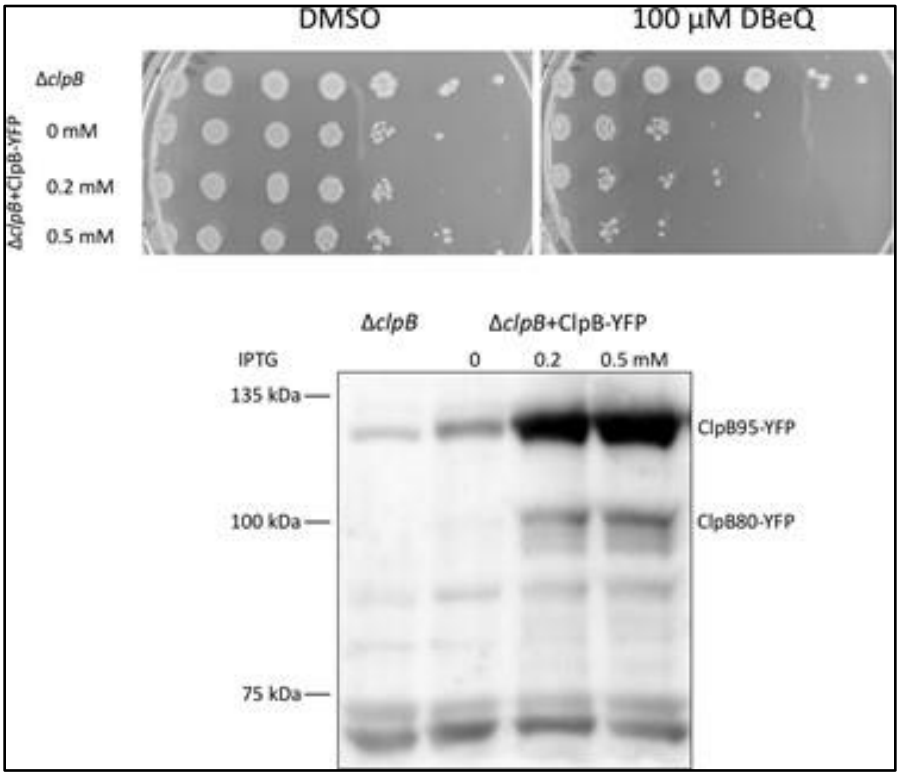
Supplementary Figure A.10. Analysis of the ClpB protein levels in *E. coli*.

Total cell lysates from MC4100 (WT) and MC4100ΔclpB (ΔclpB) *E. coli* strains grown at 37 °C and 45 °C were resolved with SDS-PAGE in an 8% gel. (Top) ClpB was detected using Western blotting. (Bottom) Coomassie Brilliant Blue staining of the same samples as in the top resolved in the control part of the same gel.



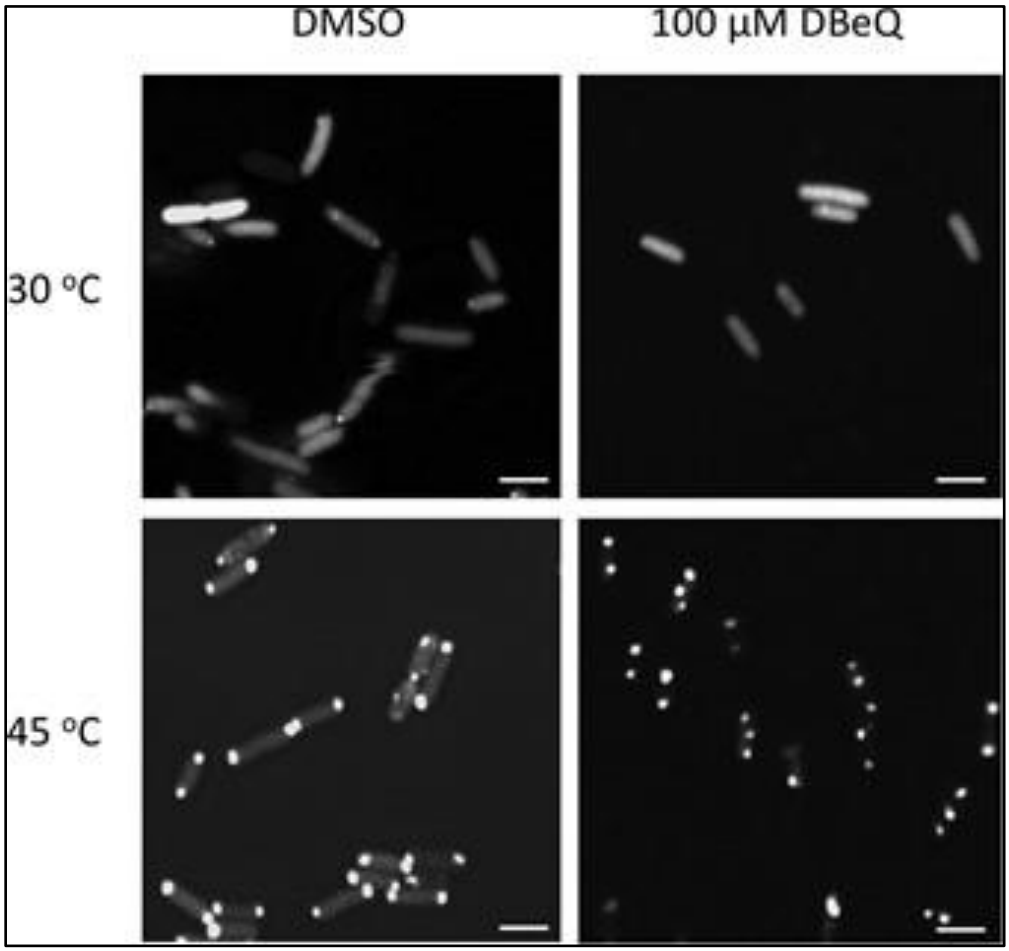
Supplementary Figure A.11. Viability of ΔclpB *E. coli* strain upon expression of ClpB- YFP.

(Top) The *E. coli* cultures were incubated at 45 °C in the presence of DBeQ or with only DMSO for 3 hours then spotted on agar plates and incubated overnight at 37 °C. Expression of ClpB-YFP was induced with the IPTG concentration indicated on the left. Each spot on the agar plates represents a viable culture after a 10 fold serial dilution (from left to right). (Bottom) Western blot analysis of the bacterial lysates with anti-ClpB antibody. The cultures were incubated with the indicated IPTG concentration for 3 h. The detected signals correspond to the full-length ClpB-YFP (ClpB95-YFP) and the N-terminally truncated ClpB80-YFP (1,2). The nonspecific band below 75 kDa serves as a loading control.



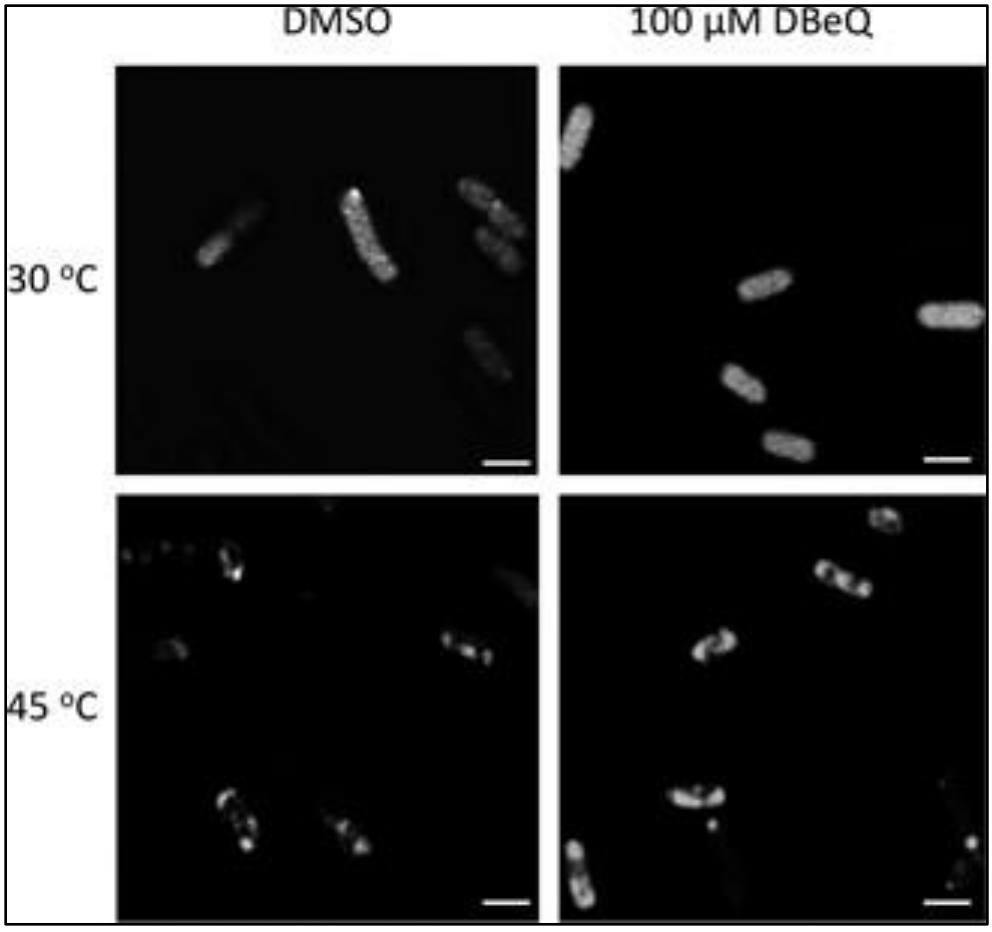
Supplementary Figure A.12. Localization of YFP-luciferase in *E. coli* cells under heat shock.

MC4100 cells were grown at 30 °C and then shifted to 45 °C for 30 min. The images show YFP fluorescence signal. Representative images from 3 independent experiments are shown. The white bar in each panel corresponds to 2 μm.



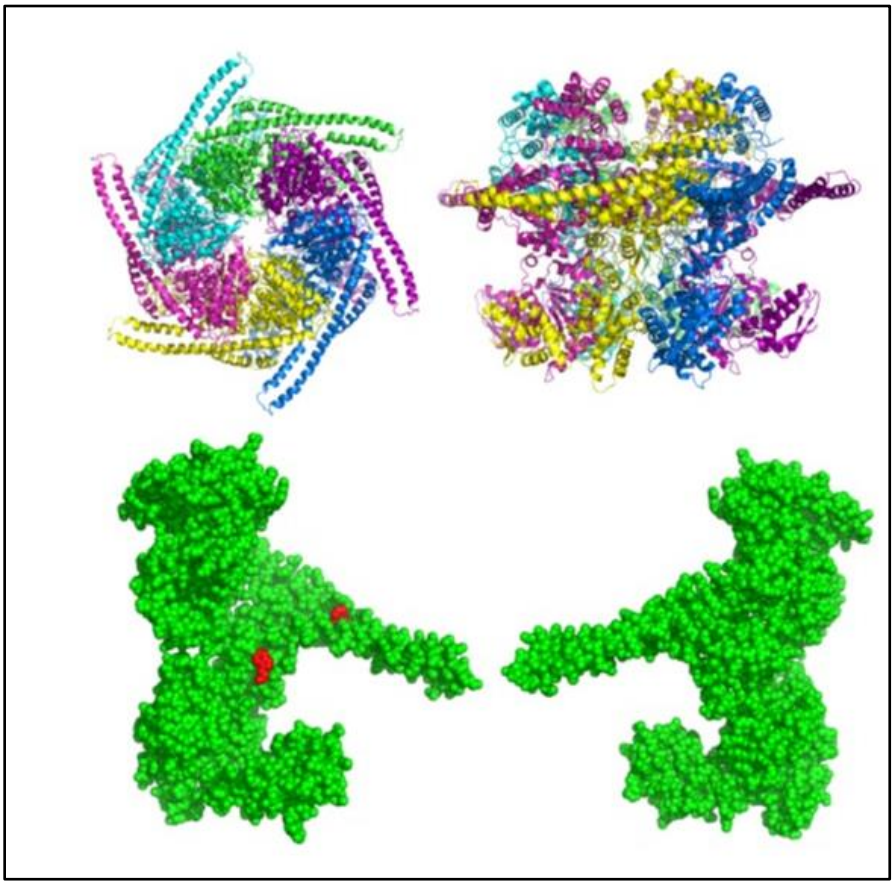
Supplementary Figure A.13. Localization of ClpB-YFP in the dnaK756 *E. coli* strain under heat shock.

MC4100 cells were grown at 30 °C and then shifted to 45 °C for 30 min. The images show YFP fluorescence signal. Representative images from 3 independent experiments are shown. The white bar in each panel corresponds to 2 μm.



Supplementary Figure A.14. (Top) Model of the ClpB hexamer (top and side view) assembled from the *Thermus thermophilus* ClpB monomer structure (PDB Id: 1QVR, chain A).

Each subunit in the hexamer is shown in a different color. (Bottom) Space-filling model of the *T. thermophilus* ClpB monomer viewed from the inside of the hexamer (left) and from the outside of the hexamer (right). The side chains of the two Trp residues are shown in red (see Supplementary Figure 4). The indicated Trp residues in ClpB are conserved between E. coli and T. thermophilus. Image generated by Dr Przemyslaw Glaza, using PyMOL 1.3 (Schrödinger LLC, www.pymol.com).



References for supplementary material:

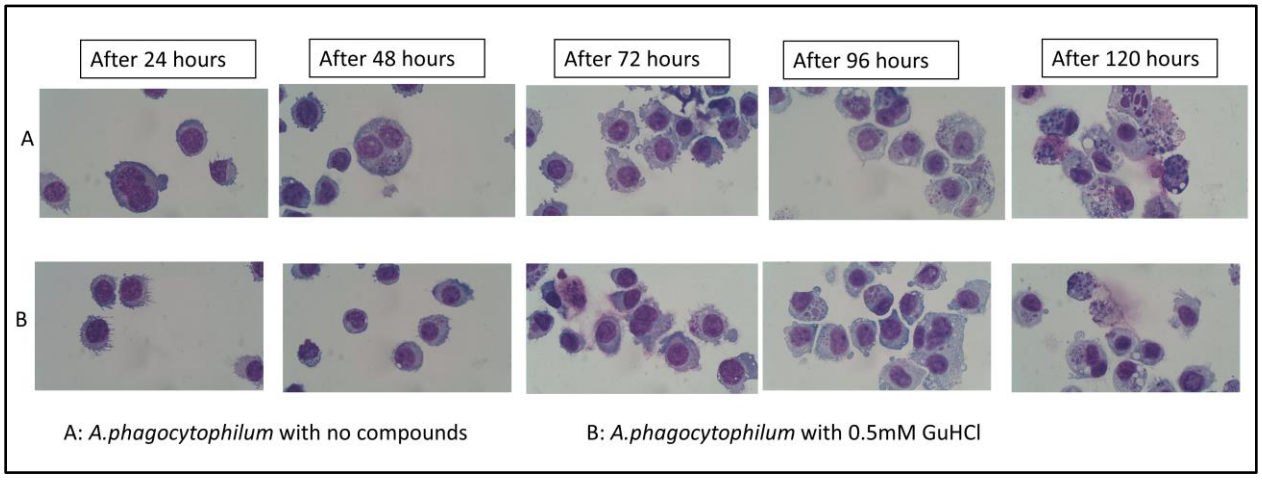
1. Squires, C.L., Pedersen, S., Ross, B.M., and Squires, C. (1991) ClpB is the Escherichia coli heat shock protein F84.1. J. Bacteriol. 173, 4254-4262.

2. Chow, I.T., Barnett, M.E., Zolkiewski, M., and Baneyx, F. (2005) The N-terminal domain of Escherichia coli ClpB enhances chaperone function. FEBS Lett. 579, 4242- 4248.

Appendix B - Chapter 4 Supplementary Data.

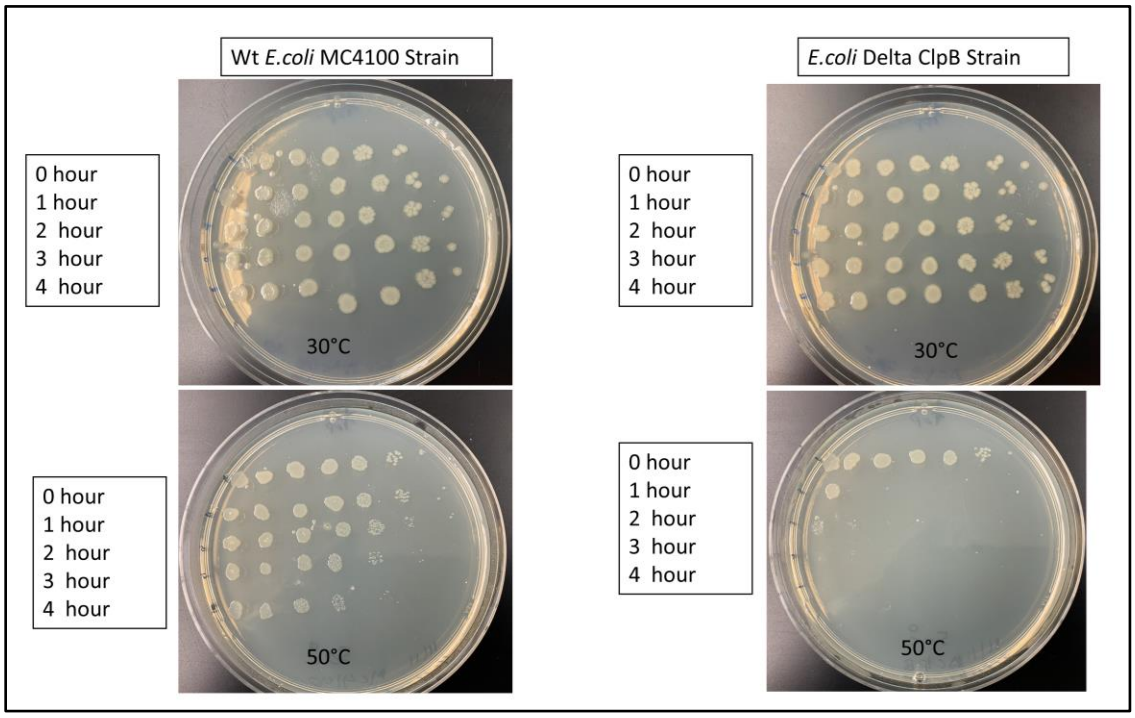
B.1 *A. phagocytophilum* replication in HL60 cells with or without 0.5 mM GuHCl.

Cells were grown (with or without GuHCl) as described earlier in this chapter and 200 µl sample was withdrawn from the T25 flask at each post infection time point. Sample was stained by hematoxylin stain and observed under a light microscope. Image was captured and generated using ACCU-SCOPE Excelis camera and CaptaVision+™ Software. Image magnification of 1000.Under these different growth conditions, *A.phagocytophilum* managed to invade HL60 cells and replicate successfully.



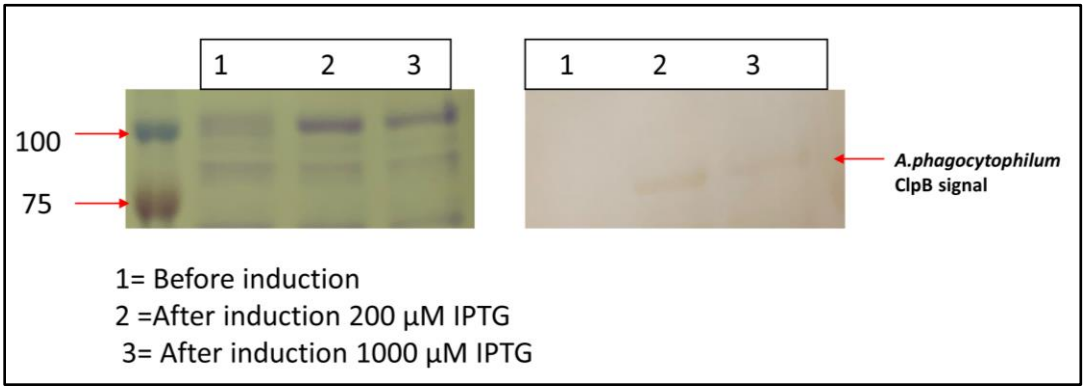
B.2 Survival of wild type *E. coli* (MC4100) and *E. coli* ClpB deleted (MC4100ΔclpB) strains at 50°C.

Cells were grown either at 30°C or under severe heat shock conditions at 50°C for the period indicated on the left (in hours) and then spotted on agar plates and incubated overnight at 37 °C. Each spot on the agar plates represents a viable culture after a 10-fold serial dilution (from left to right). Wild type *E. coli* cells managed to withstand severe heat shock conditions successfully while ClpB deleted *E. coli* strain died within first two hours. ClpB gene is essential for survival under severe heat shock conditions.



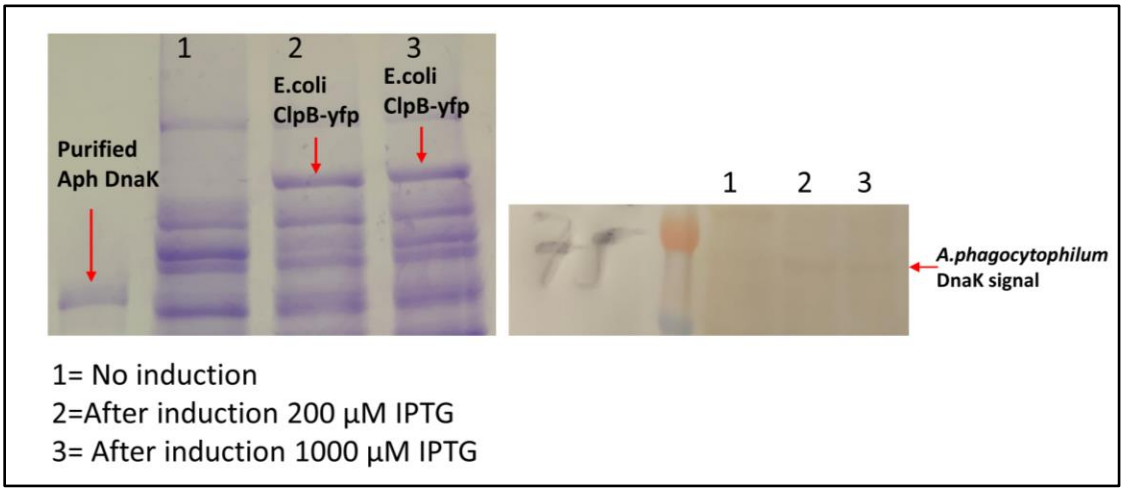
B.3 Recombinant *Anaplasma* ClpB expression in *E. coli* ClpB deleted (MC4100ΔclpB) strain.

Recombinant *Anaplasma* ClpB plasmid and plasmid pCS6 (for T7 RNA polymerase production) was transformed into MC4100ΔclpB strain and induced with IPTG at 37°C for six hours. SDS- PAGE gel stained by Coomassie staining and Western blot membrane are shown here. Our results demonstrated that Anaplasma ClpB protein was successfully induced in *E. coli* (MC4100ΔclpB) strain. Equal amounts of proteins were loaded in each well.



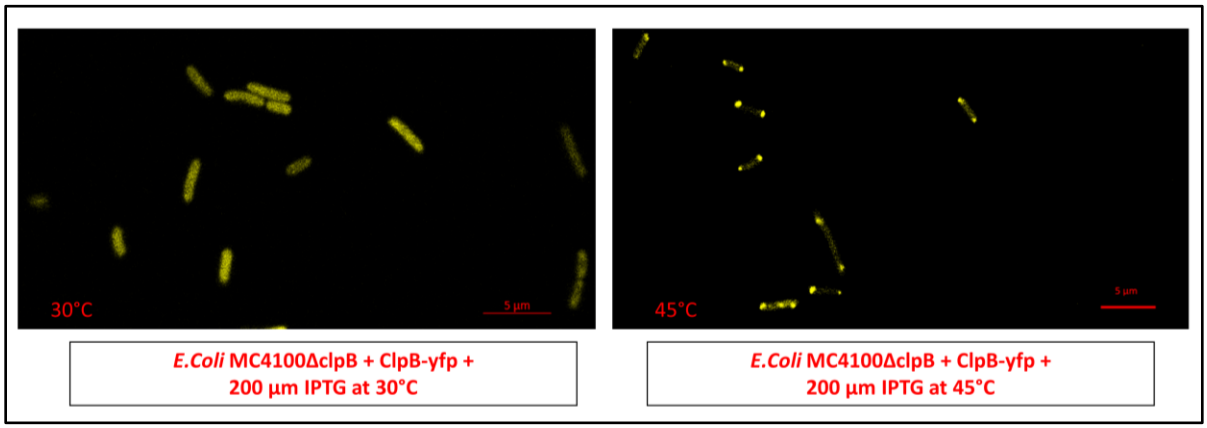
B.4 Recombinant *Anaplasma* DnaK expression in *E. coli* DnaK103 strain.

Recombinant *Anaplasma* ClpB plasmid ,plasmid pCS6 (for T7 RNA polymerase production) , ClpB-YFP plasmids were transformed into *E.coli* DnaK103 strain and induced with IPTG at 37°C for six hours. SDS- PAGE gel stained by Coomassie staining and Western blot membrane are shown here. Even though protein induction signals were not that prominent in both SDS-PAGE gel and Western blot , there was specific signal for the production of *Anaplasma* DnaK protein in *E. coli* Dnak103 cells. ClpB-YFP (~122 kDa) was also induced by IPTG. Equal amount of proteins were loaded in each well.



B.5 Migration of *E. coli* ClpB-YFP proteins under heat shock.

E. coli delta ClpB strain (MC4100ΔclpB) was transformed with ClpB-YFP plasmid and ClpB-YFP protein induction was induced by 200µM IPTG. After 04 hours of induction at 37°C, protein translation was stopped by addition of erythromycin (30 µg/ml), and each culture was divided into two samples. One sample remained at 30 °C (control), whereas the other sample was transferred to 45 °C for 30 minutes (heat-shock conditions).Confocal Image processing was done as mentioned in material and methods. As shown in earlier publications (Winkler, Seybert et al. 2010, Glaza, Ranaweera et al. 2020)When there is no heat shock, ClpB-YFP is seen diffused throughout the cytoplasm. But when cells are under heat shock, ClpB-YFP gets localized to the ends of the cell (where protein aggregates are located). The scale bar is located at bottom right-hand corner of each image and represents 5µm.

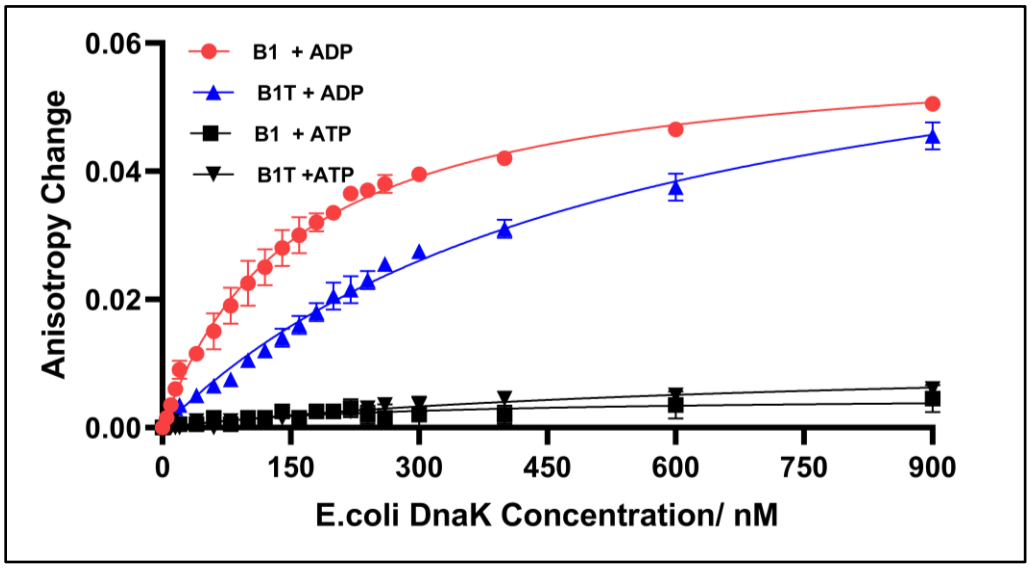


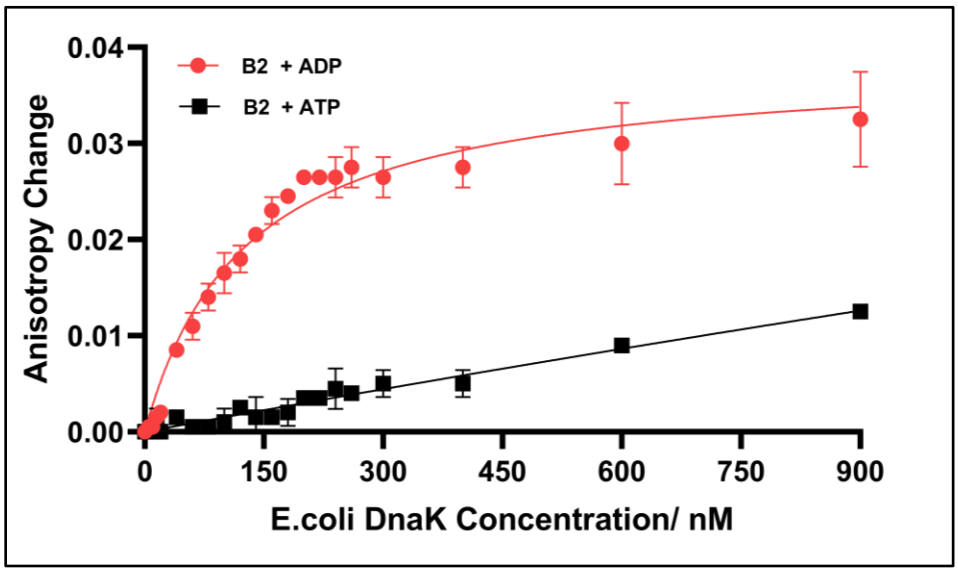
B.6 *E. coli* DnaK binding isotherms for substrate mimicking peptides.

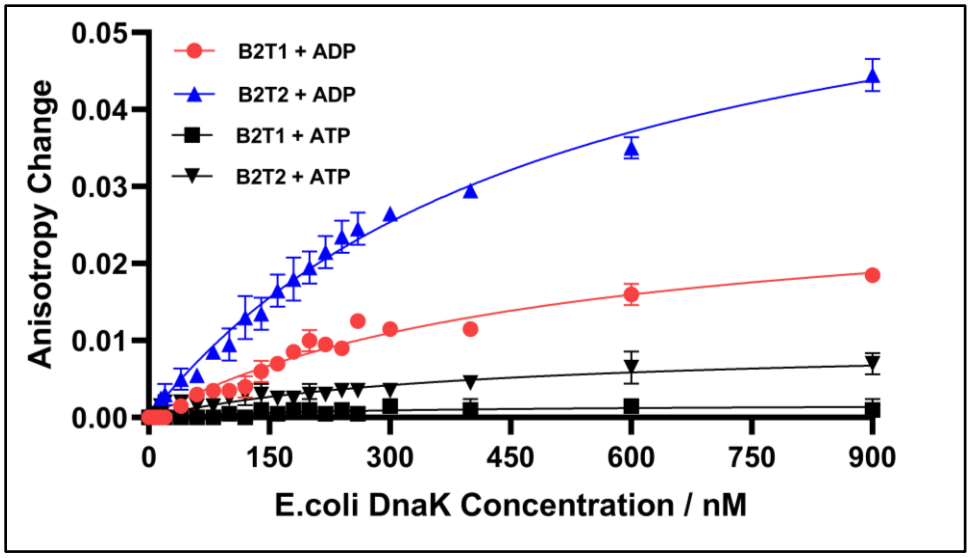
Purified *E. coli* DnaK were used for fluorescence anisotropy binding experiments. These experiments were performed as described with the same Fluorescein isothiocyanate (FITC) labelled peptides mentioned in our earlier publication (Ranaweera, Glaza et al. 2018). Peptides, peptide sequences, binding isotherms and apparent dissociation constants (Kd) values for *E.coli* ClpB and DnaK are shown here.

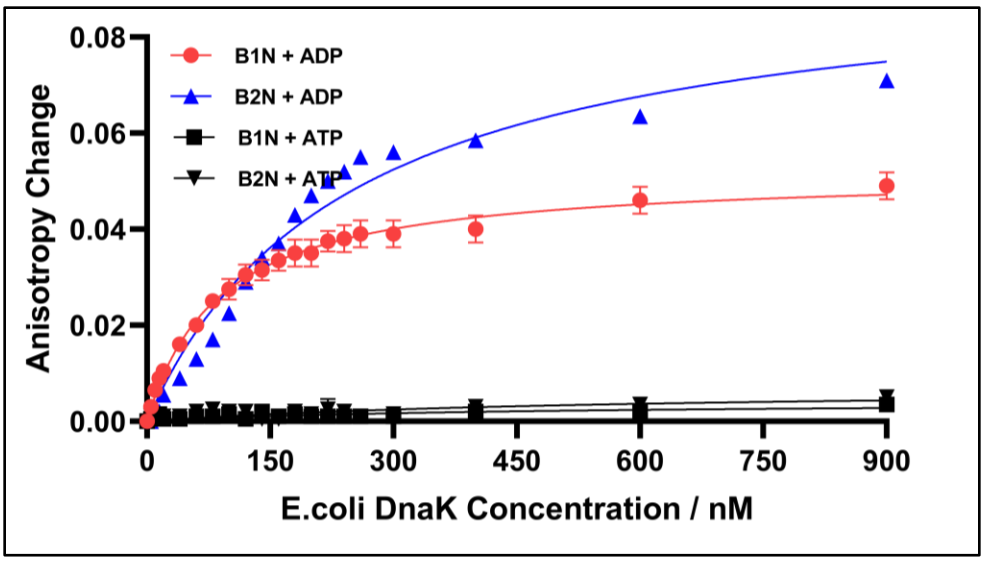
Peptide	Sequence
B1	AHAWQHQGKTLFISRKTYRI
B1T	KTLFISR
B1N	FHAWQHQGKTLAISRKTYRI
B2	QRKLFFNLRKTKQRLGWFNQ
B2T1	KLFFNLR
B2T2	RLGWFNQ
B2N	QLWQGQFRNLRRKTFFKKLN
B2A	QGQFRNLRRKQLWTFFLNKK
B2B	QQFRNRRKLTFLKKLGWFNQ

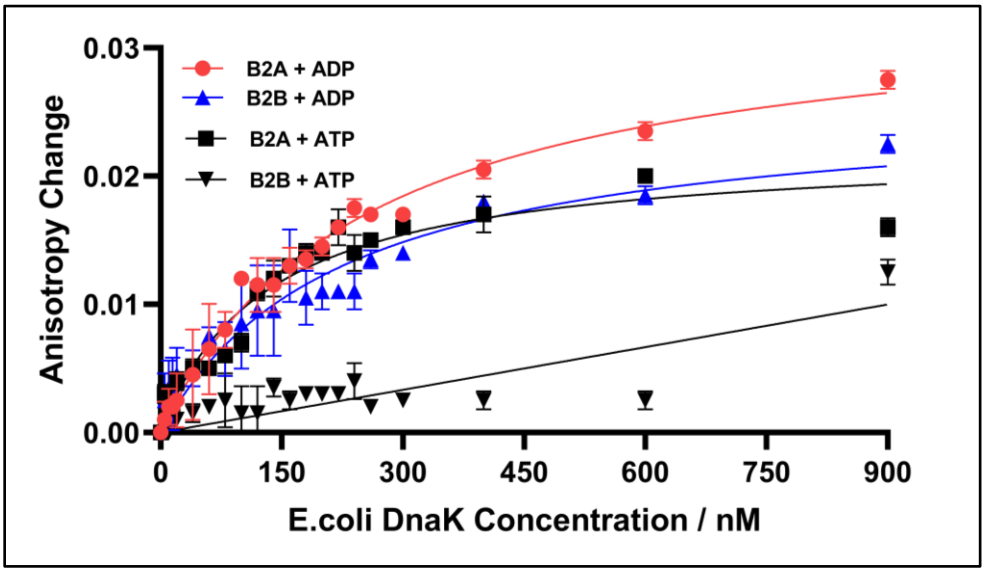
Peptides	*E.coli* ClpB Kd [nM]	*E.coli* DnaK Kd [nM]
B1	46.8±1.5	157.6
B1T	329 ±39	546.3
B1N	7.26 ±0.32	87.8
B2	13.24 ±0.42	126.7
B2T1	912 ±231	498.3
B2T2	>1000	513.2
B2N	17.54 ±0.92	245.7
B2A	10.17 ±0.80	251.6
B2B	11.51 ±0.52	224.3




0.04
0.03
0.02
0.01
0.00
Anisotropy Change
B2 + ADP
B2 + ATP
0
150
300
450
600
750
900
E.coli DnaK Concentration/ nM


0.05
0.04
0.03
0.02
0.01
0.00
Anisotropy Change
B2T1 + ADP
B2T2 + ADP
B2T1 + ATP
B2T2 + ATP
0
150
300
450
600
750
900
E.coli DnaK Concentration / nM


0.08
0.06
0.04
0.02
0.00
Anisotropy Change
B1N + ADP
B2N + ADP
B1N + ATP
B2N + ATP
0
150
300
450
600
750
900
E.coli DnaK Concentration / nM


0.03
0.02
0.01
0.00
Anisotropy Change
B2A + ADP
B2B + ADP
B2A + ATP
B2B + ATP
0
150
300
450
600
750
900
E.coli DnaK Concentration / nM

B.7 Sequence alignment of ClpB proteins from several human pathogens.

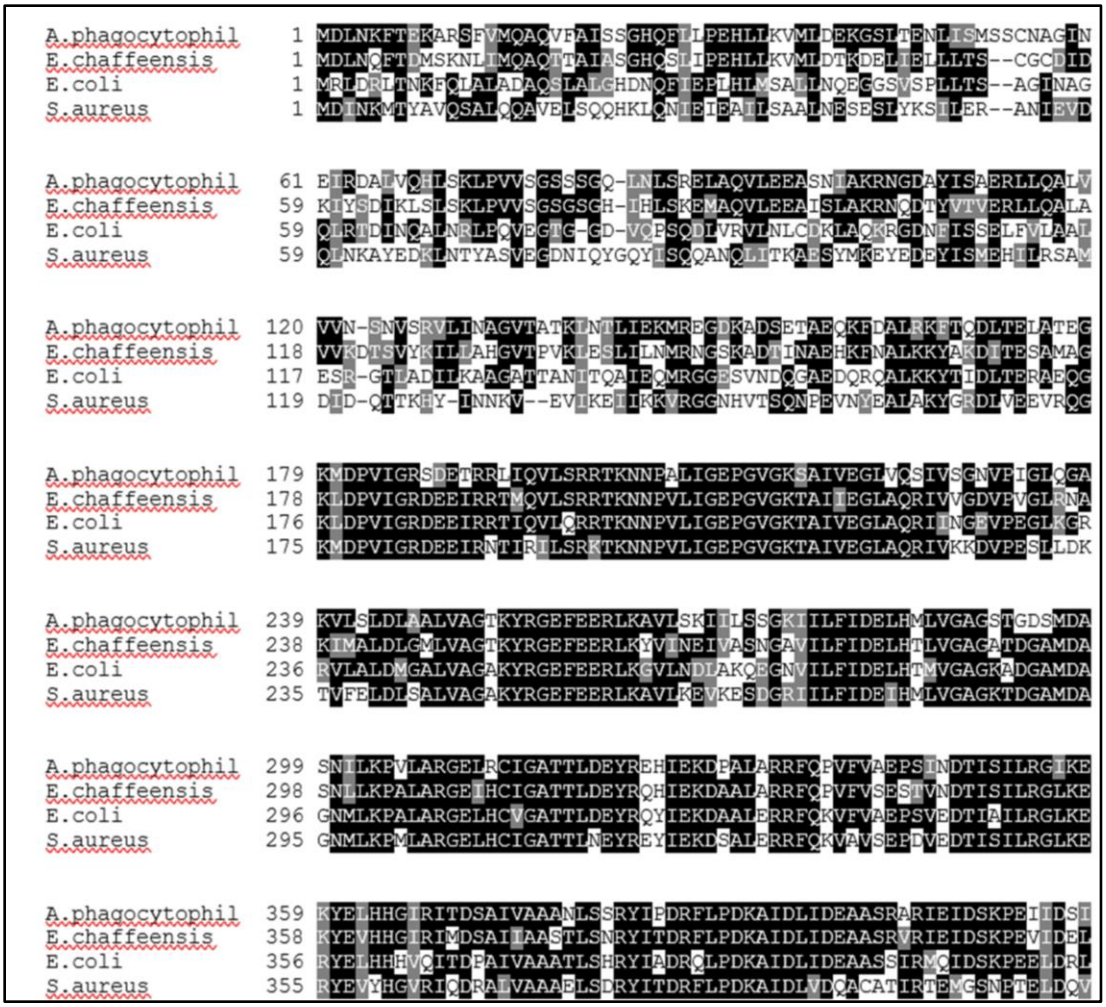
ClpB sequence from different pathogens have been compared using Clustal Omega tool provided by Bioinformatics Institute (EMBL-EBI, https://www.ebi.ac.uk/Tools/msa/clustalo/), followed by BOXSHADE 3.21 tool, from the Swiss Institute of Bioinformatics (https://embnet.vital-it.ch/software/BOX_form.html) which shaded multiple alignments. Black shaded boxes indicate conserved residues, grey indicates similar amino acids and white indicates different amino acids. (-) indicates the gaps introduce by Clustal Omega program when aligning these sequences. Full details of ClpB sequence sources are listed below.

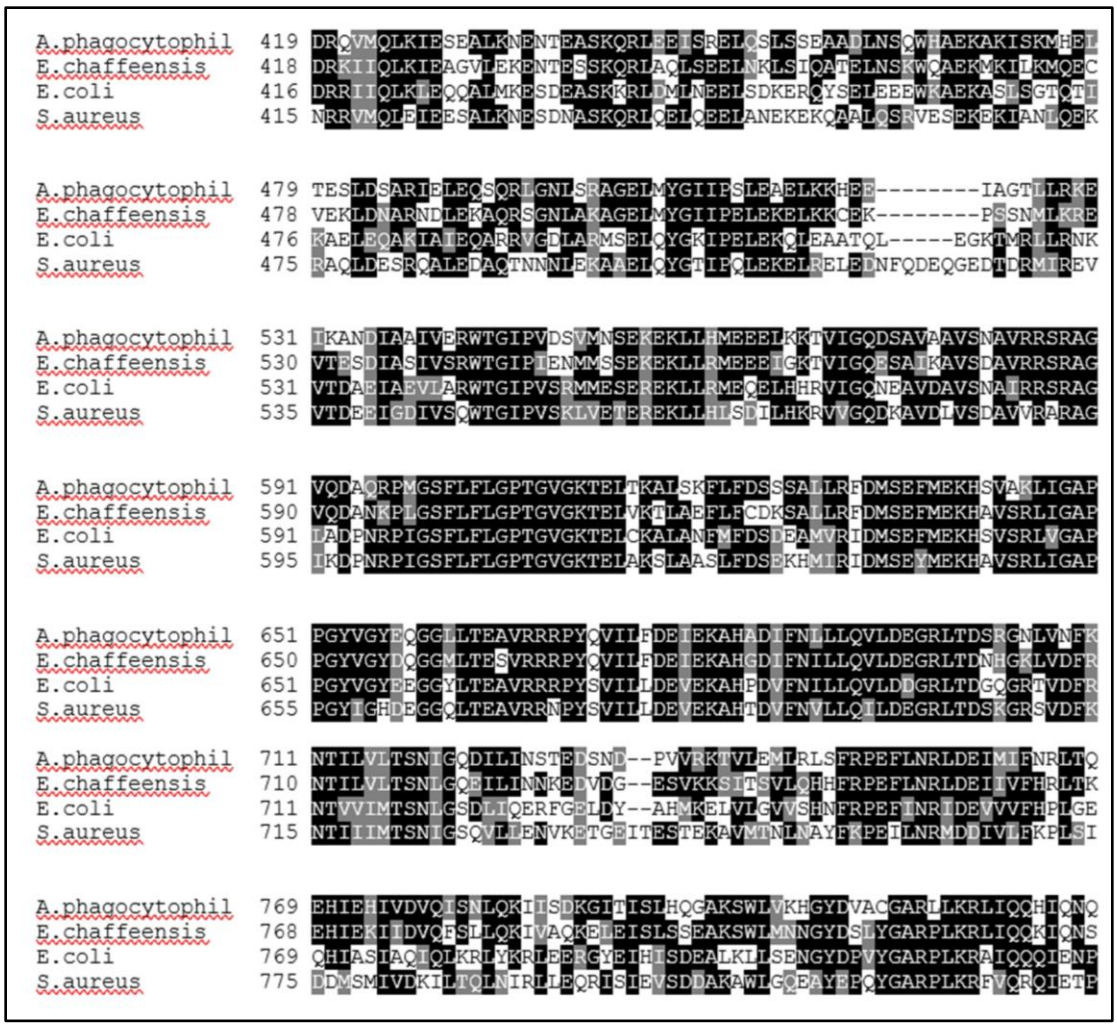
>tr|Q2GL99|Q2GL99_ANAPZ ATP-dependent chaperone protein ClpB OS=*Anaplasma phagocytophilum* (strain HZ) OX=212042 GN=clpB PE=3 SV=1

>tr|Q2GH96|Q2GH96_EHRCR Chaperone protein ClpB OS=*Ehrlichia chaffeensis* (strain ATCC CRL-10679 / Arkansas) OX=205920 GN=clpB PE=3 SV=1

>sp|P63284|CLPB_ECOLI Chaperone protein ClpB OS=*Escherichia coli* (strain K12) OX=83333 GN=clpB PE=1 SV=1

>sp|Q7A6G6|CLPB_STAAN Chaperone protein ClpB OS=*Staphylococcus aureus* (strain N315) OX=158879 GN=clpB PE=1 SV=1


A.phagocytophil 1 MDLNKFTEKARSFVMQAQVFAISSGHQFLPEHLLKVMLDEKGSLTENLISMSSCNAGIN
E.chaffeensis 1 MDLNQFTDMSKNLIMQAQTTAIASGHQSLIPEHLLKVMLDTKDELIELLLTS--CGCDID
E.coli 1 MRLDRLTNKFQLALADAQSLALGHDNQFIEPLHLMSALLNQEGGSVSPLLTS--AGINAG
S.aureus 1 MDINKMTYAVQSALQQAVELSQQHKLQNIEIEAILSAALNESESLYKSILER--ANIEVD
A.phagocytophil 61 EIRDALVQHLSKLPVVSGSSSGQ-LNLSRELAQVLEEASNIAKRNGDAYISAERLLQALV
E.chaffeensis 59 KIYSDIKLSLSKLPVVSGSGSGH-IHLSKEMAQVLEEAISLAKRNQDTYVTVERLLQALA
E.coli 59 QLRTDINQALNRLPQVEGTG-GD-VQPSQDLVRVLNLCDKLAQKRGDNFISSELFVLAAL
S.aureus 59 QLNKAYEDKLNTYASVEGDNIQYGQYISQQANQLITKAESYMKEYEDEYISMEHILRSAM
A.phagocytophil 120 VVN-SNVSRVLINAGVTATKLNTLIEKMREGDKADSETAEQKFDALRKFTQDLTELATEG
E.chaffeensis 118 VVKDTSVYKILLAHGVTPVKLESLILNMRNGSKADTINAEHKFNALKKYAKDITESAMAG
E.coli 117 ESR-GTLADILKAAGATTANITQAIEQMRGGESVNDQGAEDQRQALKKYTIDLTERAEQG
S.aureus 119 DID-QTTKHY-INNKV--EVIKEIIKKVRGGNHVTSQNPEVNYEALAKYGRDLVEEVRQG
A.phagocytophil 179 KMDPVIGRSDETRRLIQVLSRRTKNNPALIGEPGVGKSAIVEGLVQSIVSGNVPIGLQGA
E.chaffeensis 178 KLDPVIGRDEEIRRTMQVLSRRTKNNPVLIGEPGVGKTAIIEGLAQRIVVGDVPVGLRNA
E.coli 176 KLDPVIGRDEEIRRTIQVLQRRTKNNPVLIGEPGVGKTAIVEGLAQRIINGEVPEGLKGR
S.aureus 175 KMDPVIGRDEEIRNTIRILSRKTKNNPVLIGEPGVGKTAIVEGLAQRIVKKDVPESLLDK
A.phagocytophil 239 KVLSLDLAALVAGTKYRGEFEERLKAVLSKIILSSGKIILFIDELHMLVGAGSTGDSMDA
E.chaffeensis 238 KIMALDLGMLVAGTKYRGEFEERLKKVINEIVASNGAVILFIDELHTLVGAGATDGAMDA
E.coli 236 RVLALDMGALVAGAKYRGEFEERLKGVLNDLAKQEGNVILFIDELHTMVGAGKADGAMDA
S.aureus 235 TVFELDLSALVAGAKYRGEFEERLKAVLKEVKESDGRIILFIDEIHMLVGAGKTDGAMDA
A.phagocytophil 299 SNILKPVLARGELRCIGATTLDEYREHIEKDPALARRFQPVFVAEPSINDTISILRGIKE
E.chaffeensis 298 SNLLKPALARGEIHCIGATTLDEYRQHIEKDAALARRFQPVFVSESTVNDTISILRGLKE
E.coli 296 GNMLKPALARGELHCVGATTLDEYRQYIEKDAALERRFQKVFVAEPSVEDTIAILRGLKE
S.aureus 295 GNMLKPMLARGELHCIGATTLNEYREYIEKDSALERRFQKVAVSEPDVEDTISILRGLKE
A.phagocytophil 359 KYELHHGIRITDSAIVAAANLSSRYIPDRFLPDKAIDLIDEAASRARIEIDSKPEIIDSI
E.chaffeensis 358 KYEVHHGIRIMDSAIIAASTLSNRYITDRFLPDKAIDLIDEAASRVRIEIDSKPEVIDEL
E.coli 356 RYELHHHVQITDPAIVAAATLSHRYIADRQLPDKAIDLIDEAASSIRMQIDSKPEELDRL
S.aureus 355 RYEVYHGVRIQDRAIVAAAELSDRYITDRFLPDKAIDLVDQACATIRTEMGSNPTELDQV


A.phagocytophil 419 DRQVMQLKIESEALKNENTEASKQRLEEISRELQSLSSEAADLNSQWHAEKAKISKMHEL
E.chaffeensis 418 DRKIIQLKIEAGVLEKENTESSKQRLAQLSEELNKLSIQATELNSKWQAEKMKILKMQEC
E.coli 416 DRRIIQLKLEQQALMKESDEASKKRLDMLNEELSDKERQYSELEEEWKAEKASLSGTQTI
S.aureus 415 NRRVMQLEIEESALKNESDNASKQRLQELQEELANEKEKQAALQSRVESEKEKIANLQEK
A.phagocytophil 479 TESLLDSARIELEQSQRLGNLSRAGELMYGIIPSLEAELKKHEE--------IAGTLLRKE
E.chaffeensis 478 VEKLDNARNDLEKAQRSGNLAKAGELMYGIIPELEKELKKCEK--------PSSNMLKRE
E.coli 476 KAELEQAKIAIEQARRVGDLARMSELQYGKIPELEKQLEAATQL-----EGKTMRLLRNK
S.aureus 475 RAQLLESRQALEDAQTNNNIEKAAELQYGTIPQLEKELRELEDNFQDEQGEDTDRMIREV
A.phagocytophil 531 IKANDIAAIVERWTGIPVDSVMNSEKEKLLHMEEELKKTVIGQDSAVAAVSNAVRRSRAG
E.chaffeensis 530 VTESDIASIVSRWTGIPIENMMSSEKEKLLRMEEEIGKTVIGQESAIKAVSDAVRRSRAG
E.coli 531 VTDAEIAEVLARWTGIPVSRMMESEREKLLRMEQELHHRVIGQNEAVDAVSNAIRRSRAG
S.aureus 535 VTDEEIGDIVSQWTGIPVSKLVETEREKLLHLSDILHKRVVGQDKAVDLVSDAVVRARAG
A.phagocytophil 591 VQDAQRPMGSFLFLGPTGVGKTELTKALSKFLFDSSSALLRFDMSEFMEKHSVAKLIGAP
E.chaffeensis 590 VQDANKPLGSFLFLGPTGVGKTELVKTLAEFLFCDKSALLRFDMSEFMEKHAVSRLIGAP
E.coli 591 IADENRPIGSFLFLGPTGVGKTELCKALANFMFDSDEAMVRIDMSEFMEKHSVSRLVGAP
S.aureus 595 IKDENRPIGSFLFLGPTGVGKTELAKSLAASLFDSEKHMIRIDMSEYMEKHAVSRLIGAP
A.phagocytophil 651 PGYVGYEQGGLLTEAVRRRPYQVILFDEIEKAHADIFNLLLQVLDEGRLTDSRGNLVNFK
E.chaffeensis 650 PGYVGYDQGGMLTESVRRRPYQVILFDEIEKAHGDIFNILLQVLDEGRLTDNHGKLVDFR
E.coli 651 PGYVGYEEGGYLTEAVRRRPYSVILLDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFR
S.aureus 655 PGYIGHDEGGQLTEAVRRNPYSVILLDEVEKAHTDVFNVLLQILDEGRLTDSKGRSVDFK
A.phagocytophil 711 NTILVITSNIGQDILINSTEDSND--PVVRKTVLEMLRLSFRPEFLNRLDEIMIFNRLTQ
E.chaffeensis 710 NTILVLTSNLGQEILINNKEDVDG--ESVKKSITSVLQHHFRPEFLNRLDEIIVFHRLTK
E.coli 711 NTVVIMTSNLGSDLIQERFGELDY--AHMKELVLGVVSHNFRPEFINRIDEVVVFHPLGE
S.aureus 715 NTIIIMTSNIGSQVLLENVKETGEITESTEKAVMTNLNAYFKPEILNRMDDIVLFKPLSI
A.phagocytophil 769 EHIEHIVDVQISNLQKILSDKGITISLHQGAKSWLVKHGYDVACGARLLKRLIQQHIQNQ
E.chaffeensis 768 EHIEKIIDVQFSLLQKIVAQKELEISLSSEAKSWLMNNGYDSLYGARPLKRLIQQKIQNS
E.coli 769 QHIASIAQIQLKRLYKRLEERGYEIHISDEALKLLSENGYDPVYGARPLKRAIQQQIENP
S.aureus 775 DDVSMIVDKILTQLNIRLLEQRISIEVSDDAKAWLGQEAYEPQYGARPLKRFVQRQIETP
A.phagocytophil 829 LACLLLGDKITEGSKLVVFEENNSLVIKE-A---T
E.chaffeensis 828 LAKLILANQVSKGDKLEVVLINDDLIINK-----L
E.coli 829 LAQQILSGELVPGKVIRIEVNEDRIVAV------Q
S.aureus 835 LARMMIKEGFPEGTTIKVNLNSDNNLTFNVEKIHE

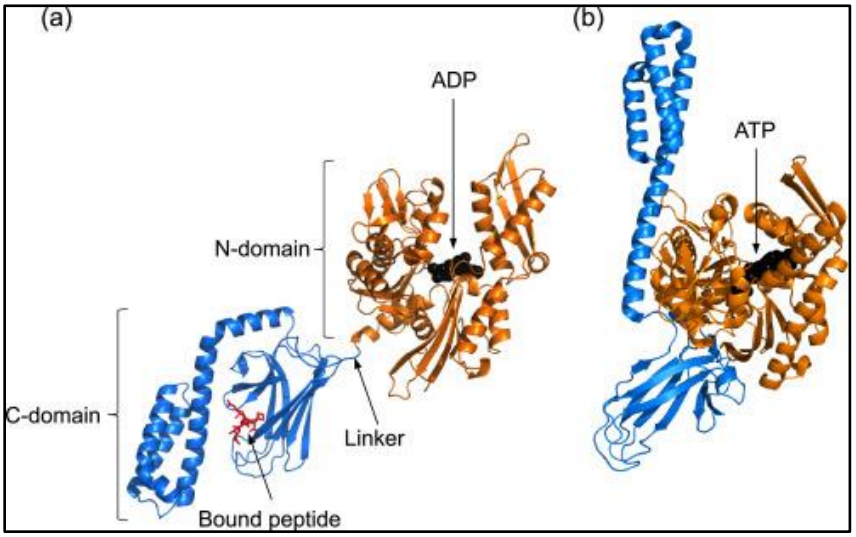
B.8 Amino acid sequence of ClpB monomer from *Escherichia coli.*

E. coli ClpB sequence from UniProtKB - P63284 (CLPB_ECOLI) from *Escherichia coli* (strain K12). Residues which are part of Walker A of first NBD are highlighted in pink color box, Walker B of first NBD are highlighted in green color box, Walker A of second NBD are highlighted in red color box and Walker B of second NBD are highlighted in yellow color box. Pore loops residues of first NBD are highlighted in blue colored box and pore loops residues of second NBD are highlighted in magenta color box. Residues of Sensor 1 are underlined, and Sensor 2 residues are shown in italics and underlined. Arginine finger residue are shown in brown color and underlined.

>sp|P63284|CLPB_ECOLI Chaperone protein ClpB
OS=Escherichia coli (strain K12) GN=clpB PE=1 SV=1
MRLDRLTNKFQLALADAQSLALGHDNQFIEPLHLMSALLNQEGGSVSPLLTSAGINAGQL
RTDINQALNRLPQVEGTGGDVQPSQDLVRVLNLCDKLAQKRGDNFISSELFVLAALESRG
TLADILKAAGATTANITQAIEQMRGGESVNDQGAEDQRQALKKYTIDLTERAEQGKLDPV
IGRDEEIRRTIQVLQRRTKNNPVLIGEPGVGKTAIVEGLAQRIINGEVPEGLKGRRVLAL
DMGALVAGAKYRGEFEERLKGVLNDLAKQEGNVILFIDELHTMVGAGKADGAMDAGNMLK
PALARGELHCVGATTLDEYRQYIEKDAALERRFQKVFVAEPSVEDTIAILRGLKERYELH
HHVQITDPAIVAAATLSHRYIADRQLPDKAIDLIDEAASSIRMQIDSKPEELDRLDRRII
QLKLEQQALMKESDEASKKRLDMLNEELSDKERQYSELEEEWKAEKASLSGTQTIKAELE
QAKIAIEQARRVGDLARMSELQYGKIPELEKQLEAATQLEGKTMRLLRNKVTDAEIAEVL
ARWTGIPVSRMMESEREKLLRMEQELHHRVIGQNEAVDAVSNAIRRSRAGLADPNRPIGS
FLFLGPTGVGKTELCKALANFMFDSDEAMVRIDMSEFMEKHSVSRLVGAPPGYVGYEEGG
YLTEAVRRRPYSVILLDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFRNTVVIMTSNL
GSDLIQERFGELDYAHMKELVLGVVSHNFRPEFINRIDEVVVFHPLGEQHIASIAQIQLK
RLYKRLEERGYEIHISDEALKLLSEN*GYDPVYGAR*PLKRAIQQQIENPLAQQILSGELVP
GKVIRLEVNEDRIVAVQ

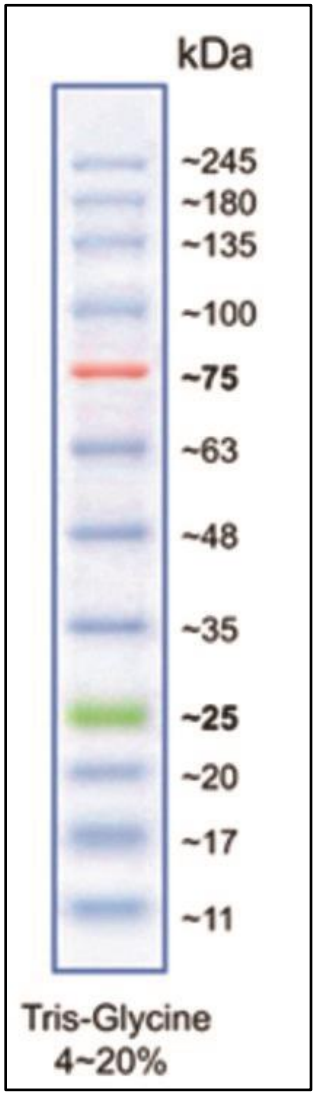
B.9 *E. coli* DnaK structure.

E. coli DnaK has two main domains. One is known as N-terminal nucleotide binding domain (NBD, 44 kDa, shown in gold color) that contains ATPase activity. C-terminal domain is known as substrate binding domain (SBD, 25kDa, shown in blue color). These two domains are connected by a short flexible interdomain linker (approximately 12 residues in length). N-terminal adenosine triphosphate (ATP) binding allosterically regulates substrate binding in the C terminal SBD. As ATP is hydrolyzed, in adenosine diphosphate (ADP) bound DnaK, alpha helical lid adheres strongly (closed lid state, a) to SBD. Substrate affinity of SBD is high in ADP bound DnaK. When ADP is replaced with ATP in the next round by a nucleotide exchange factor , the lid opens (opened lid state, b)and release the substrate. Substrate affinity is lower in ATP bound DnaK. Original figure is from (Wickner, Camberg et al. 2017). Printed with permission.



B.10 BLUeye Prestained Protein Ladder.

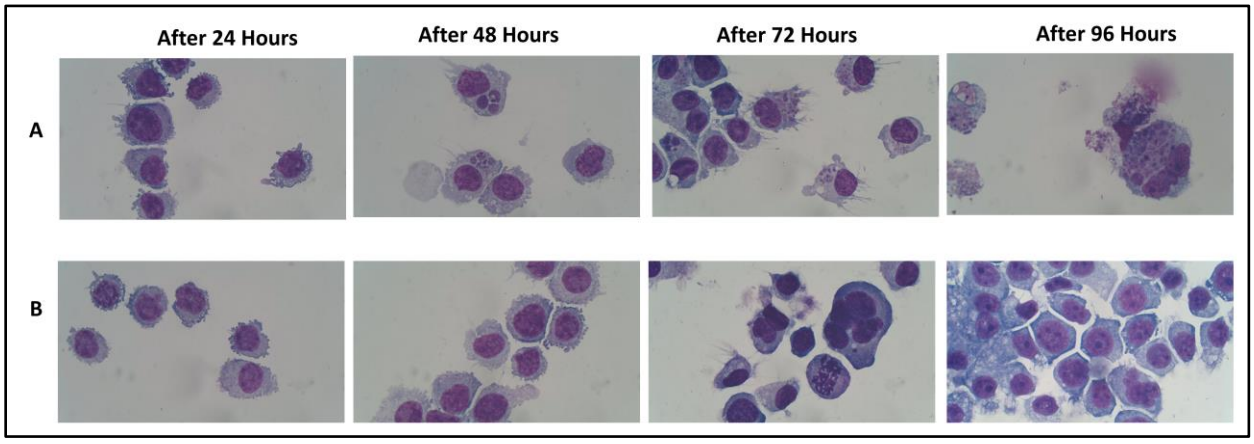
BLUeye Prestained Protein Ladder (94964, Sigma-Aldrich) is shown here with its corresponding molecular weight marker positions in Tris- Glycine 4-20 % SDS-PAGE gel. This is the markers used for all the gels and western blots presented in this thesis. Picture adapted from https://www.sigmaaldrich.com/catalog/product/sigma/94964?lang=en&region=US, accessed on 03/20/2021.



Appendix C - Chapter 5 Supplementary Data.

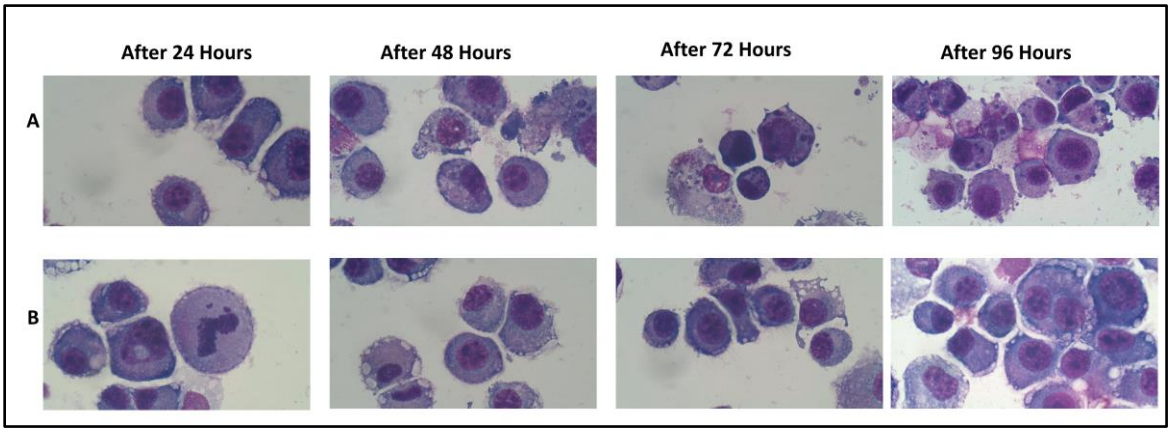
C.1. Inhibition of *Anaplasma phagocytophilum* infection propagation with 0.5 µM DBeQ.

HL60 cells were infected with *A. phagocytophilum* in T25 flasks and 0.5 µM DBeQ was added to one flask and DMSO was added to the other flask immediately . Infection propagation was monitored every day by hematoxylin staining. Panel A shows infected HL60 cells with DMSO and panel B shows infected HL60 cells with 0.5 µM DBeQ at different time points. There was no infection propagation in the flask treated with 0.5 µM DBeQ. Clearly infected and lysed cells were seen in DMSO treated flask after 96 hours. Image was captured and generated using ACCU-SCOPE Excelis camera and CaptaVision+™ Software. Image magnification of 1000.



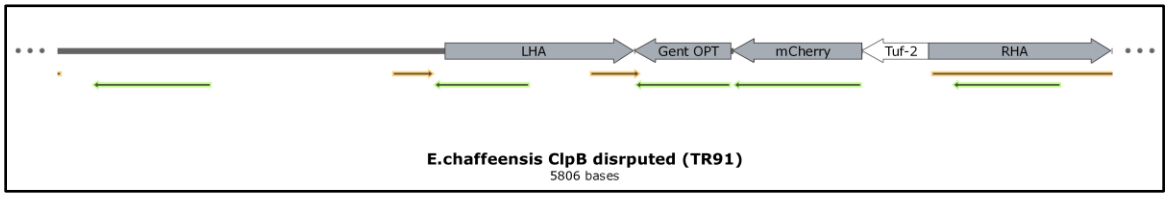
C.2. Inhibition of *Ehrlichia chaffeensis* infection propagation with 0.1 µM DBeQ.

DH82 cells were infected with *Ehrlichia chaffeensis* in T25 flasks and 0.1 µM DBeQ was added to one flask and DMSO was added to the other flask immediately . Infection propagation was monitored every day by hematoxylin staining. Panel A shows infected DH82 cells with DMSO and panel B shows infected DH82 cells with 0.1 µM DBeQ at different time points. There was no infection propagation in the flask treated with 0.1 µM DBeQ. Clearly infected and lysed cells were seen in DMSO treated flask after 96 hours. Image was captured and generated using ACCU-SCOPE Excelis camera and CaptaVision+™ Software. Image magnification of 1000.



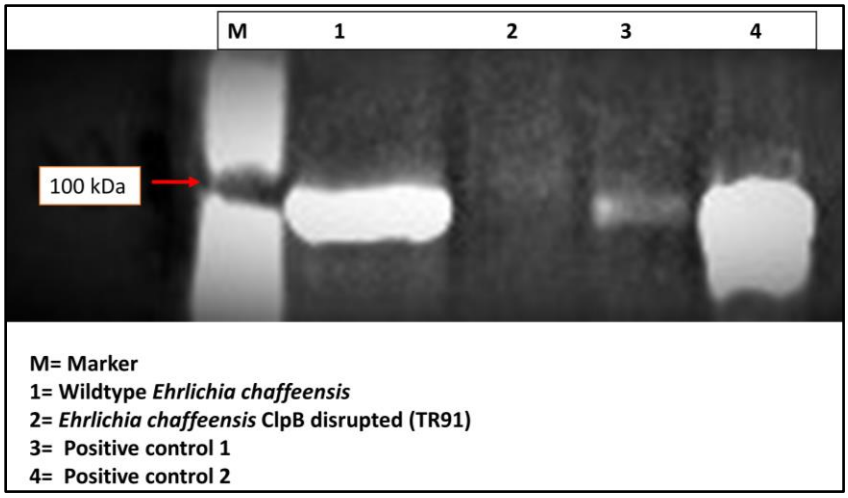
C.3. ClpB gene disruption in *Ehrlichia chaffeensis* TR91.

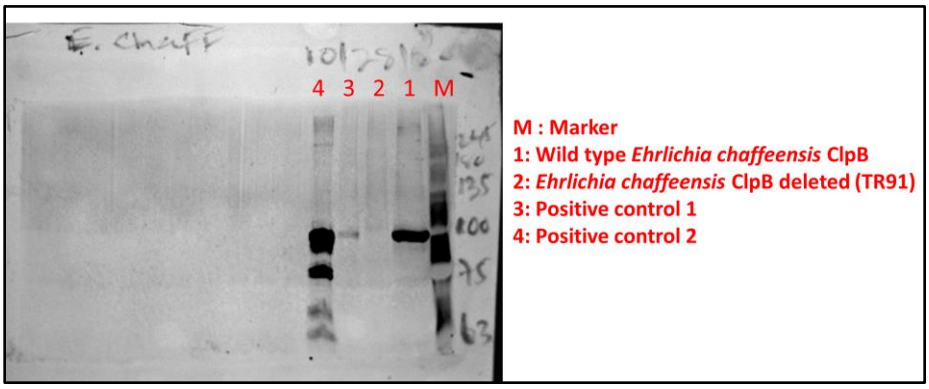
Ehrlichia chaffeensis ClpB deleted strain was generated at Dr Roman Ganta's lab. This strain was generated by inserting a mCherry and gentamicin cassettes within ClpB coding sequence by homologous recombination. In TR91 strain, ClpB gene is disrupted and hence no ClpB protein is being produced. mCherry and gentamicin and expression is controlled by Tuf promoter. RHA stands for right homolog arm of ClpB gene. LHA stands for left homolog arm of ClpB gene.



C.4. ClpB gene disruption conformation in *Ehrlichia chaffeensis* TR91.

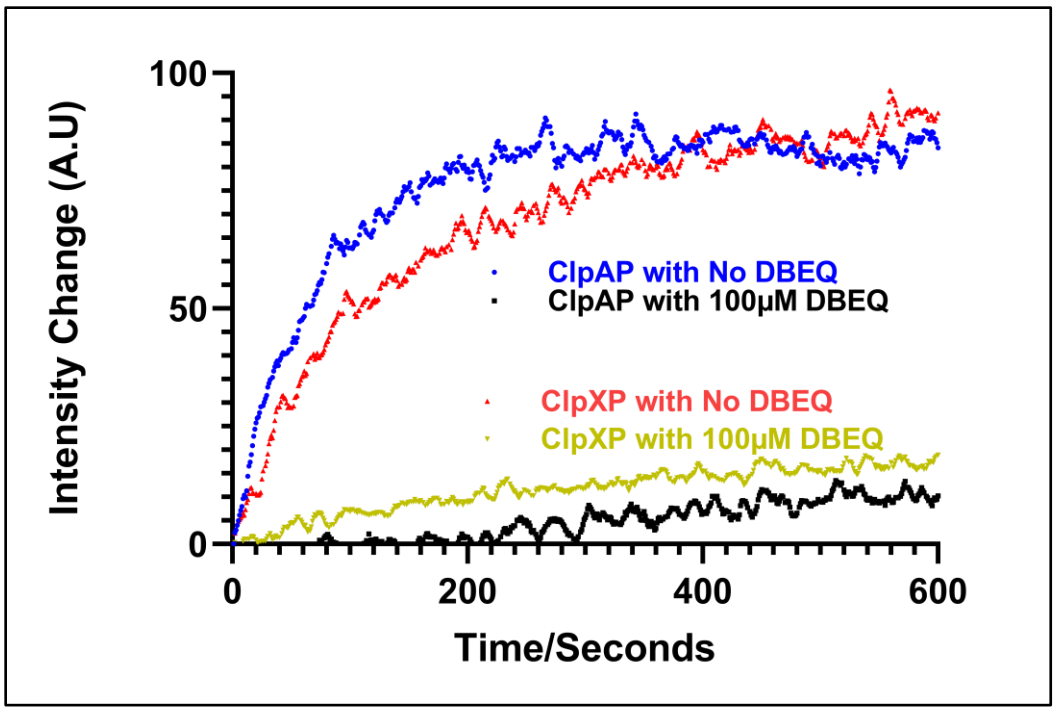
Western blot was performed as explained in materials and methods. Immunodetection was done using rabbit polyclonal anti-ClpB IgG and anti-rabbit horseradish peroxidase conjugated secondary antibodies. Signal detection was performed using SuperSignal West Pico Chemiluminescent Substrate (Pierce) and image acquiring was done using an Invitrogen iBright FL1500 Imaging System (ThermoFisher Scientific, USA). Location of 100 kDa band is shown in the figure. Positive control 1 was purified recombinant *Ehrlichia chaffeensis* ClpB and positive control 2 was purified recombinant *E.coli* ClpB. There was no signal recorded for TR91 sample and confirmed that there was no ClpB protein found in TR91 sample. Original blot is also shown below.





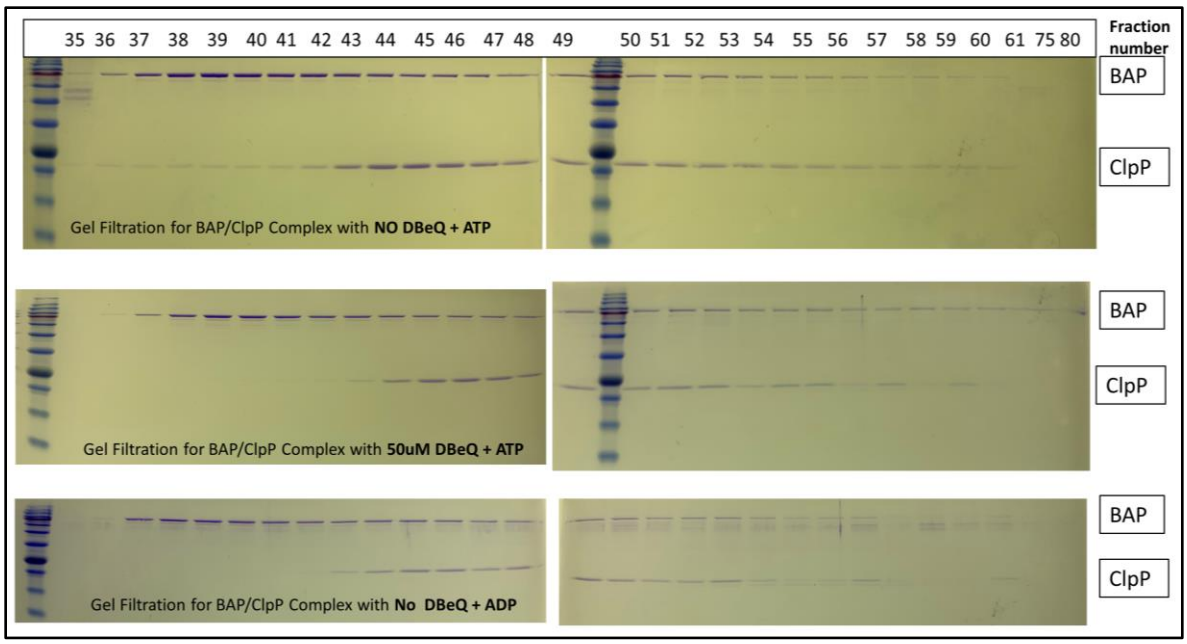
Appendix D - DBeQ destabilizes the ClpAP and ClpXP complexes.

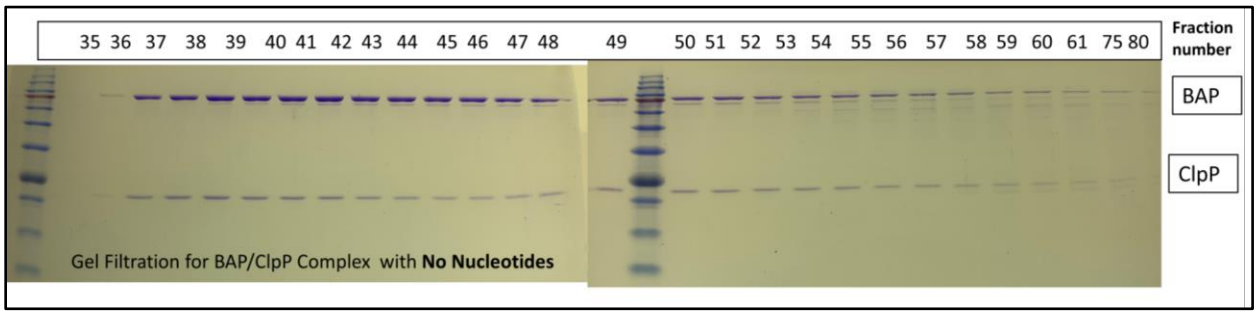
D.1. The appendices include additional results of the experiments performed to explain the effects of DBeQ on ClpB, which were not included in the publication described in Chapter 3,4 and 5.We investigated the effect of DBeQ on the assembly of ClpAP and ClpXP protease complexes. ClpA and ClpX are AAA+ proteins similar to ClpB, while ClpP is a serine protease. A degradation substrate is translocated through the central channel in ClpA or ClpX and is handed over to ClpP for degradation. To display a protein degradation activity, ClpA and ClpX need to interact with ClpP and form biologically active ClpAP and ClpXP complexes. We discovered that DBeQ inhibits formation of ClpAP and ClpXP complexes *in vitro*. We monitored degradation of FITC-labelled casein over time with or without 100 µM DBeQ. Experimental conditions: 20 µl of ATP from 100 mM stock, 10 µl of wild type ClpX from 13.3 mg/ml stock, 10 µl of wild type ClpP from 1.5 mg/ml stock, 10 µl of FITC casein from 130 µM stock, excitation wavelength at 498 nm, emission wavelength at 521 nm. The experiments were performed at room temperature in buffer C (50-mM Tris-HCl, pH 7.4, 20-mM $MgCl_2$, 1-mM EDTA, 0.5-mM tris(2-carboxyethyl)phosphine [TCEP]). Fluorescence intensity was measured with a PerkinElmer LS55 spectrometer. Emission increase corresponds to degradation of FITC-casein.

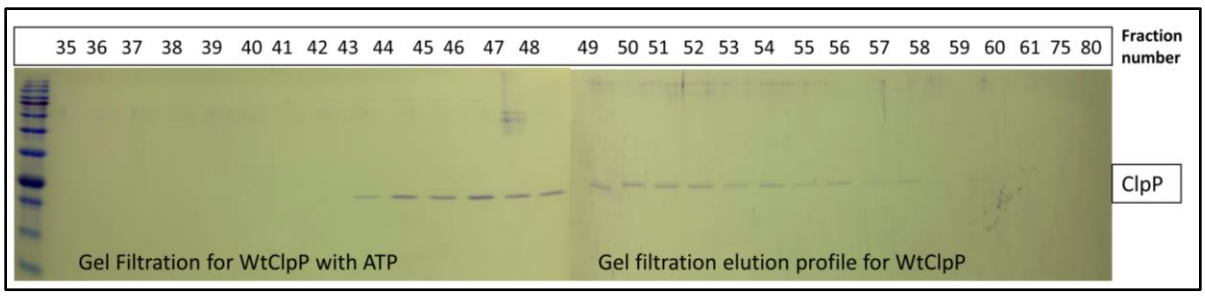

Intensity Change (A.U)
100
50
0
0
200
400
600
Time/Seconds
ClpAP with No DBEQ
ClpAP with 100μM DBEQ
ClpXP with No DBEQ
ClpXP with 100μM DBEQ

Appendix E - DBeQ destabilizes the BAP/ClpP complex.

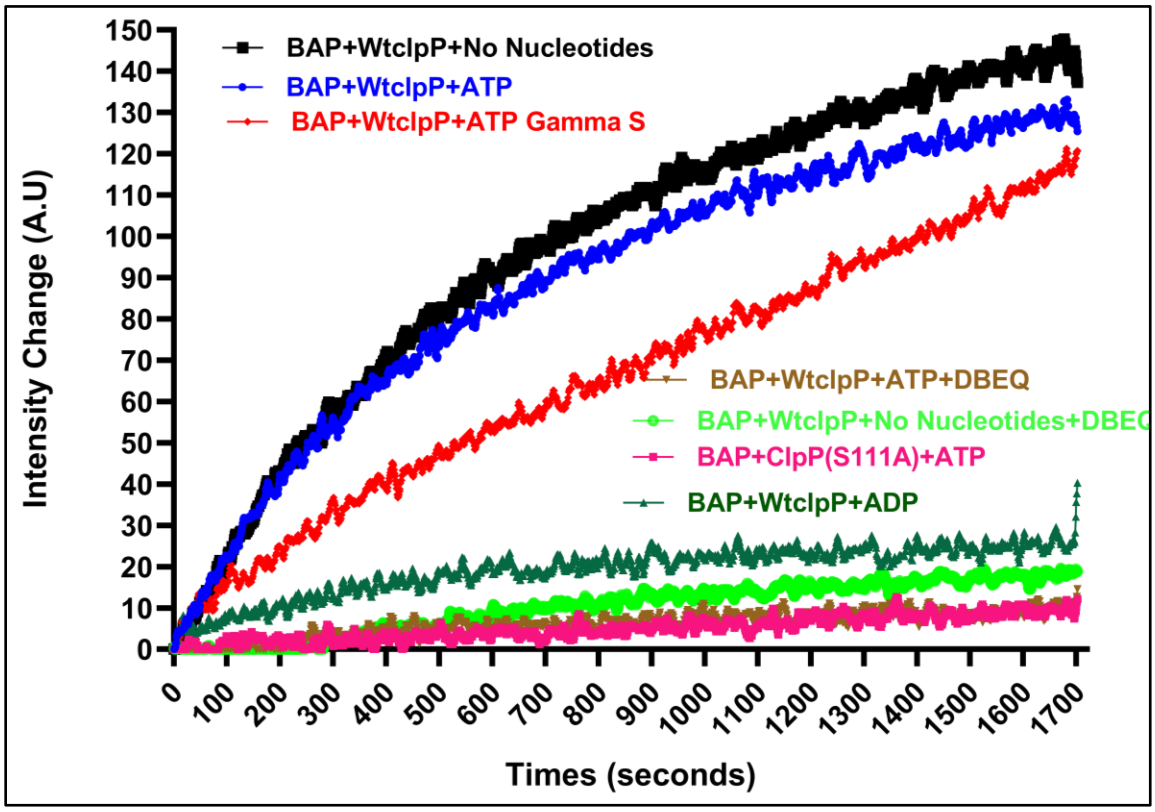
E.1 We used a ClpB variant known as BAP (a ClpB variant that, like ClpA associates with ClpP). BAP is an engineered protein and not found in nature (chapter 2 reference to (Weibezahn, Tessarz et al. 2004)). ClpB does not interact with ClpP because it lacks the essential structural fragment found in ClpA and ClpX. Like native ClpB, BAP translocates substrates through its central pore, but unlike in ClpB, the substrates of BAP directly enter ClpP and are subject of degradation. We monitored degradation of FITC-labelled casein over time with or without 100 µM DBeQ. The FITC degradation assay was performed under the same conditions as in appendix A. Gel filtration was performed with purified 440 µg of BAP with 195 µg of wildtype ClpP. The column was washed with 5mM ATP and the running buffer included 5 mM ATP or ADP. The flow rate was 0.3 ml/min. BAP (95 kDa) and ClpP (23 kDa) eluted together without DBeQ in the presence of ATP or without nucleotides (fractions 36-42), but their complex was not observed in presence of 50 µM DBeQ. The BlueEye prestained Protein Marker (refer to Appendix B.10) is also shown here.


35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 75 80
Fraction number
BAP
ClpP
Gel Filtration for BAP/ClpP Complex with NO DBeQ + ATP
BAP
ClpP
Gel Filtration for BAP/ClpP Complex with 50uM DBeQ + ATP
BAP
ClpP
Gel Filtration for BAP/ClpP Complex with No DBeQ + ADP

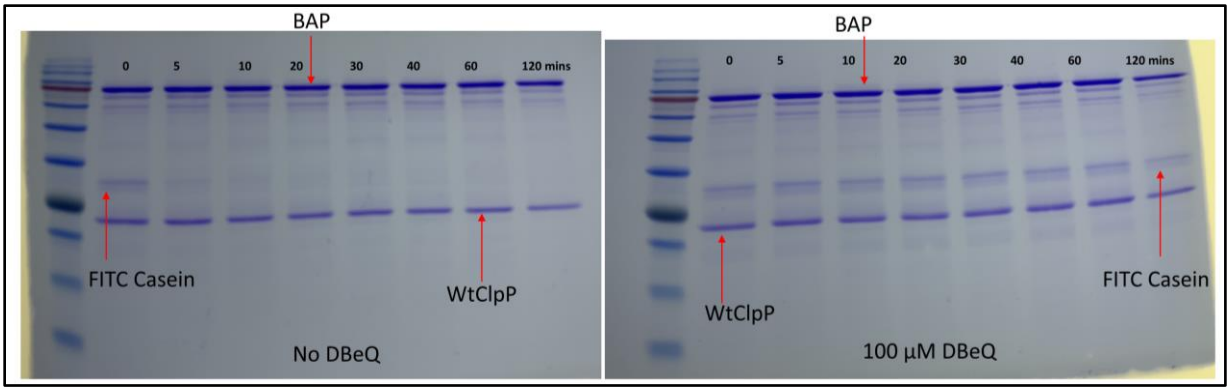

35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 75 80
Fraction number
BAP
ClpP
Gel Filtration for BAP/ClpP Complex with No Nucleotides


35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 75 80
Fraction number
ClpP
Gel Filtration for WtClpP with ATP
Gel filtration elution profile for WtClpP

E.2. FITC-labeled casein degradation by BAP/ClpP complex is shown below. 1.56 µM FITC casein was incubated with 2µL BAP from 22mg/ml stock and 2µL ClpP from 1.3mg/ml stock. ClpPS111A is an inactive version of ClpP. An increase in emission intensity indicates degradation of casein.

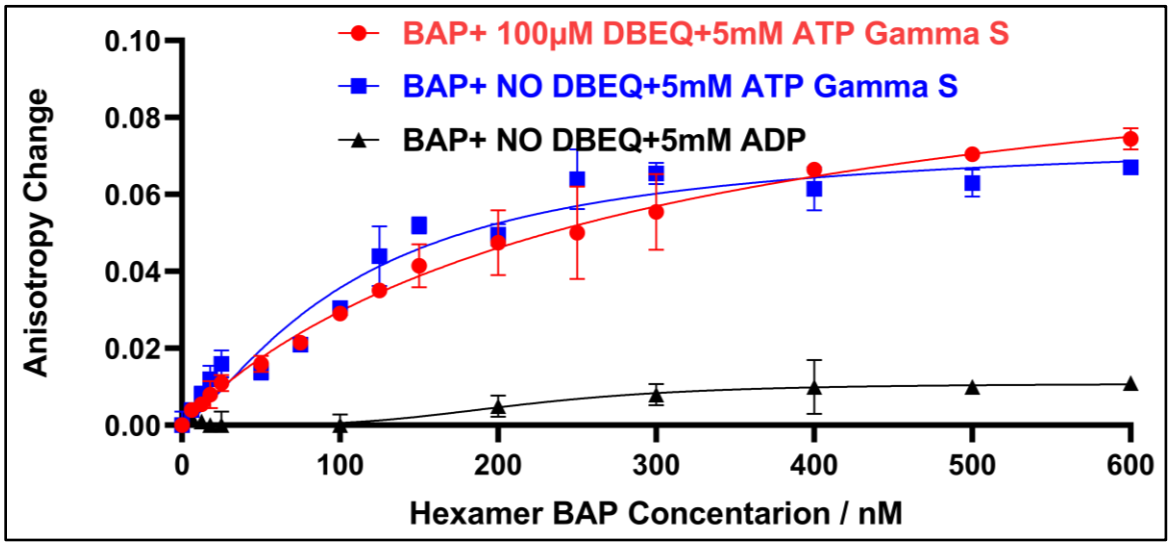


E.3. The FITC-labeled casein degradation by BAP/ClpP is also shown below with SDS-PAGE. 2 µM FITC casein, 76 µM BAP and 48 µM wildtype ClpP and 5 mM ATP were used. The degradation of casein is inhibited by 100 µM DBeQ.



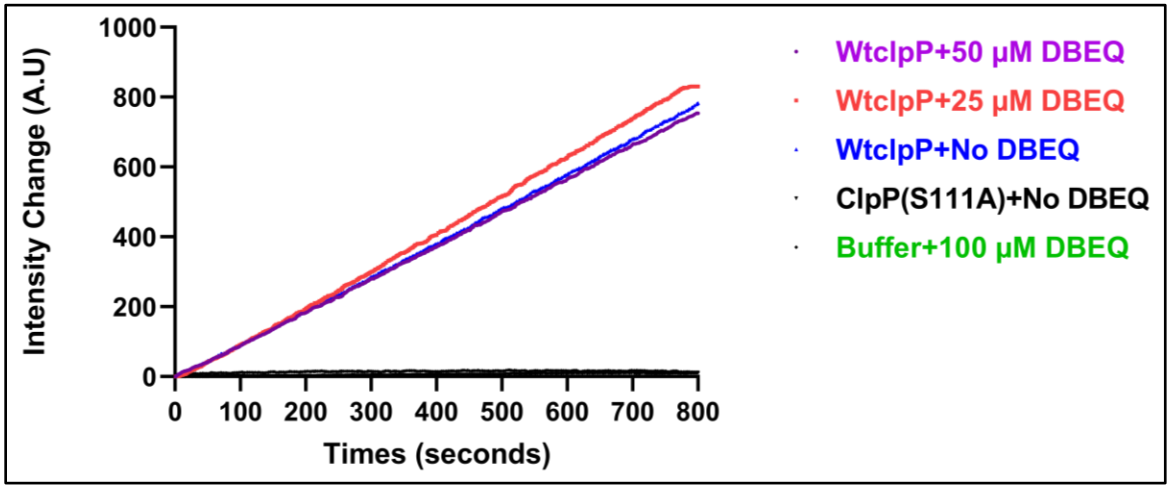
Appendix F - DBeQ does not inhibit ATP-dependent binding of casein to BAP.

F.1. The titration of FITC-casein with BAP was done with or without DBeQ and the fluorescence anisotropy measurements were performed as previously described (see Chapter 2). We used 100 nM FITC-casein in duplicate experiments. The average anisotropy values are plotted in the graph below with standard deviations. A specific binding model with Hill number was used to fit the data. The binding isotherms gave the following values in the presence of ATP-gamma-S and DBeQ: Kd, 299.1 nM; saturation signal, 0.1137; Hill number, 0.9537; r2, 0.9770 and for ATP-gamma-S without DBeQ: Kd, 108.1 nM; saturation signal, 0.07536; Hill number, 1.352; r2, 0.9477. Thus, DBeQ did not inhibit binding of FITC-casein binding to BAP.



Appendix G - DBeQ does not inhibit ClpP protease activity.

G.1. Purified wild type ClpP was used with or without DBeQ in presence of L-Leucine-7-amido-4-methylcoumarin hydrochloride (AMC substrate purchased from Sigma-Aldrich). During degradation of this substrate, its fluorescence increases. The excitation wavelength of 380 nm and the emission wavelength of 455 nm were used. The experiment was performed at room temperature in buffer Z (25 mM Tris.HCl , 10 mM $Mgcl_2.6H_2O$, 2 mM DTT, 75 mM KCl). ClpP S111A is an inactive version of ClpP. Experimental conditions: 10 µl from 10 mM AMC substrate stock, 1.5 µl of ClpP from 1.3 mg/ml stock and the total reaction volume was 400 µl . The results show that DBeQ does not inhibit the protein degradation activity of ClpP.



Appendix H - Neither BAP nor ClpP shows a standalone protein degradation activity.

H.1. The FITC-casein degradation experiment was conducted in buffer Z (25 mM Tris.HCl , 10 mM $Mgcl_2.6H_2O$, 2 mM DTT, 75 mM KCl). 4 µl of FITC casein from 156 uM stock was used. The total reaction volume was 400 µl. 2 µl BAP from 22 mg/ml stock and 2 µl wild type ClpP was used from 1.3 mg/ml stock. The experiments were performed in the presence of 5 mM ATP at room temperature. No degradation of FITC-casein was observed up to 90 min of the reaction, which confirmed that the BAP/ClpP complex formation is essential for FITC-casein degradation. The BlueEye prestained Protein Marker (refer to Appendix B.10) is also shown.