Investigation of ClpXP protease mechanism of function and its interaction with the folding chaperone trigger factor

by

Angela Yeou Hsiung Yu

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Department of Biochemistry

University of Toronto

© Copyright by Angela Yeou Hsiung Yu 2013
Investigation of ClpXP protease mechanism of function and its interaction with the folding chaperone trigger factor

Angela Yeou Hsiung Yu
Doctor of Philosophy
Department of Biochemistry
University of Toronto
2013

Abstract

The major chaperones identified in *Escherichia coli* that assist in protein folding include trigger factor (TF), DnaK/DnaJ/GrpE and GroEL/GroES systems. The main ATP-dependent proteases are ClpXP, ClpAP, HslUV, Lon, and FtsH. From detailed sequence analysis, we found that *tig* (gene for TF), *clpX*, and *clpP* genes co-localize next to each other in most examined bacteria. We hypothesized that TF and ClpXP are functionally associated. TF is a ribosome-associated folding chaperone whereas ClpXP is a degradation complex. ClpX serves as the regulatory ATPase that recognizes substrates, unfolds and translocates polypeptides into ClpP for degradation. I found that TF physically interacts with ClpX, and that they collaborate to enhance degradation of certain ClpXP substrates. It is estimated that TF enhances the degradation of about 2% of newly synthesized *E. coli* proteins. One of the ClpXP substrates with degradation enhanced by TF was λO, the λ phage replication protein. Furthermore, TF also enhanced the degradation of ribosome-stalled λO nascent chains. Experiments suggest that TF transfers ribosome-stalled λO to ClpX for degradation by ClpP, demonstrating the existence of co-translational protein degradation in *E. coli*.

To understand ClpXP mechanism, we had previously proposed that the degraded peptides are released from ClpP through transient equatorial side pores. To further understand ClpP dynamics, we determined the structure of ClpP(Ala153Cys) in its oxidized state. The structure shows that each opposing pair of protomers is linked by a disulfide bond. Unexpectedly, this structure resembles the compact structures of *Streptococcus pneumoniae*, *Mycobacterium tuberculosis*, and *Plasmodium falciparum* ClpPs, rather than the extended states seen in previous *E. coli* ClpP structures. Normal mode analysis of ClpP structures suggested that the
compact structure is a naturally sampled conformation of WT ClpP. My findings provide insights for understanding ClpP dynamics as well as reveal a novel association between ClpXP protease and TF folding chaperone.

(300 words)
Acknowledgments

I am truly appreciative for all the help and support I received during my graduate training. There are so many people who helped me along the way and it is impossible to name them all, but my appreciation is sincere to all. First, I would like to express my sincere gratitude to my supervisor Dr. Walid Houry. I feel very fortunate to have conducted research in his lab. He has demonstrated for me how to think like a scientist and do good research. I have learned tremendously from him. I thank him for all his guidance and support for all these years. I hope that one day I will be able to think like him and to generate very novel ideas myself too.

My committee members Dr. John Glover and Dr. Avi Chakrabarty have been extremely helpful for giving me directions in my research. They are very supportive and encouraging to me. I thank them for helping me along the way.

I would like to thank Dr. James Rini, the graduate coordinator at the time when I first joined the Biochemistry Department. I was transferring from another department to Biochemistry, and he provided lots of advice and helped me transition into the Biochemistry Department and Dr. Houry’s lab. I thank him for helping a lost student get back on track.

My undergraduate professor at Queen’s University Dr. Michael Neshem has inspired me greatly both in science and in life. He taught me the first Biophysics course and mentored me in his lab. I thank him for being such an inspiration and opened the door of research for me.

I would like to thank everyone in the Houry lab, past and present, short term and long term. Thank you for making the lab such a friendly and stimulating environment. Guillaume was guiding when I first entered the lab, and how he strived for excellence often reminds me that I should try harder. I thank Guillaume, Deji, Sarah, Elisa, and my collaborators Dr. Elisabeth Tillier, Dr. Matt Kimber, and Dr. Mikael Borg for their help and contributions to my research.

I have developed great friendships with everyone in the lab, especially Jen, Elisa and Usheer. Thank you for listening to me, for giving me advice in science and in life, and for being such nice people! Special thanks to Jen for being so kind and generous to me. I cherish the time we spent together in the lab and in your house.

I thank the researchers whom I have never met in person but had sent me strains,
antibodies, plasmids and information. Because of you, I feel that researchers are not only the most brilliant but also the nicest people on earth and it motivates me to become one.

My soon to be husband Jonathan Lowe has been supportive for the longest time. I thank him for driving me to the lab on numerous weekends and waiting in the car while I tried to finish a “fast experiment” which always turned out to be much longer. I am truly grateful for his patience and sorry for making him wait every time. Last and most importantly, I want to thank my parents. They are always there for me and give me all the support I ever need and even more. Because of them, I feel brave enough to make my own choices. I want to thank them for being such great parents. Their support and encouragements are the most important fuels that keep me going and accomplish anything.
Table of Contents

Contents

Acknowledgments ................................................................................................................ iv
Table of Contents ................................................................................................................ vi
List of Tables ......................................................................................................................... viii
List of Figures ....................................................................................................................... ix
1 Introduction of protein folding and degradation ................................................................. 1
   1.1 Summary ....................................................................................................................... 2
   1.2 Molecular chaperones assist protein folding in vivo ....................................................... 3
   1.3 TF assists in co-translational protein folding in E. coli ............................................... 5
      1.3.1 TF domain arrangement and structure ................................................................. 5
      1.3.2 Mechanism of functions of TF .............................................................................. 8
   1.4 Co-translational degradation ....................................................................................... 9
   1.5 Clp ATPases and their role in protein unfolding and degradation in E. coli ............. 11
      1.5.1 ClpXP degradation system .................................................................................. 13
      1.5.2 ClpX AAA+ ATPase ............................................................................................ 15
      1.5.3 ClpP: A Distinctive Family of Cylindrical Energy-Dependent Serine Proteases. 25
   1.6 Thesis rationale ........................................................................................................... 45
2 Functional cooperativity between trigger factor and ClpXP in degradation ................. 47
   2.1 Summary ....................................................................................................................... 48
   2.2 Introduction .................................................................................................................. 49
   2.3 Experimental Procedures ............................................................................................ 52
   2.4 Results ......................................................................................................................... 58
   2.5 Discussion ................................................................................................................... 82
3 Structural and Theoretical Studies Indicate that the Cylindrical Protease ClpP Samples
Extended and Compact Conformations .................................................................................. 84

3.1 Summary .......................................................................................................................... 85
3.2 Introduction ..................................................................................................................... 86
3.3 Experimental Procedures .............................................................................................. 91
3.4 Results and Discussion ................................................................................................. 95

4 Conclusions and future directions ..................................................................................... 121

4.1 Conclusions .................................................................................................................. 122
4.2 Future directions .......................................................................................................... 123

References .......................................................................................................................... 129

Appendices .......................................................................................................................... 147
List of Tables

Table 1. The distribution of ClpP and ClpR across different organisms .................................. 27

Table 2. Results of the sedimentation velocity analytical ultracentrifugation experiments. ....... 98

Table 3. Peptidase activity of WT ClpP and Clp(A153C).......................................................... 100

Table 4. Data collection and structure refinement statistics. ..................................................... 104

Table 5. Disordered residues in the ClpP\textsuperscript{SS} structure........................................... 105
List of Figures

Figure 1. A simplified schematic illustrating \textit{de novo} protein synthesis in \textit{E. coli}......................... 4

Figure 2. TF domain arrangement and structure................................................................. 6

Figure 3. ClpXP degradation cycle...................................................................................... 14

Figure 4. Structure of ZBD and ClpX.................................................................................. 18

Figure 5. The structures of tetradecameric ClpP. ................................................................. 30

Figure 6. The structures of the ClpP subunit. ....................................................................... 33

Figure 7. Alignment of ClpP and ClpR sequences. .............................................................. 35

Figure 8. Model of ClpP mechanism of function. ................................................................. 38

Figure 9. The genomic positions of \textit{tig}, \textit{clpP}, \textit{clpX} genes across all bacterial protein clusters. ... 60

Figure 10. The effect of chaperones on the ClpXP-mediated degradation of model substrates... 64

Figure 11. Effect of TF on ClpXP-mediated degradation...................................................... 66

Figure 12. SPR analysis of the binding between TF, ClpX and substrates. ......................... 70

Figure 13. Binding of TF domains to ClpX and \textit{\lambda}O. ...................................................... 72

Figure 14. The effect of TF on the ATPase activity of ClpX .................................................. 74

Figure 15. Effect of TF domains on ClpXP-mediated \textit{\lambda}O degradation. ......................... 75

Figure 16. The role of TF in the cellular degradation of \textit{\lambda}O............................................. 79

Figure 17. Degradation of ribosome-stalled \textit{\lambda}O-SecM \textit{in vivo} ........................................... 81

Figure 18. Graphical representation of the dynamics of ClpP cylinder............................... 85

Figure 19. The sequence and secondary structure of \textit{E. coli} ClpP ................................. 87
Figure 20. The extended and compact structures of ClpP. ......................................................... 88

Figure 21. Biophysical characterization of ClpP(A153C).............................................................. 97

Figure 22. Binding of ClpP(A153C) to the AAA+ Chaperones.................................................... 102

Figure 23. The overall structure of ClpP$^{SS}$ and comparison with WT ClpP.............................. 108

Figure 24. Overlay of E. coli ClpP structures.............................................................................. 109

Figure 25. Overlay of the catalytic sites in ClpP......................................................................... 110

Figure 26. ENM normal mode analysis of ClpP......................................................................... 115

Figure 27. TF structure showing potential binding sites to ClpX............................................... 125
1 Introduction of protein folding and degradation

1.1 Summary

Proteins are folded into their precise three-dimensional conformations to fulfill their biological functions. In the classic protein folding experiment, Anfinsen demonstrated that proteins can fold spontaneously in a test tube (Anfinsen 1973) and that the primary sequence of the polypeptide determines the three-dimensional structure of the protein. However, protein folding in the cell is often much more complicated. The cellular environment is crowded with protein concentrations of ~300-400 mg/ml (Zimmerman and Trach 1991), which makes spontaneous protein folding frequently error-prone and inefficient. Folding intermediates and misfolded proteins exposing hydrophobic regions to the cytosol are prone to aggregation. Cells maintain protein homeostasis by employing molecular chaperones that assist in the folding of polypeptide chains and by proteases that regulate protein levels through degradation. In *Escherichia coli*, the major chaperones identified that assist in protein folding include trigger factor (TF), the Hsp70/Hsp40 system (DnaK/DnaJ/GrpE), and the Hsp60/Hsp10 system (GroEL/GroES). The major ATP-dependent proteases are ClpXP, ClpAP, HslU, Lon and FtsH (Figure 1) with each protease having discrete substrate specificities (Dougan, Mogk et al. 2002; Hartl and Hayer-Hartl 2002; Gottesman 1996).
1.2 Molecular chaperones assist protein folding \textit{in vivo}

A nascent chain starts to fold into its tertiary structure co-translationally as the polypeptide chain emerges from the ribosomal tunnel. However, for single domain proteins, only when a full domain of a protein is present outside the ribosome can folding occur. \textit{De novo} protein folding is a complicated process since the sequence information is not only incomplete but also continuously changing during translation. Moreover, the fast translation rate in \textit{E. coli} of about 20 amino acids per second compared to about 5 amino acids per second in eukaryotes makes the co-translational folding process in \textit{E. coli} even more challenging (Netzer and Hartl 1997; Agashe, Guha et al. 2004; Chang, Kaiser et al. 2005). The translating nascent chains that expose hydrophobic regions to the cytosol are prone to protein misfolding and aggregation in the crowded cellular environment (Ellis and Minton 2006). Therefore it is extremely important for the cell to prevent protein aggregation by safeguarding the folding process. Molecular folding chaperones are employed by the cell to bind to the aggregation prone regions and folding intermediates that expose hydrophobic regions to prevent misfolding (Hartl and Hayer-Hartl 2009). The folding chaperones form a network in the cytosol functioning at different stages of polypeptide translation and folding. In \textit{E. coli}, when a polypeptide chain emerges from the ribosome, the first chaperone it encounters is the ribosome-associated TF. TF potentially interacts with most nascent chains during ongoing synthesis and assists nascent chains to fold in a co-translational manner. Small proteins fold rapidly upon synthesis without further assistance. Longer chains subsequently interact with DnaK/DnaJ/GrpE system to reach the native state. After translation, folding intermediates that still expose hydrophobic patches can interact with GroEL/GroES system or DnaK/DnaJ/GrpE system to fold into stable conformations (Figure 1) (Hartl and Hayer-Hartl 2009). The ribosome-associated and cytosolic chaperones form a strong network and cooperate in assisting \textit{de novo} protein folding.
Figure 1. A simplified schematic illustrating *de novo* protein synthesis in *E. coli*

Misfolded or unfolded proteins can be refolded by the folding chaperones or degraded by the proteases in the cell.
1.3 TF assists in co-translational protein folding in *E. coli*

1.3.1 TF domain arrangement and structure

TF was first identified in 1987 by William Wickner and his student Elliott Crooke. TF was named based on its speculated function at the time; it was thought to trigger translocation of membrane protein proOmpA (Crooke, Guthrie et al. 1988). Later in 1990, it was found by the same group that the presence of TF is not crucial in assisting membrane protein translocation since the translocation of proOmpA was not abolished in the TF deletion strain (Δtig), though TF did aid long term stability of the protein (Guthrie and Wickner 1990). It is now generally accepted that TF is a key ribosome-associated chaperone that assists in nascent chain folding co-translationally (Deuerling, Patzelt et al. 2003; Kaiser, Chang et al. 2006; Oh, Becker et al. 2011).

In bacteria, TF is the only ribosome-associated chaperone known so far. TF associates with the large 50S ribosomal subunit through interactions with the ribosomal proteins L23 and L29 located at the polypeptide exit tunnel in *E. coli* (Figure 2A-C) (Ferbitz, Maier et al. 2004), thus it is the first chaperone to interact with nascent chains to prevent misfolding during translation (Deuerling, Schulze-Specking et al. 1999; Teter, Houry et al. 1999). TF is found in bacteria and chloroplast. There is no TF homolog in eukaryotes, but functionally similar chaperones that assist in co-translational protein folding like the nascent chain associated complex (NAC), ribosome associated complex (RAC), and ribosome quality control complex (RQC) are present in yeast and higher eukaryotes (Craig, Eisenman et al. 2003; Wegrzyn and Deuerling 2005). TF in prokaryotes, NAC, RAC and RQC in eukaryotes are considered as the molecular guardians that ensure proper folding of nascent chains as they emerge from the ribosome.
Figure 2. TF domain arrangement and structure.

(A) Domain arrangement of TF. (B) Crystal structure of monomeric *E. coli* TF (Protein Data Bank file: 1W26) (Ferbitz, Maier et al. 2004) was drawn using PyMOL (http://pymol.sourceforge.net). (C) Structure of *E. coli* TF bound to *Haloarcula marismortui* 50S ribosomal subunit, adapted from (Ferbitz, Maier et al. 2004). The colouring in both TF structures is the same as the colouring in domain arrangement.
TF in *E. coli* is a 48 kDa protein containing an N-terminal domain, a central peptidyl-prolyl cis/trans isomerase (PPIase) domain, and a C-terminal domain with two arm motifs (Figure 2A) (Patzelt, Rudiger et al. 2001). The N-terminal domain is responsible for binding to the ribosome; it carries a signature motif (GFRxGxxP) which mediates ribosome docking (Ferbitz, Maier et al. 2004). The PPIase domain is responsible for the high peptidyl-prolyl *cis/trans* isomerase activity *in vitro*, but the significance of the PPIase domain is unclear since the PPIase activity of TF is dispensable *in vivo* (Genevaux, Keppel et al. 2004; Kramer, Patzelt et al. 2004; Kramer, Rutkowska et al. 2004). TF is not an ATP-dependent chaperone. The C-terminal domain of TF exhibits chaperone activity on its own. The individual C domain or C domain-containing combinations (NC and PC) illustrated chaperone activity both *in vitro* and *in vivo* (Merz, Hoffmann et al. 2006). Deletion of 53 residues from the C domain resulted in complete loss of TF chaperone activity *in vitro* and greatly reduced the activity *in vivo* (Merz, Hoffmann et al. 2006; Zeng, Yu et al. 2006).

Several co-crystal structures of TF and the ribosome were solved in different organisms. The first TF structure was solved with *E. coli* TF complexed with the archaeon *Haloarcula marismortui* 50S large ribosomal subunit at resolution of 2.7 Å (Ferbitz, Maier et al. 2004). The unique fold of TF resembles a ‘crouching dragon.’ The N-terminal domain is connected to the PPIase domain via a long linker at the opposite end of the molecule with the C-terminal domain inserted between the two domains (Figure 2B). The overall dimensions of TF are 122×59×63Å (Hesterkamp and Bukau 1996; Zarnt, Tradler et al. 1997; Ferbitz, Maier et al. 2004). This co-crystal structure of the ribosome-bound TF showed that TF binds to the ribosomal proteins L23 and L29 through its N-terminal domain. TF arcs over the exit tunnel and creates a shielding pocket which allows nascent chains to fold. The “folding space” was estimated to accommodate a domain of ~14 kDa (Ferbitz, Maier et al. 2004). However, this model is controversial since TF is absent in archaea and the shielding space was not observed in the structure of *Deinococcus radiodurans* TF co-crystallized with *Deinococcus radiodurans* ribosome (Schlunzen, Wilson et al. 2005). From the structure of the homologous complex of *Deinococcus radiodurans* TF and 50S subunit, it shows a long extension of L24 protein (absent in archaea) that occupies the proposed “folding space” of TF (Schlunzen, Wilson et al. 2005). A more recent cryo-EM structure of full length *E. coli* TF bound to a translating *E. coli* ribosome agrees with the initial crystallographic model, in which TF arches over the ribosomal exit tunnel and there is space...
between TF and the ribosome large enough to accommodate a small folded protein domain (Merz, Boehringer et al. 2008). It is proposed that TF is a flexible protein and that it can swing around its attachment point to the ribosome in all directions, which may account for the different orientations seen in different crystal structures.

In the interior of TF, there are multiple substrate binding sites. The hydrophobic side chains in the binding sites provide hydrophobic contacts for unfolded nascent chains (Ferbitz, Maier et al. 2004; Merz, Boehringer et al. 2008; Lakshmipathy, Tomic et al. 2007). It was found that the residence time of TF on translating ribosomes correlates with the hydrophobicity of the nascent chains (Kaiser, Chang et al. 2006). It was suggested that TF assist in protein folding in a dynamic way, in which it first binds to the hydrophobic regions of polypeptides, and through binding and releasing substrates, TF helps them achieve their native states (Kaiser, Chang et al. 2006; Rutkowska, Mayer et al. 2008). In addition to hydrophobic interactions, TF cavity also exposes hydrophilic side chains which may be involved in electrostatic interactions with substrates. From the co-crystal structure of *T. maritima* TF and its substrate S7, which is a ribosomal protein, it was shown that the tmTF:tmS7 binding interface is highly charged and exceptionally large (Martinez-Hackert and Hendrickson 2009). Therefore, the wide range of substrate specificities of TF may come from its broad and versatile binding interfaces.

### 1.3.2 Mechanism of functions of TF

TF binds to the ribosome in a 1:1 stoichiometry through the ribosomal protein L23 (Patzelt, Kramer et al. 2002), which is located at the peptide exit tunnel of 50S large ribosomal subunit; thus TF virtually interacts with all nascent chains that are emerging of the ribosomal tunnel (Lill, Crooke et al. 1988; Valent, Kendall et al. 1995; Hesterkamp and Bukau 1996; Beck, Wu et al. 2000; Kramer, Rauch et al. 2002). *In vivo*, TF is estimated to be 2- to 3- fold molar excess over ribosomes in the *E. coli* cytosol (Lill, Crooke et al. 1988). The $K_d$ between TF and non-translating ribosomes is approximately 1.2 µM (Patzelt, Kramer et al. 2002) but the affinity increases up to 30-fold in the presence of nascent chains (Rutkowska, Mayer et al. 2008). A recent study on ribosome profiling revealed that TF withholds full engagement to the ribosome until ~100 amino acids are translated to allow other modifying enzymes to interact with the nascent chain first (Oh, Becker et al. 2011).
TF and DnaK have overlapping functions as shown by genetic studies because deletion of TF and DnaK genes results in synthetic lethality to *E. coli* cells grown at temperatures higher than 37°C (Deuerling, Schulze-Specking et al. 1999; Teter, Houry et al. 1999). TF and DnaK collaborate to promote folding of newly synthesized proteins since deletion of *tig* gene results in doubling the number of nascent chains interacting with DnaK (Teter, Houry et al. 1999). However, due to the redundant functions of TF and DnaK, it has been difficult to identify the physiological function of TF as *tig* deletion strain shows no defects under normal growth conditions. The *in vivo* function and substrates remained unclear until it was shown in 2011 that TF targets membrane proteins to be translocated to the outer membrane and many substrates of TF included outer membrane proteins (Oh, Becker et al. 2011). It was found that TF promotes the folding and translocation of β-barrel outer membrane proteins (Lee and Bernstein 2002; Ullers, Houben et al. 2003; Oh, Becker et al. 2011). It is worth noting that this finding corresponds with the pioneering studies by Wickner and coworkers, who first proposed the function of TF was to stabilize proOmpA in a transport-competent conformation for translocation into outer membrane (Crooke, Guthrie et al. 1988).

Martinez-Hackert and Hendrickson proposed that TF might have other functions, possibly linked to degradation. It was shown in the chaperone deficient background (Δ*tig*Δ*dnaKdnaJ*) cells, overexpressing TF alleviates aggregation and restores cell viability at 37°C. Surprising, TF has greater effect on cell viability than on aggregation. TF was able to keep protein aggregation levels low *in vivo*, however it was not through resolubilizing protein aggregates but more likely through degradation since the total soluble components did not increase. The authors suggested that TF may direct proteins that are prone to aggregation to the ClpXP degradation system (Martinez-Hackert and Hendrickson 2009). However no direct evidence has yet proven any association between TF and the ClpXP system. My work in this regard proves such a mechanism.

### 1.4 Co-translational degradation

Co-translational degradation is well established in eukaryotes (Schubert, Anton et al. 2000; Turner and Varshavsky 2000). Nascent polypeptides emerging from the ribosome and not completely folded might transiently present degradation signals which are recognized by the
ubiquitin system. These nascent proteins can be ubiquitinated and subsequently degraded by the proteasome while bound to the ribosome (Sato, Ward et al. 1998), demonstrating a link between the pathways of protein synthesis and degradation. Previous studies showed that both stable and unstable proteins were susceptible to co-translational degradation by the proteasome. It was first estimated that about 20-55% of newly synthesized nascent chains bearing an N-terminal degradation signal (N-degron) are degraded co-translationally in yeast (Turner and Varshavsky 2000). However, the high number of co-translational degradation reported may have resulted from experimental conditions such as overexpressing E3 ubiquitin ligases (Turner and Varshavsky 2000) which are responsible for recognition of N-degron (Deshaiies and Joazeiro 2009). It was then suggested that under normal circumstances that only a small fraction of nascent chains (<10%) are co-translationally degraded in eukaryotes (Schubert, Anton et al. 2000).

The mechanisms that couple protein synthesis and degradation remain unclear. It has been proposed that the eukaryotic elongation factor 1A (eEF1A) can bind damaged nascent chains that are ubiquitinated and facilitate their delivery to the proteasome. It was also discovered that the eukaryotic initiation factor 3 (eIF3) was assembled into a large supercomplex, termed the “translasome”, which contains the proteasomes, chaperones and components of the translation machinery (including elongation factors, tRNA synthetases, 40S and 60S ribosomal proteins). The data suggested that factors involved in protein synthesis, quality control and degradation are physically linked to facilitate efficient protein biogenesis (Sha, Brill et al. 2009). Furthermore, it was shown that non-stop proteins which have escaped translational repression are removed via co-translational degradation (Chiabudini, Conz et al. 2012) and that RAC-Ssb and two ribosome-bound E3 ligases, Not4 and Ltn1 in yeast, are involved in this process. Deletion of Ltn1 results in accumulation of abortive translation products, suggesting a significant role in co-translational degradation in yeast (Bengtson and Joazeiro 2010). In a parallel study, a Ribosome Quality Control Complex (RQC) comprising the Ltn1 E3 ubiquitin ligase, two other proteins Tae2 and Rqc1, and Cdc48 and its cofactors has been identified to associate with 60S ribosomal subunits containing stalled polypeptides to trigger their degradation (Brandman, Stewart-Ornstein et al. 2012). Hence, in eukaryotes, multiple mechanisms and complexes seem to be involved in facilitating co-translational protein degradation.
The reason why cells degrade newly synthesized proteins is poorly understood. It was suggested that co-translational degradation is another form of protein quality control, and nascent chains that fail to fold rapidly which expose degradation signals are destroyed (Turner and Varshavsky 2000). Such substrates are termed as defective ribosomal products (DRiPs) (Yewdell, Anton et al. 1996). It is proposed that DRiPs are used as sources to generate antigenic peptides in a fast and efficient way. The presence of DRiPs enables major histocompatibility complex (MHC) class I molecules to monitor protein synthesis rates in cells and not protein concentrations. In the case of a viral infection, it can be rapidly detected since the co-translationally degraded viral proteins are readily presented by MHC class I molecules as antigenic peptides to the immune system, particularly since many viruses rapidly highjack the host’s protein synthesis machinery to produce their own proteins. It is important to understand how cells decide the fate of nascent chains and whether protein biogenesis is an inefficient process or it might have other biological functions to degrade newly synthesized proteins. Prior to my work, no strong evidence has been presented for co-translational protein degradation in prokaryotes.

1.5 Clp ATPases and their role in protein unfolding and degradation in E. coli

The Clp ATPases are unique ATP-dependent chaperones responsible for assembly and disassembly of protein complexes. Clp members are found in the cytoplasm in bacteria, yet they are also found in mitochondria and chloroplasts in higher eukaryotes. Clp proteins are members of the AAA+ (ATPases Associated with diverse cellular Activities) superfamily (Confalonieri and Duguet 1995). As the name suggests, it is a large class of ATP hydrolyzing proteins that are involved in different functions of the cell (Vale 2000). Some examples of AAA+ proteins include the motor protein dynein, which is involved in microtubule regulation, helicases in DNA replication, metal chelatases, and regulatory units of proteases including the proteasome (Snider, Thibault et al. 2008). The sequence feature of the AAA+ superfamily is a highly conserved AAA+ module consisted of 200-250 amino acids, and each AAA+ module contains a conserved Walker A motif and a Walker B motif (Snider, Thibault et al. 2008). Walker A motif is
associated with ATP binding while Walker B motif is involved in metal binding and ATP hydrolysis (Schirmer, Glover et al. 1996). Structurally, AAA+ proteins have a unifying feature of forming into ring-shaped hexameric or heptameric complexes (Guenther, Onrust et al. 1997; Lenzen, Steinmann et al. 1998; Bochtler, Hartmann et al. 2000; Zhang, Shaw et al. 2000). Within the AAA+ superfamily, the Clp proteins are divided into two classes: (1) Class 1 Clp ATPase - Clp proteins that contain two nucleotide-binding domains (AAA+) and (2) Class 2 Clp ATPase – Clp proteins that contain only a single AAA+ domain.

In *E. coli*, the main Clp ATPases include ClpA (758 aa), ClpB (857 aa), ClpX (423 aa), and ClpY (also known as HslU, 443 aa). ClpA and ClpB belong to Class I AAA+ proteins which have 2 AAA+ modules while ClpX and ClpY are members of Class II AAA+ with one AAA+ domain that is homologous to the second AAA+ module of ClpA and ClpB (Schirmer, Glover et al. 1996). All four of them belong to the Hsp 100 family, though ClpA is not heat shock induced. ClpA, ClpX and ClpY can function alone as molecular chaperone and also serve as regulatory components of the proteolytic complexes of ClpXP, ClpAP, and ClpYQ (also known as HslUV) (Gottesman, Wickner et al. 1997). The chaperone units of the degradation systems are involved in recognizing, unfolding substrates and translocating them into the proteolytic chamber; unfolding and translocating processes require ATP binding and hydrolysis. The regulatory chaperone units are often found on the same operon as the proteases. On *E. coli* genome, *clpP* and *clpX* constitute the *clpPX* operon while *clpQ* (*hslV*) and *clpY* (*hslU*) form the *clpQY* (*hslUV*) operon (Gottesman, Clark et al. 1993; Wojtkowiak, Georgopoulos et al. 1993; Rohrwild, Coux et al. 1996). The protease units ClpP and HslV do not belong to the AAA+ family. ClpP is a serine protease composed of 14 subunits which oligomerizes into two heptameric rings. ClpQ is a threonine protease containing 12 subunits, and it oligermizes into two hexameric rings. ClpB itself does not associate with proteases and it is found to be essential for cell survival at high temperatures (Squires, Pedersen et al. 1991; Woo, Kim et al. 1992; Kim, Woo et al. 1998). Deletion of *clpB* in *E. coli* results in a slower growth rate at 44°C and a higher death rate above 50°C (Squires, Pedersen et al. 1991; Woo, Kim et al. 1992; Kim, Woo et al. 1998). ClpB chaperone is not responsible for direct protein folding, instead, ClpB functions as a disaggregase that solubilizes aggregates. DnaK transfers preformed protein aggregates to ClpB, ClpB disaggregates the aggregates and transfers them to the Hsp70 system (DnaK/DnaJ/GrpE) which directly assists in protein folding (Goloubinoff, Mogk et al. 1999; Zolkiewski 1999). *clpB* gene is
on the same operon as *grpE*, *dnaK*, and *dnaJ* in *T. thermophilus* where the protein products function cooperatively to protect the cell from thermal stress. The ClpB homolog in yeast, Hsp104, also acts as a disaggregase and works in conjunction with the Hsp70 system in yeast (Parsell, Kowal et al. 1994; Glover and Lindquist 1998). Hsp104, together with the Hsp70 system are the only rescue mechanisms known to date to restore protein aggregates back to soluble proteins *in vivo*.

### 1.5.1 ClpXP degradation system

The ClpXP system consists of the ClpX AAA+ ATPase and ClpP cylindrical protease (Figure 3). ClpP is an active peptidase which can degrade small peptides up to ten residues. ClpX ATPase can bind to one or both ends of the ClpP protease and together the ClpXP system can degrade well folded protein substrates (Grimaud, Kessel et al. 1998; Ortega, Lee et al. 2002). ClpX recognizes substrates, denatures them, and then translocates the polypeptide chains into the proteolytic chamber of ClpP for degradation. The denaturation and translocation processes by ClpX require ATP, while degradation by the protease ClpP is nucleotide independent. ClpX is a hexameric ATPase and ClpP is a cylindrical serine protease composed of two heptameric rings stacked on top of each other (Figure 3). Orthologues of ClpX and ClpP are found in bacteria, mammals (in mitochondria), plants (in plastids), and some yeasts (Wojtyra, Thibault et al. 2003). My study focuses on the ClpXP from *Escherichia coli*, which is the first ClpX family protease to be isolated and studied.
Figure 3. ClpXP degradation cycle

ClpXP degradation system is composed of ClpX ATPase and ClpP protease. ClpX has an N-terminal zinc binding domain (ZBD) and an AAA+ domain. The processes of recognition, denaturation and translocation of substrates are performed by the AAA+ protein ClpX. The unfolded polypeptide chains are degraded in the proteolytic chamber of ClpP.
1.5.2 ClpX AAA+ ATPase

1.5.2.1 ClpX discovery

The discovery of ClpX came after the discovery of another AAA+ ATPase ClpA, which was the first Clp chaperone identified as part of the two component ATP-dependent degradation system (Hwang, Park et al. 1987; Katayama-Fujimura, Gottesman et al. 1987). ClpA belongs to the Class 1 of the Clp ATPase family which contains two AAA+ modules. The protein casein was first discovered as a substrate of ClpAP, thus the protease was named Caseinolytic Protease (Clp) (Katayama-Fujimura, Gottesman et al. 1987). It was found that degradation of casein requires ClpA, ClpP and ATP hydrolysis. Later on, ClpX was discovered as another AAA+ ATPase that functioned with the ClpP protease but recognized different substrates. The differences in substrate specificity between ClpXP and ClpAP suggested that the AAA+ ATPases are involved in substrate recognition. ClpXP was identified as the major protease that degraded λO bacteriophage replication protein (Wojtkowiak, Georgopoulos et al. 1993). The half-life of λO protein is only one to two minutes in wildtype E. coli cells; the stability of λO dramatically increased in clpX or clpP (but not clpA) deletion mutants where λO remained stable for over forty minutes (Gottesman, Clark et al. 1993). ClpX alone also acts as a molecular chaperone towards λO. ClpX is capable of protecting λO from thermal inactivation and aggregation in the same manner as the DnaK/DnaJ/GrpE system (Wawrzynow, Wojtkowiak et al. 1995). MuA is a transposase protein that catalyzes DNA cleavage and joining reactions in Mu phage. ClpX functions as an ATP-dependent disassembly chaperone for MuA transposase. ClpX disassembles the hyperstable MuA transposase tetramers from DNA, destabilizing the MuA-DNA complex, which permits Mu phage DNA replication. The remodeling mechanism is entirely dependent on ClpX activity and not ClpP as Mu phage growth was severely compromised in ΔclpX strain but unaffected in ΔclpP strain (Mhammedi-Alaoui, Pato et al. 1994; Burton and Baker 2005). More details on ClpXP substrates are discussed further below.

1.5.2.2 ClpX structure

ClpX is composed of the zinc binding domain (ZBD) at the N-terminus and an AAA+ domain. Full length ClpX structure has not been solved yet but the structure of the ZBD and the AAA+
domain have been determined independently (Donaldson, Wojtyra et al. 2003; Wojtyra, Thibault et al. 2003; Kim and Kim 2003; Park, Lee et al. 2007; Glynn, Martin et al. 2009). Both the full length ClpX and ClpX lacking the ZBD (ClpXΔZBD) hexamerize in the presence of nucleotides and both constructs bind to ClpP, though recognition of certain substrates (eg. λO) requires the presence of ZBD. Structurally, the AAA+ domain (ClpXΔZBD) can be further divided into a large and a small AAA+ subdomain (Figure 4A and E). The large and small AAA+ subdomains function together and form hexameric rings. Nucleotides bind in between the large and small AAA+ subdomains (Figure 4E) (Kim and Kim 2003; Glynn, Martin et al. 2009). The structure of AAA+ domain was initially solved in the ADP-bound state of monomeric Helicobacter pylori ClpXΔZBD (Kim and Kim 2003). The structure was found similar to that of HsIu, and the hexameric model of Helicobacter pylori ClpXΔZBD was constructed using the hexameric structure of E. coli HsIu as a template (Kim and Kim 2003). The model shows that ClpXΔZBD is a six-fold symmetrical hexamer, and the corresponding atoms of all six subunits are in planes perpendicular to the symmetry axis (Figure 4B and C) (Kim and Kim 2003).

However subsequent crystal structure of E.coli ClpXΔZBD, in which three subunits were connected by linkers, showed that ClpX forms an asymmetric ring hexamer (Glynn, Martin et al. 2009). Four subunits (loadable or L) bind to nucleotides and the nucleotide binding sites of the other two subunits (unloadable or U) are destroyed due to large rotations (~80°) between the large and the small AAA+ domains (Figure 4E). The nucleotide-binding competent and non-competent subunits of ClpX are organized as L-L-U-L-L-U, resulting in a highly asymmetrical hexamer.

The structure of the 60 amino acids long ZBD was first solved by NMR (Donaldson, Wojtyra et al. 2003; Wojtyra, Thibault et al. 2003). ZBD folds independently of the AAA+ module as a dimer where each protomer has two β-hairpins and one α-helix. Each monomer of ZBD has a treble clef fold with a zinc ion chelated by four conserved cysteines which are essential for the ZBD dimerization (Figure 4D) (Donaldson, Wojtyra et al. 2003; Wojtyra, Thibault et al. 2003). Crystal structure of ZBD with a peptide corresponding to the C-terminus of the ClpX cofactor SspB also show similar structure as the solution NMR structure (Park, Lee et al. 2007).
Figure 4. Structure of ZBD and ClpX

(A) Domain arrangement and functional motifs of *E. coli* ClpX. (B) Side and (C) top view of the hexameric model of the AAA+ domain of ClpX based on the solved monomeric *H. pylori* ClpX structure (PDB file: 1UM8) (Kim and Kim 2003) and the hexameric *E. coli* HslU structure (Bochtler, Hartmann et al. 2000). One monomer is shown in orange while the other 5 subunits are shown in grey. (D) NMR structure of ZBD dimer from *E. coli* ClpX (PDB file: 1OVX) (Donaldson, Wojtyra et al. 2003). Monomers are shown in green and blue. Four cysteines chelating Zn$^{2+}$ are shown as sticks and Zn$^{2+}$ atoms are shown as brown dots. (E) Top view of hexameric AAA+ domain of ClpX with three domains linked (PDB file: 3HWS), coloured by subunits.
1.5.2.3 ClpX function – substrate recognition

1.5.2.3.1 Sequence recognition – SsrA tag

ClpX recognizes short unstructured sequences called degradation tags or degrons. For example, the *E. coli* SsrA tag (AANDENYALAA) is a degradation tag recognized by several proteases *in vitro* and *in vivo*. SsrA tag is added to translating proteins by tmRNA when ribosomes stall. tmRNA, also known as SsrA RNA or 10SaRNA, functions as a rescue system in *E. coli* when the ribosome is stalled. It has the dual function of tRNA and mRNA to mediate modification of nascent chains when ribosomes stall during mRNA translation (Keiler, Waller et al. 1996; Roche and Sauer 1999). The SsrA tag is added co-translationally by tmRNA to the C-terminus of incomplete nascent chains on the stalled ribosomes (Roche and Sauer 1999). SsrA tagged proteins are universal substrates of ClpXP. In *E. coli*, cytoplasmic proteases ClpAP, and Lon protease also recognize the SsrA degradation tag and target the tagged proteins for degradation *in vivo* and *in vitro* (Keiler, Waller et al. 1996; Choy, Aung et al. 2007). *In vivo*, ClpXP is the major protease involved in degrading SsrA-tagged proteins. Model substrates with the SsrA tag added to the C-terminus have been used extensively to study the mechanism of function of ClpXP and ClpAP systems. ClpX binds directly to the last three residues (LAA) in the SsrA tag at the C-terminus of the SsrA tag (Flynn, Levchenko et al. 2001). Replacing the last two residues AA by DD abolishes ClpX recognition (Gottesman, Roche et al. 1998).

The binding affinity of SsrA tagged proteins to ClpX is approximately 1 µM (Bolon, Wah et al. 2004). The GYVG pore in the axial channel of ClpX hexamer is responsible for binding SsrA-tagged proteins (Figure 4C). Mutations of the pore loops decreases the affinity of ClpXP for SsrA-tagged substrates, suggesting that the primary binding site of SsrA peptides is located in the GYVG pore (Siddiqui, Sauer et al. 2004). The RKH loops which surround the entrance to the central pore of ClpX play important roles in substrate specificity (Figure 4B). It has been demonstrated that the human ClpX which contains the same GYVG loops as *E. coli* ClpX could not recognize the SsrA tag, but a mutant containing the RKH loops transplanted from *E. coli* ClpX acquires this activity (Martin, Baker et al. 2008). Furthermore, mutations of the RKH loops change the substrate specificity of ClpX to recognize different protein substrates.
(Farrell, Baker et al. 2007), suggesting that the ClpX substrate specificity is not optimized for a single class of proteins but rather contains a broad range of substrates.

1.5.2.3.2 Sequence recognition – substrate λO

Other than SsrA tagged proteins, there are natural substrates of ClpX that are recognized by their internal sequences. For example, the λ phage replication protein O and Mu phage transposase MuA contain degradation signals in the N-terminal and C-terminal sequences, respectively. Bacteriophage λ DNA replication initiates from the replication origin (oriλ) and proceeds bidirectionally. The bacteriophage λO protein is required for the initiation of λ DNA replication during the lytic cycle of λ phage (Skalka 1977; Anderl and Klein 1982; Tsurimoto and Matsubara 1982; Wold, Mallory et al. 1982). λO proteins bind to four repeating sequences of the λ replication origin (oriλ), forming the O-some DNA-nucleoprotein structure (Tsurimoto, Hase et al. 1982; Roberts and McMacken 1983; Dodson, Roberts et al. 1985; Liberek, Georgopoulos et al. 1988; Echols 1990). Subsequently, λ replication protein P and the host helicase DnaB are recruited to bind to the O-some, forming a large oriλ:O:P:DnaB preprimosomal complex (Wegrzyn, Wegrzyn et al. 1995; Wegrzyn, Wegrzyn et al. 1996). The preprimosomal complex is activated by DnaK/DnaJ/GrpE system by dissociating λP from the complex which allows DnaB helicase to start the unwinding process. The unwound λ DNA strands provide room for DNA polymerase III holoenzyme to bind, which leads to the assembly of a functional replication complex (Zylicz, Ang et al. 1989; Wegrzyn, Pawlowicz et al. 1992; Wegrzyn, Szalewska-Palasz et al. 1995; Wegrzyn, Wegrzyn et al. 1995; Wegrzyn, Wegrzyn et al. 1996). The free λO proteins are degraded rapidly by ClpXP in vivo but they become protected from proteolysis when they form a complex with λ DNA. From O-some formation to the preprimosome complex, the stability of λO increases as it forms larger complexes, which could provide an explanation on why inactivation of the clpX gene does not seem to affect the lysis or lysogeny decision of λ phage since λO is protected when bound to λ DNA (Zylicz, Liberek et al. 1998).

ClpX also functions as a molecular chaperone towards λO by protecting λO from thermal inactivation and aggregation (Wawrzynow, Wojtkowiak et al. 1995). Moreover, ClpX disaggregates preformed λO aggregates and reactivates denatured λO. λO protein is composed of an N-terminal DNA-binding domain and a C-terminal domain that mediates protein-protein
interactions. It was found that the N-terminal sequences of λO are essential for targeting λO degradation by ClpXP (Gonciarz-Swiatek, Wawrzynow et al. 1999). However, there are residues throughout λO sequence that affect its degradation and there is no exact consensus recognition sequence identified. Yet, the first 12 residues at the N-terminus of λO are essential for the recognition by ClpX (Gonciarz-Swiatek, Wawrzynow et al. 1999). The resistance of λO protein assembled in DNA-protein complexes (O-some and preprimosome) to the ClpXP degradation suggests that the ClpX recognition motifs in the nucleoprotein structures are inaccessible to ClpX. It is also possible that protein-protein and protein-DNA interactions within DNA-protein complex stabilize λO against the critical unfolding step mediated by ClpX.

1.5.2.3.3 Sequence recognition – substrate MuA

Replication of Mu phage requires ClpX by disassembling the transpososomes, which consist of a MuA tetramer bound to recombined DNA. Mu phage propagates its genome into the host by transposing the Mu phage DNA into the bacterial chromosome during replication. MuA directly catalyzes the recombination of Mu phage DNA and host DNA. First, MuA transposase binds to specific sequences at the left and right ends of the Mu genome, and interactions between MuA subunits bring the Mu DNA ends together to form the transpososome, which consists of four MuA subunits (Mizuuchi 1992). Second, MuA subunits nick one DNA strand at each end of the Mu genome, generating the cleaved donor complex (CDC). Subsequently, the 3' hydroxyl group at the Mu DNA ends attacks a target DNA molecule of the host genome and generates the strand transfer complex (STC1). The STC1 is inhibitory to the DNA-replication machinery and blocks the completion of replicative transposition (Mhammedi-Alaoui, Pato et al. 1994; Nakai and Krukltis 1995). ClpX recognizes MuA, unfolds MuA, and remolds STC1 into a less stable complex STC2. STC2 ensures the recruitment of bacterial replication machinery to the recombination site to replicate Mu DNA. Therefore, the remodeling of STC1 to STC2 is essential for Mu replication (Burton and Baker 2005). MuA protein is monomeric in solution and forms a tetramer upon DNA binding. When bound in the transpososomes, MuA becomes resistant to ClpXP degradation but remains susceptible to ClpX unfoldase activity. Mu phage propagation is solely dependent on ClpX unfolding activity and not ClpXP degradation activity since the growth of Mu phage is only affected in the *clpX* but not *clpP* deletion strain.
ClpX recognizes the specific C-terminal sequence QNRRKKAI of MuA which has been classified as one of the degradation signals that targets the protein for ClpXP degradation (Levchenko, Luo et al. 1995; Flynn, Neher et al. 2003).

### 1.5.2.3.4 Other substrates

Proteomic studies revealed more than one hundred ClpXP substrates by trapping the endogenous substrates in the proteolytic chambers of inactive ClpP proteases *in vivo* (Flynn, Neher et al. 2003). The trapped proteins were then purified and identified by mass spectrometry. The analysis established five classes of ClpXP degradation tags based on recurring motifs of the trapped substrates. Two classes of C-terminal degradation signals share similar sequences with the SsrA tag (LAA) and MuA tag (RRKKAI) respectively. The N-motif 1 shares similar homology with the N-terminus of λO (polar-T/ϕ-ϕ-basic-ϕ). The N-motif 2 has the consensus sequence of NH2-M-basic-ϕ3-X5-ϕ and the consensus sequence of N-motif 3 is ϕ-X-(polar-X)2-basic-polar, where the symbol ϕ represents hydrophobic amino acids (Flynn, Neher et al. 2003). The large diversity of ClpXP substrate specificity indicates that ClpXP can degrade proteins with diverse structures and stabilities, allowing ClpXP to function in protein quality control and regulatory activities.

### 1.5.2.3.5 SspB adaptor mediated recognition

SspB (stringent starvation protein B) is a dimeric adaptor protein that helps deliver SsrA-tagged (AANDENYALAA) proteins for ClpXP degradation (Levchenko, Seidel et al. 2000). Each subunit of SspB dimer contains a groove at the N-terminus which binds to part of the SsrA tag (AAND and Y) that is different from the ClpX binding site (LAA) (Flynn, Levchenko et al. 2001). One SspB dimer is found to bind to one ClpX hexamer. The C-terminal tail of each SspB subunit docks to the zinc binding domain of ClpX and the N-terminus of SspB tethers to substrate (Wojtyra, Thibault et al. 2003; Dougan, Weber-Ban et al. 2003; Wah, Levchenko et al. 2003; Bolon, Wah et al. 2004; Hersch, Baker et al. 2004; Park, Lee et al. 2007). The three components of SspB, ClpX and SsrA-tagged substrate form a stable complex, and the tight
affinity between ClpX, SspB adaptor protein and the substrate allows for efficient degradation of SsrA-tagged proteins in the cell (Wah, Levchenko et al. 2002; Bolon, Grant et al. 2004; Bolon, Wah et al. 2004; Hersch, Baker et al. 2004). This stable complex is disrupted by ClpX pulling the SsrA tag on the C-terminus of a substrate to initiate degradation and thereby breaking contacts with SspB at the same time. SspB targets SsrA-tagged proteins specifically to ClpXP and keeps the substrates away from ClpAP, which also recognizes the SsrA tag. ClpA recognizes the residues AA (SspB also binds to these residues) and ALA of the SsrA tag. Since SspB has higher binding affinity towards SsrA tag than ClpA, the presence of SspB inhibits the binding of SsrA-tagged substrates to ClpA and directs them to ClpXP (Levchenko, Seidel et al. 2000).

### 1.5.2.4 ClpX mechanism of function

ClpX recognizes substrates, denatures them, and then translocates the unfolded polypeptide chains into the proteolytic chamber of ClpP for degradation (Figure 3). The denaturation and translocation processes by ClpX require ATP, while degradation by the protease ClpP is nucleotide independent. ATP binding and hydrolysis drive the movements in the ClpX AAA+ ring and thus are closely coupled with degradation (Burton, Baker et al. 2003). ATP hydrolysis rate increases upon the addition of SsrA tagged proteins. Among the degradation processes of substrate binding, unfolding, translocation, degradation and product release, unfolding is the rate limiting step and it consumes the most energy (Kim, Burton et al. 2000; Kenninston, Baker et al. 2003). The degradation of native titin-SsrA by ClpXP was about 16 fold slower than the degradation of an unfolded mutant (Kenninston, Baker et al. 2003), suggesting that unfolding is the rate limiting step for degradation.

The stability of substrate structure near the degradation tag affects unfolding and degradation rate of the substrate. For example, ClpXP degrades N-terminally tagged titin about 4 times faster than the same substrate that is C-terminally tagged, suggesting that faster unfolding rate is generated from pulling on the N terminus tag and the neighbour structures (Kenniston, Baker et al. 2005). ClpX unfolds substrates by pulling on the end of the substrate in a probabilistic manner, subsequently translocating the polypeptide chain through ClpX chamber, which results in the destabilization of the rest of the substrate (Lee, Schwartz et al. 2001). If a
protein resists the unfolding force of ClpX, ClpX can either maintain the binding and continue to pull on the substrate, or ClpX can dissociate from the substrate first, rebind and eventually degrade the substrate (Lee, Schwartz et al. 2001; Kenniston, Baker et al. 2005). The ability of ClpXP to freely dissociate from substrates that resist unfolding allows it to preferentially degrade unfolded or unstable native proteins in a mixture of substrates (Kenniston, Baker et al. 2005).

After the substrate is unfolded, the polypeptide chain is translocated by ClpX into the ClpP chamber for degradation. Translocation is also ATP-dependent but requires less energy than unfolding. The distance from the top of ClpX to the ClpP chamber that an unfolded substrate traverses is approximately 90 Å (Ortega, Singh et al. 2000). Translocation occurs through a threading mechanism by ClpX where the unfolded polypeptide is pushed down from ClpX and threaded into the proteolytic chamber of ClpP where degradation takes place. ClpXP unfolds and degrades a wide variety of proteins with very different structures and stabilities, making it a central degradation system for protein quality control.
1.5.3 ClpP: A Distinctive Family of Cylindrical Energy-Dependent Serine Proteases

1.5.3.1 Summary

Processes maintaining protein homeostasis in the cell are governed by the activities of molecular chaperones that mainly assist in the folding of polypeptide chains and by a large class of proteases that regulate protein levels through degradation. ClpP proteases define a distinctive family of cylindrical, energy-dependent serine proteases that are highly conserved throughout bacteria and eukaryota. They typically interact with ATP-dependent AAA+ chaperones that bind and unfold target substrates and then translocate them into ClpP for degradation. Structural and functional studies have provided a detailed view of the mechanism of function of this class of proteases.

1.5.3.2 Overview

Proteases play an essential role in protein quality control by removing short-lived regulatory proteins, as well as proteins that are misfolded and damaged, thus maintaining cellular homeostasis. A significant proportion of protein degradation in the cell is carried out by oligomeric, cylindrical, self-compartmentalized, energy-dependent proteases. ClpP (Caseinolytic Protease) is a representative member of these cylindrical proteases. Other members include HslV and the 20S proteasome core particle.

ClpP is a highly conserved serine protease present throughout bacteria and eukaryota; it seems to be absent in archaea (Wong and Houry 2004), mollicutes (Wong and Houry 2004), and some fungi (Table 1). ClpP was first identified in Escherichia coli (Katayama-Fujimura, Gottesman et al. 1987; Katayama, Gottesman et al. 1988), and numerous studies have been published on it since its discovery. The X-ray structure of ClpP has been solved from several different organisms (Figure 5 and Figure 6). All the structures show similar features: the protease is comprised of 14 subunits arranged into two heptameric rings forming a cylindrical-like structure which encloses a large chamber containing the protease active sites. Entrance into the
chamber occurs through axial pores in the cylinder. ClpP typically forms complexes with AAA+ (ATPases Associated with various cellular Activities) chaperones that denature substrates and then translocate them through the axial pores into the proteolytic chamber of the protease for degradation. ClpP degrades proteins into peptides of about 7 to 8 residues (Choi and Licht 2005) which are subsequently released from the chamber. In this section, I will describe the distribution of ClpP proteases in different kingdoms of life and then discuss in more detail ClpP structure and mechanism of function.
Table 1. The distribution of ClpP and ClpR across different organisms

<table>
<thead>
<tr>
<th>Superkingdom</th>
<th>Phylum</th>
<th>Class</th>
<th>Species</th>
<th>ClpP</th>
<th>ClpR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaea</td>
<td>Actinobacteria</td>
<td>Actinobacteria</td>
<td>Mycobacterium tuberculosis H37Rv</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Bacteria</td>
<td>Actinobacteria</td>
<td>Actinobacteria</td>
<td>Bacteroides fragilis NCTC 934</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorobi</td>
<td>Chlorobi</td>
<td>Chlorobium chlorochromatii DD3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cyanoacteria</td>
<td>Cyanoacteria</td>
<td>Prochlorococcus marinus str. MIT 9313</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Cyanobacteria</td>
<td>Cyanobacteria</td>
<td>Synechococcus elongatus</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Fimbicutes</td>
<td>Fimbicutes</td>
<td>Synechocystis PCC6803</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>Firmicutes</td>
<td>Barrella azzelli (strain PKo)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>Firmicutes</td>
<td>Barrella atzelli (strain PKo)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>Firmicutes</td>
<td>Barrella azzelli (strain PKo)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Agrobacterium tumefaciens str. C58</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Anaplasma phagocytophilum HZ</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Rickettsia felis URRWXCal2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Bordetella pertussis Tohama I</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Bordetella parapertussis 12822</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Burkholderia cenocepacia AU 1054</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Ralstonia eutropha JMP134</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Chromobacterium violaceum ATCC 12472</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Neisseria gonorhoeae FA 1090</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Neisseria meningitidis FAM18</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Desulfotalea psychrophila LSv54</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Geobacter metallireducens GS-15</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Campylobacter jejuni RM1221</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Helicobacter pylori J99</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Escherichia coli K12</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Salmonella typhimurium LT2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Yersinia pestis Antiqua</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Haemophilus ducreyi 35000HP</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Pasteurella multocida subsp. Multocida str. Pm70</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Pseudomonas aeruginosa 2192</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Vibrio cholerae O1 biovar eltor str. N16961</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Proteobacteria</td>
<td>Xanthomonas axonopus pv. citri str. 306</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Apicomplexa</td>
<td>Aconoida</td>
<td>Plasmodium falciparum 3D7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Ascomycota</td>
<td>Saccharomyces</td>
<td>Saccharomyces cerevisiae</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ascomycota</td>
<td>Schizosaccharomyces</td>
<td>Schizosaccharomyces pombe 972h-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ascomycota</td>
<td>Sordariomycetes</td>
<td>Gibberella zeae PH-1 (incomplete sequence)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ascomycota</td>
<td>Eurotomyces</td>
<td>Neurospora crassa OR74A (incomplete sequence)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ascomycota</td>
<td>Eurotomyces</td>
<td>Aspergillus fumigatus AF293</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorophyta</td>
<td>Chlorophyceae</td>
<td>Chlamydomonas reinhardtii</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Streptophyta</td>
<td>Streptophyta</td>
<td>Arabidopsis thaliana</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Nematoda</td>
<td>Chromadorea</td>
<td>Caenorhabditis elegans</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arthropoda</td>
<td>Insecta</td>
<td>Drosophila melanogaster</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Mus musculus</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Rattus norvegicus</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Homo sapiens</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

E. coli ClpP, Synechocystis PCC6803 ClpR, and Arabidopsis ClpR protein sequences were used to BLAST against all archaeal genomes and selected organisms from bacteria and eukaryota in NCBI. A cut-off of 40% identity with at least 150 amino acids aligned was typically used.
1.5.3.3 The distribution of ClpP proteases in different kingdoms of life

1.5.3.3.1 Bacterial ClpPs

ClpP is found in all bacteria sequenced to date except for Mollicutes (Wojtyra, Thibault et al. 2003; Yu and Houry 2007). Table 1 lists the number of ClpP copies present in selected organisms. The majority of bacteria contain only one copy of ClpP, although there are exceptions such as for actinobacteria, chlamydiae, cyanobacteria, and others (Table 1) which have multiple isoforms of the protease. In addition to ClpP, cyanobacteria contain a ClpP paralog that does not have all 3 residues of the Ser-His-Asp catalytic triad. This inactive version of ClpP is called ClpR (Peltier, Ytterberg et al. 2001) and is mostly present in plants in addition to cyanobacteria. For example, cyanobacteria Synechocystis PCC6803 contains three copies of ClpP and one copy of ClpR (Table 1 and Figure 7) (Peltier, Ytterberg et al. 2001; Schelin, Lindmark et al. 2002).

ClpP was first identified and mostly studied in *Escherichia coli*, which contains only one copy of the protease (Table 1). *E. coli* ClpP consists of 207 amino acids including an N-terminal prosequence which acts as a regulatory peptide and undergoes autocatalytic cleavage during folding to yield a mature ClpP of 193 residues (Maurizi, Clark et al. 1990). In *E. coli*, ClpP forms complexes with AAA+ chaperones, ClpX and ClpA, which belong to the Clp/Hsp100 family. ClpA is an 83 kDa ATPase that contains two AAA+ domains, while ClpX contains only one AAA+ domain (46 kDa), which is homologous to the second AAA+ domain of ClpA. ClpA and ClpX are hexameric chaperones that can stack onto one or both ends of ClpP to form ClpAP or ClpXP holoenzyme complexes. Substrates are recognized, unfolded by ClpX or ClpA, and then threaded into the ClpP proteolytic chamber through the narrow axial pores for degradation (Figure 8). Only the unfolding and threading by the chaperones require ATP binding and hydrolysis, while proteolysis by ClpP does not require nucleotide hydrolysis. Other bacteria contain ClpA paralogs, such as ClpC, ClpE, and ClpL (Butler, Festa et al. 2006).

The most general substrates for the *E. coli* ClpXP or ClpAP system are those with an SsrA C-terminal tag (Keiler, Waller et al. 1996). The SsrA tag consists of 11 hydrophobic residues that are added by a tmRNA, also called SsrA or 10Sa RNA, to the C-terminus of nascent chains whose translation is stalled on the ribosome, targeting them for degradation (Gottesman, Roche et al. 1998; Weber-Ban, Reid et al. 1999; Gillet and Felden 2001). SsrA
tagging can be thought of as the prokaryotic analogy to eukaryotic ubiquitination.

Without ATPase components, only small peptides can enter the ClpP chamber (Thompson, Singh et al. 1994; Gottesman, Maurizi et al. 1997); hence, ClpP on its own cannot degrade folded proteins. Recently, however, a group at Bayer described the discovery of acyldepsipeptides (ADEPs) that, when added to ClpP, allow the protease to degrade folded native proteins in the absence of its cognate chaperones (Brotz-Oesterhelt, Beyer et al. 2005). The drug was initially found to rescue mice challenged with lethal infections of *Enterococcus faecalis* and *Staphylococcus aureus*, and the target for the drug was then identified to be ClpP. ClpP is activated by ADEPs to degrade, or at least to cleave, folded proteins in the absence of the cognate Clp ATPases. The consequences of the action of the drug *in vivo*, is that the ADEPs cause unregulated proteolysis by ClpP, subsequently, triggering cell death in gram-positive bacteria. Gram-negative bacteria, on the other hand, were found to be able to survive drug treatment because they have efficient efflux pumps to remove the drug from the cell, but deletion of those pumps or the use of permeabilizing agents made gram-negative bacteria vulnerable to ADEP treatment as well (Brotz-Oesterhelt, Beyer et al. 2005). The mechanism of how ADEPs activate ClpP in the absence of its cognate chaperones is currently unknown. The discovery of ADEPs may lead to development of a novel class of antimicrobial agents.
Figure 5. The structures of tetradecameric ClpP.

Shown are the side and top views of the X-ray structures of tetradecameric ClpPs from *E. coli* (1YG6), *S. pneumoniae* (1Y7O), *M. tuberculosis* (2CE3), *P. falciparum* (2F61), and human (1TG6). The structures were drawn using PyMOL (http://pymol.sourceforge.net/).
1.5.3.3.2 Human ClpP

In the human genome, ClpP is encoded on chromosome 19 (Corydon, Wilsbech et al. 2000). Immunofluorescence studies illustrated that human ClpP is located in the mitochondrial matrix (de Sagarra, Mayo et al. 1999), although its role is still not known. The mature human ClpP has about 56 residues removed from its N-terminus (von Heijne, Steppuhn et al. 1989; de Sagarra, Mayo et al. 1999), which include the mitochondrial targeting sequence and the prosequence. Mature human ClpP shares high sequence identity (56%) and similarity (71%) with *E. coli* ClpP. However, human ClpP has an additional 28 residues present at its C-terminus. The function of the extended C-terminus is unknown although it seems to be a unique feature for mammalian ClpPs (Kang, Maurizi et al. 2004). The only ATPase component identified so far for human ClpP is ClpX, which is encoded on chromosome 15 (Corydon, Wilsbech et al. 2000). Human ClpX also contains an extended N-terminal sequence which targets ClpX to the mitochondria; the mature human ClpX shares 44% identity and 62% similarity with *E. coli* ClpX (Figure 7).

Unlike *E. coli* ClpP that exists solely as a double-ring tetradecamer, human ClpP exists as a single heptameric ring under physiological conditions (Kang, Ortega et al. 2002). The heptameric human ClpP does not exhibit protease activity and has a very low peptidase activity compared to *E. coli* ClpP (Kang, Dimitrova et al. 2005). The lack of activity is more likely due to an inactive orientation of the catalytic triads in human ClpP heptamer. A single heptameric ring cannot compartmentalize its catalytic triads; indeed, the active sites in human ClpP have been shown to be solvent accessible by trypsin digestion (Kang, Dimitrova et al. 2005). The single-ring human ClpP forms a double ring upon binding to ClpX in the presence of ATP, which suggests that binding of the ATPase component might induce conformational changes in the ClpP heptameric rings leading to the formation of the double ring. The human ClpXP complex has protease activity as well as higher peptidase activity compared to heptameric human ClpP, implying a realignment of the catalytic residues in ClpP into an active conformation (Kang, Dimitrova et al. 2005). Hence, in the human ClpP heptamer, the accessible catalytic sites are rendered inactive to prevent uncontrolled proteolysis until the ClpP cylinder is properly formed and the active sites are sequestered from the environment upon binding of ClpX and the formation of ClpXP complexes. Such a paradigm of functional regulation has been observed for ClpPs in other organisms such as *Bacillus subtilis* (Kirstein, Schlothauer et al. 2006).
Surprisingly, it was found that human ClpP can form a complex with *E. coli* ClpX *in vitro*. The structure of the heterogeneous complex as visualized by electron microscopy was highly similar to that of the *E. coli* ClpXP complex. However, the inverse hetero-complex of human ClpX and *E. coli* ClpP could not be formed. Also, the *E. coli* ClpA ATPase has no affinity to human ClpP. It is known that human ClpXP does not recognize the same substrates as the *E. coli* ClpXP system. Interestingly, however, when human ClpP is associated with *E. coli* ClpX, the heterogeneous system can degrade *E. coli* ClpXP substrates. These data further confirm that substrate recognition is only dependent on the chaperone component of the system.
Figure 6. The structures of the ClpP subunit.

(A) Secondary structure elements of *E. coli* ClpP monomer are highlighted, the axial loops, head domain, and handle region are also indicated. Residues of the catalytic triad are shown as dotted spheres. (B) – (F) Overlay of the monomers from *S. pneumoniae* (1Y7O), *M. tuberculosis* (2CE3), *P. falciparum* (2F6I), and human (1TG6) ClpPs with that from *E. coli* ClpP (1YG6). The overlays were done using PyMOL. The *S. pneumonia* ClpP has the mutation A153P in the E helix.
1.5.3.3.3 Plant ClpPs

The diversity and complexity of the Clp family is evident in higher plant organisms (Peltier, Ripoll et al. 2004). There are at least 10 ClpP-like proteins identified in the model plant Arabidopsis thaliana (Sjogren, Stanne et al. 2006) (Table 1). There are 6 ClpP paralogs (ClpP1-6) and 4 ClpR paralogs (ClpR1-4). Phylogenetic studies indicate that some ClpR proteins (ClpR1, 3, 4) of Arabidopsis thaliana may have evolved from the cyanobacterial ClpR (Peltier, Ytterberg et al. 2001). In addition, there are 10 Clp/Hsp100 AAA+ chaperones found in A. thaliana including 7 Class I (ClpB1-4, ClpC1-2, ClpD) and 3 Class II (ClpX1-3) which are associated with ClpPR complexes (Peltier, Ripoll et al. 2004). ClpT is an ortholog of the bacterial ClpS, which is a cofactor that binds to ClpA and affects substrate specificity in E. coli (Erbse, Schmidt et al. 2006). A new group of plant Clp chaperones with unknown function were discovered and also named as ClpS (ClpS1, ClpS2), but they have no similarity with bacterial ClpS and should not be confused with them. Although the function of plant ClpS1-2 is not clear, their sequences are homologous to the N-terminus of the chaperone ClpC but lack the AAA+ domains (Peltier, Ytterberg et al. 2001), which led to the suggestion that they may bind to the apical surface of ClpPR core complex, preventing association with the ATPases (Peltier, Ripoll et al. 2004).

Most Clp ATPases and proteases in Arabidopsis are found in the chloroplast stroma including ClpB3, ClpC1-2, ClpD, ClpP1, ClpP3-6, ClpR1-4, ClpS1-2, and ClpT. All of them are encoded in the nuclear genome, except for ClpP1 which is plastid-encoded (Peltier, Ripoll et al. 2004; Sjogren, Stanne et al. 2006). The complexity of Clp proteins was also seen in other plants. For example, all Clp proteins except for ClpT were also identified in the plastids of non-green plants Brassica rapa roots and Brassica oleracea petals by mass spectrometry (Peltier, Ytterberg et al. 2001; Peltier, Ripoll et al. 2004).
Figure 7. Alignment of ClpP and ClpR sequences.

Sequence alignments of ClpP from *E. coli*, *S. pneumoniae*, *M. tuberculosis*, *P. falciparum*, and human with ClpR from *Synechocystis PCC6803* and ClpR3 from *A. thaliana*. The alignment was generated using ClustalW and the Blosum62 matrix [49]. Identical residues in all seven sequences are highlighted in green, while identical residues in the five ClpP sequences are highlighted in yellow. The catalytic triad residues are indicated by blue dots. The boundaries for the axial loop and handle region and the numbering of the helices and strands refer to that of *E. coli* ClpP [34] and [37].
The ClpP oligomer in the chloroplast of *Arabidopsis* has been identified as composed of the subunits ClpP1, ClpP3-6, ClpR1-4 and ClpS1-2, forming a complex of 325-350 kDa (Peltier, Ripoll et al. 2004). Modeling studies indicate that ClpP and ClpR contribute to formation of the tetradecamer, whereas ClpS1-2 do not fit in the ring structure but rather seem to be associated with hydrophobic pockets on the axial sites on top of the ClpPR core complex (Peltier, Ripoll et al. 2004). In the mitochondria, the ClpP complex is a homotetradecamer of ClpP2 subunits (Peltier, Ripoll et al. 2004; Sjogren, Stanne et al. 2006). From sequence alignments, ClpP2 is found to be the most closely related to human mitochondrial ClpP, and, hence mitochondrial plant ClpP2 possibly forms a complex with plant ClpX which is also targeted to the plant mitochondria (Halperin, Zheng et al. 2001).

ClpS1-2 are only found in land plants but not in prokaryotes or green algae, which suggests that they probably have special functional roles in proteolysis in higher plants. Binding of ClpS1-2 to the axial site of ClpPR core is likely to inhibit the binding of the cognate ATPase such as ClpC or ClpD. Therefore ClpS1-2 might regulate degradation by competitive binding with the ATPase chaperones (Majeran, Friso et al. 2005). The heterocomplex of ClpPR core is also found in the green alga *Chlamydomonas reinhardtii* stroma, but ClpS1-2 do not exist in *C. reinhardtii* genome (Peltier, Ripoll et al. 2004). Instead, a 30-kDa insertion sequence (IS1) is found in ClpP1 of *C. reinhardtii*, which is proposed to have the same effect as ClpS1-2. The insertion domain located between helix 2 and strand 2 (Figure 6) should protrude over the apical surface thus hindering ATPase binding (Majeran, Friso et al. 2005).

The inactive ClpR is thought to have a regulatory role in proteolysis. ClpRs in *A. thaliana* have extended C-termini compared to *E. coli* ClpP (Figure 7). The C-terminal extension in ClpR might fold on top of the proteolytic core and, hence, might control the interaction between the core and its chaperones. An insertion loop of 9-10 residues in ClpR1, ClpR3, and ClpR4 is found when aligned with *A. thaliana* ClpPs. According to homology modeling, the insertion loop is part of the substrate-binding cleft in the active site, thus this loop of the non-catalytic ClpR proteins might participate in presenting substrates to the catalytic triads of active ClpP neighbours (Peltier, Ripoll et al. 2004).

The cellular functions of most ClpPs in plants are not yet known, but all ClpP paralogs are essential for chloroplast development. In *C. reinhardtii*, ClpP1 is responsible for degrading
fully or incompletely assembled cytochrome $b_{6}f$ and for high CO$_2$ tolerance (Majeran, Wollman et al. 2000). ClpP1 is also vital for plastid development and plant viability in tobacco (Nicotiana tabacum) (Shikanai, Shimizu et al. 2001; Kuroda and Maliga 2002). In Arabidopsis, ClpR2 is essential for Clp core complex assembly since reducing the levels of ClpR2 resulted in reduction of the ClpPRS protease complex levels, which led to a pale-green phenotype, delayed shoot development, reduced chloroplast size, decreased thylakoid accumulation, and increased plastoglobule (lipoproteins in chloroplasts) levels (Rudella, Friso et al. 2006).

In summary, the plant Clp system is much more complicated than in other organisms. This might be a reflection of the multiple severe stresses that stationary organisms encounter. Studying the plant Clps is a new and exciting area of research that is still at its infancy and needs to be further explored.
Figure 8. Model of ClpP mechanism of function.

Shown is a cartoon model of the ClpP tetradecamer with the axial loops indicated. For simplicity, the chaperones bound to ClpP are drawn as simple ellipses. In the proposed model, substrate proteins are unfolded and then translocated into the ClpP cylinder by the bound chaperone. Unfolded polypeptides enter the ClpP proteolytic chamber through the axial pores which are lined by the axial loops. The polypeptide chains are then degraded into small peptides. The peptides are proposed to be released through axial pores that are transiently formed as a result of the dynamics in the handle region of the ClpP subunits.
1.5.3.3.4 Yeast ClpPs

In yeast, ClpP homologs are found in some but not all strains. There are no ClpP sequences identified in the common laboratory strains of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Table 1). However, *Gibberella zeae*, *Neurospora crassa*, and *Aspergillus fumigatus*, among others, contain ClpP-like proteins that typically have sequences shorter than that of *E. coli* ClpP. Multiple sequence alignment reveals that the yeast ClpPs have about 50 – 70 residues deleted from their N-termini when compared to *E. coli* ClpP. Little is known about the cellular function of these yeast ClpPs or, indeed, whether these shorter ClpPs are active.

1.5.3.4 ClpP structure and function

1.5.3.4.1 ClpP X-ray structure

To date, ClpP structures have been solved from 5 different organisms (Figure 5) including *E. coli* (Wang, Hartling et al. 1997; Bewley, Graziano et al. 2006; Szyk and Maurizi 2006), *Streptococcus pneumoniae* (Gribun, Kimber et al. 2005), the malaria agent *Plasmodium falciparum* (Vedadi, Lew et al. 2007), *Mycobacterium tuberculosis* (Ingvarsson, Mate et al. 2007), and human (Kang, Maurizi et al. 2004). The sequence alignment of these different ClpPs is given in Figure 7. As expected, the overall ClpP structures from these organisms are very similar (Figure 5 and Figure 6).

The X-ray structure of ClpP from *E. coli* was the first to be solved (1TYF, not shown in Figure 5) (Wang, Hartling et al. 1997); it shows a cylindrical-shaped tetradecamer of about 300 kDa in molecular weight and 90 Å in both height and diameter (Wang, Hartling et al. 1997). The monomer (Figure 6) is mainly composed of six repeats of α/β fold (αA/β1/β2, αB/β3/β4, αC/β5/β6, αD/β7/β8, αF/β10, and αG/β11) with an additional protruding α/β unit (αE/β9). The subunits are held together mainly by hydrophobic interactions. Each monomer can be divided into a handle region (β strand 9 and E helix), which mediates ring-ring interaction, and a head domain comprised of residues 28-120 and 160-188 (Wang, Hartling et al. 1997). A heptameric ring is formed from the packing of head domains, exposing 7 handle regions, which intercalate with 7 handle regions from the opposite heptameric ring, forming a tetradecamer with a spherical
internal chamber of about 51 Å in diameter (Wang, Hartling et al. 1998). Although according to the X-ray structure, the handle region is the only area where two ClpP rings have contact, surprisingly, truncations in this region do not lead to the dissociation of the 2 rings, indicating a high degree of plasticity of this region (Gribun, Kimber et al. 2005). Charge-charge interaction networks involving residues in the head domain seem to contribute to stabilizing the interactions between the two ClpP heptameric rings (Maurizi, Singh et al. 1998; Gribun, Kimber et al. 2005). The two rings are found to dissociate under high sulfate concentration at low temperature (Maurizi, Singh et al. 1998). A mutant ClpP(A153P) from *S. pneumoniae* was solved at 2.5 Å resolution, and the major difference in the structure of this mutant from the first solved *E. coli* ClpP was in the N-terminal region (Figure 6). In *S. pneumoniae* ClpP(A153P), it was observed that the N-terminal residues Met16-Ser30 (using *E. coli* SwissProt numbering, Figure 7) form an extended loop protruding from the body of the ClpP cylinder. These axial loops were not detected in the original *E. coli* ClpP structure due to the weak electron density in that region; however, recently solved X-ray structures of *E. coli* ClpP [e.g. 2FZS (Szyk and Maurizi 2006) and 1YG6 (Bewley, Graziano et al. 2006), shown in Figure 5] and human ClpP (1TG6) (Kang, Maurizi et al. 2004) clearly show the presence of these axial loops. These axial loops are an important feature of the ClpP structure and have been shown to mediate the interaction of ClpP with its cognate ATPases (Gribun, Kimber et al. 2005).

The mutation of Ala153 to Pro in the *S. pneumoniae* ClpP(A153P) structure causes disorder in the handle domain and renders the E helix shorter by two turns as compared to the E helix in wild type *E. coli* ClpP (Figure 6) (Gribun, Kimber et al. 2005). The fact that ClpP remains tetradecameric under these conditions is an explicit demonstration of the plasticity of the handle region. This region is also found to be disordered in the crystal structure of *P. falciparum* and *M. tuberculosis* ClpP. Based on biochemical and NMR-based studies (Gribun, Kimber et al. 2005; Sprangers, Gribun et al. 2005), we had proposed that this region of ClpP is highly dynamic and that the movement of the E helices results in the transient formation of equatorial side pores that allow for the exit of peptide fragments generated from the degradation of polypeptide chains inside the ClpP chamber (Figure 8). This issue is further discussed below.

Very little is known about the *P. falciparum* ClpP. However, the sequence of this protease has about 150 extra amino acids at the N-terminus of the protein when compared to the sequence of the *E. coli* ClpP. It is predicted that this extra N-terminal sequence targets the
nuclear encoded protease to the apicoplast organelle present in plasmodium, although this has not been experimentally established. The apicoplast is an organelle that is homologous to chloroplasts of plants and is found in apicomplexan parasites such as *P. falciparum* (Waller and McFadden 2005). The apicoplast is an ancient feature of this group of organisms and is thought to have been acquired by the process of endosymbiosis. It contains proteins that are nuclear encoded and then transported to this organelle and also proteins that are encoded and expressed by its own genome and expression machinery.

The structure of *M. tuberculosis* ClpP1 has two unique features (Ingvarsson, Mate et al. 2007). First, there is an extended αA helix at the N-terminus (Figure 6) which renders the axial pores smaller compared to those in other ClpPs. Second, the monomers are tilted inwards compared to other ClpPs. The rotated orientation of the monomers results in equatorial side pores, which were suggested to transiently form in ClpP (Gribun, Kimber et al. 2005; Sprangers, Gribun et al. 2005). This rotated orientation of the monomers may assist in substrate entry through the axial pores, or product release from the equatorial pores (Ingvarsson, Mate et al. 2007). As for human ClpP, *M. tuberculosis* ClpP1 is also found to form heptamers under normal conditions (Ingvarsson, Mate et al. 2007), although in the crystal it formed a tetradecamer. It should be noted, however, that the structure of *M. tuberculosis* ClpP1 might not be the predominant physiological assembly since there are two ClpPs in *Mycobacterium tuberculosis* (ClpP1 and ClpP2, Table 1) which seem to be on the same operon and might form a hetero-oligomer.

The unique feature in the structure of human ClpP is the presence of an extended C-terminus with 28 additional residues located on the periphery of the heptamer, forming a flexible loop which extends out of the surface of the oligomer. The loop is unstructured and is not observed in the solved X-ray structure of human ClpP (Figure 5 and Figure 6). This C-terminal extension is found to affect the assembly of human ClpP heptamer since deletion of this C-terminus resulted in structurally unstable ClpP (Kang, Maurizi et al. 2004). However, the deletion also resulted in an increased affinity of human ClpP to human ClpX, suggesting that the C-terminus might hinder the interaction between ClpX and ClpP (Kang, Maurizi et al. 2004).

Based on these multiple ClpP structures from different organisms, it can be seen that the ClpP cylinder maintains the same overall assembly and construction. Modifications are then
added to the ‘core’ ClpP structure to engineer specific functions that are suitable for the varied cellular environments of the different organisms.

1.5.3.4.2 ClpP mechanism of function

ClpP is a classical serine protease whereby each subunit in the ClpP homotetradecamer has an active site consisting of the three canonical residues: Ser, His, and Asp (Figure 6A and Figure 7). Hence, the ClpP homotetradecamer contains 14 active sites within its proteolytic chamber. The catalytic triad is located at the junction between the head and handle domains in each ClpP monomer (Figure 6A). For the case of the hetero ClpPR complexes in plants and cyanobacteria, the number of active sites would be reduced. It has been shown that mutations in the handle region are often found to decrease ClpP activity due to the close proximity of the active site to the handle region, resulting in the disruption of the catalytic triad alignment (Sprangers, Gribun et al. 2005). Systematic studies on ClpP substrate cleaving specificity have not yet been published, however, according to one report (Szyk and Maurizi 2006), *E. coli* ClpP was found to preferentially cleave after charged and branched-chain amino acids. There are indications that the catalytic residues undergo a rearrangement to an active configuration upon substrate binding (Szyk and Maurizi 2006).

The only access to the catalytic chamber is through the narrow axial pores which allow entry of small peptides of about 30 residues in length (Thompson, Singh et al. 1994). The N-terminal axial loops contribute to the axial pores. The axial loops can be divided into two parts (Gribun, Kimber et al. 2005; Bewley, Graziano et al. 2006; Szyk and Maurizi 2006). There are 7 to 8 hydrophobic residues, termed the axial pore lining, that line the axial pores followed by about 9 residues, the axial protrusion, that protrude from the apical surface of the ClpP cylinder (Figure 6). These axial loops are required for the interaction of ClpP with its cognate ATPases (Gribun, Kimber et al. 2005; Bewley, Graziano et al. 2006). Mutants with the first seven N-terminal residues deleted from mature ClpP were shown to be unable to degrade folded protein substrates in the presence of the ATPases, but exhibited higher peptidase activity, suggesting that the truncations did not inactivate the protease but had disrupted its interaction with the chaperones (Gribun, Kimber et al. 2005). The increased peptidase activity is probably due to the enlargement of the axial pores as a result of the deletion of the N-terminal residues.
On the apical surface of ClpP, about 54 Å away from the axial pores, there are 7 grooves of about 10 Å in diameter. These grooves are mainly composed of conserved hydrophobic residues and are thought to provide binding pockets for specific loop regions present in the structure of the chaperones. These loop regions are highly conserved in AAA+ chaperones that bind to ClpP (Kim and Kim 2003) such as the ‘IGF’ loop in *E. coli* ClpX or ‘IGL’ loop in *E. coli* ClpA. Hence, the chaperone-protease interaction is mainly mediated by loop-groove interactions rather than by interactions between large surfaces. This mode of interaction might circumvent the symmetry mismatch that exists between ClpP, which has a seven fold symmetry, and its interacting chaperones, which typically have six fold symmetry. Furthermore, the loop-groove interaction would provide the chaperone-protease complex with more structural flexibility that might be necessary for the complex to achieve substrate protein unfolding, translocation, protein degradation, and peptide product release.

ClpP processively degrades substrates and generates peptides of average length of about 7 to 8 residues (Thompson, Singh et al. 1994; Choi and Licht 2005). The mechanisms by which unfolded polypeptides enter ClpP and how the degraded products are released from the protease are still poorly understood. Crystallographic data (Wang, Hartling et al. 1997; Szyk and Maurizi 2006) indicate that the substrate protein is probably captured close to the active site by a hydrogen bonding network involving residues in β4 and β9 and that the interaction is further stabilized by hydrophobic interactions with the wall of the ClpP chamber. It is proposed that the substrate will be horizontally positioned in the proteolytic chamber near the equator of the ClpP cylinder and oriented in a clockwise N to C direction when viewed from the axial channel through which the substrate entered (Szyk and Maurizi 2006). This would be in agreement with a C to N translocation of substrates into the ClpP proteolytic chamber.

Different hypotheses exist about the mechanism of product release. One model proposes that the generated peptides exit the ClpP proteolytic chamber by passive diffusion through the same axial pores used for substrate entry (Thompson and Maurizi 1994; Wang, Hartling et al. 1997; Kim, Burton et al. 2000). However, this implies that the bound chaperones have to dissociate from the ClpP cylinder to allow for peptide exit. Such a mechanism might be inefficient and would probably render the degradation of polypeptide chains nonprocessive. This mechanism has been suggested for the proteasome in which the dissociation of the ATPase units has been reported to coincide with product release (Babbitt, Kiss et al. 2005). The second model
suggests that peptides are released through side pores that are transiently formed at the interface between the two heptameric ClpP rings. It has been established that the handle region exhibits high plasticity (Gribun, Kimber et al. 2005), and NMR studies further confirmed that the E helices exchange between at least two structurally distinct conformations, suggesting the existence of transiently formed dynamic side pores which could act as exit pores for degraded products (Sprangers, Gribun et al. 2005). Biochemical data also confirmed this model for peptide exit. An Ala residue at position 153 on the E helix in *E. coli* ClpP was mutated to Cys, A153C, allowing the formation of a disulfide bond between E helices of two subunits from the two opposite heptameric ClpP rings. The cross-linking of the two rings inhibited movements of the handle regions. A small dipeptide was able to diffuse into the chamber of disulfide-linked inactive ClpP but was trapped inside the chamber under oxidizing condition. The trapped substrate was shown to be released after the system was placed in reducing conditions. The results strongly suggested that the equatorial regions of the ClpP barrel provide the exit sites of degraded products (Sprangers, Gribun et al. 2005). The observation that *M. tuberculosis* ClpP1 and human ClpP exist mainly as a single heptameric rings (Kang, Dimitrova et al. 2005; Ingvarsson, Mate et al. 2007) and that human ClpP forms the double ring structure only upon binding the chaperone ClpX (Kang, Dimitrova et al. 2005), seems to further support the model of peptide exit model through transiently formed equatorial side pores.

Understanding how the ClpP system functions has direct implications on understanding the mechanism of function of other more complex cylindrical proteases such as the proteasome. ClpP shares many similarities with the proteasome both functionally and structurally. The arrangement of subunits in ClpP is strikingly analogous to that of eukaryotic and archaeal 20S proteasomes, which is composed of 2 heptameric β-subunit rings enclosing the active site sandwiched between 2 heptameric rings of α subunits. Another protease HslV (ClpQ), also known as the ‘bacterial proteasome’, exhibits analogous packing except that the cylinder is composed of two hexameric rings. The similarity in the cylindrical structural arrangement probably underlines the commonality in the mechanism of function of these proteases. Efforts currently underway in several laboratories should further enhance our understanding of the mechanism of function of these degradative machines. These ongoing efforts might have direct clinical consequences as ClpP has now been shown to be a legitimate antibacterial drug target.
1.6 Thesis rationale

TF chaperone and ClpXP degradation system are important molecular machineries that maintain protein homeostasis. TF assists in the folding of polypeptide chains co-translationally and ClpXP protease regulates protein levels through degradation. It is generally thought that there is no functional association between folding chaperones and protein degradation systems since they belong to parallel cellular pathways: folding to the native state versus degradation. However, as part of our general interest in mapping chaperone interaction networks, we found that tig (gene for trigger factor), clpX, and clpP genes colocalize next to each in most examined bacteria. This finding was the first clue that led to our hypothesis that TF and ClpXP might be functionally associated. In Chapter 2, I demonstrated that TF physically interacts with ClpX, and they functionally collaborate to enhance degradation of certain ClpXP substrates. We estimated that TF enhances the degradation of about 2% of newly synthesized E. coli proteins. Phage replication protein λO is degraded faster in the presence of TF both in vitro and in vivo. TF was found to interact with both ClpX and λO with similar affinities and to enhance the ClpXP-mediated degradation of ribosome-stalled λO nascent chains. Experiments suggest that TF transfers ribosome-stalled λO to ClpX for degradation by ClpP protease, therefore, demonstrating the existence of co-translational protein degradation in E. coli.

The study of TF and ClpXP interaction is my main research goal, and I also collaborated with Dr. Kimber (University of Guelph, Canada) and Dr. Borg (Bioinformatics Infrastructure for Life Sciences, Sweden) to study the mechanism of peptide release from ClpXP system. The pathway of substrate delivery is well established, but the pathway of product release remains unclear. Different hypotheses exist about the mechanism of product release. One hypothesis proposes that the degraded peptides exit the ClpP proteolytic chamber by passive diffusion through the same axial pores used for substrate entry. This implies that ClpX has to dissociate from ClpP to allow for peptide exit, which might be an inefficient mechanism. Another hypothesis, proposed by our laboratory (Gribun, Kimber et al. 2005; Sprangers, Gribun et al. 2005), suggests that peptides are released through side pores that are transiently formed at the interface between the two heptameric ClpP rings. In Chapter 3, we showed that the crystal structure of the mutant E. coli ClpP (A153C) solved by Dr. Kimber resembled a compact conformation compared to the wildtype E. coli ClpP. This compact structure was also observed in S. pneumonia, M. tuberculosis, and P. falciparum ClpPs while WT E. coli and human ClpP
structures showed an extended form. We hypothesized that ClpP changes conformations to allow the protease to switch between degradation to product release. By normal mode analysis, we showed that ClpP exhibited dynamics that allow WT ClpP to switch between extended states to compact states. I contributed to this study by characterizing the mutant ClpP (A153C) and comparing it with the WT ClpP both biochemically and biophysically.
2 Functional cooperativity between trigger factor and ClpXP in degradation

Data Attribution: I performed the majority of the experiments in this chapter with the following exceptions. The initial bioinformatics analysis was done by Dr. Elisabeth Tillier; the data were further confirmed by Ms. Sarah Kim and me (Figure 9). The degradation of FITC-λO by ClpXP in the presence TF was monitored by Dr. Guillaume Thibault (Figure 10C). The degradation assays of S7, Fnr and TF by ClpXP were done by Mr. Adedeji Ologbenla (Figure 11B,C). The binding experiments of TF domains were performed by Ms. Sarah Kim, Mr. Adedeji Ologbenla and me (Figure 12C and Figure 13). The ClpX ATPase activity in the presence of TF domains was measured by Ms. Sarah Kim and me (Figure 14C). The degradation assays of λO by ClpXP in the presence of TF domains were performed by Mr. Adedeji Ologbenla and me (Figure 15). TF domains were cloned by Ms. Elisa Leung.
2.1 Summary

Co-translational protein degradation is well established in eukaryotes, but it remains unclear whether prokaryotes also possess such a system for degrading nascent chains as they emerge from the ribosome. Here we show that the ribosome-associated chaperone trigger factor (TF), which promotes co-translational protein folding, physically interacts with the ClpX unfoldase of the ClpXP proteolytic system. This reveals a new role for TF in facilitating the degradation of newly translated proteins in *Escherichia coli*. It is estimated that TF enhances the degradation of about 2% of newly synthesized *E. coli* proteins. One of the ClpXP substrates whose degradation was enhanced by TF was λO, the λ phage replication protein. TF was found to interact with both ClpX and λO with similar affinities and to enhance the ClpXP-mediated degradation of ribosome-stalled λO nascent chains. Experiments suggest that TF transfers λO to ClpX for eventual degradation by the ClpP serine protease, therefore, demonstrating the existence of co-translational protein degradation in *E. coli*.
2.2 Introduction

In the very crowded environment of the cell, molecular chaperones and proteases function to ensure proper protein homeostasis either by promoting folding of newly synthesized proteins and maintaining the correct conformation of pre-existing proteins or by degrading misfolded and unfolded proteins, respectively. Generally, some chaperones such as trigger factor (TF) and the Hsp70/Hsp40 system (DnaK/DnaJ) act early during the folding process, while other chaperones such as GroEL/GroES, TRiC, and Hsp90 act at later stages of folding (Hartl and Hayer-Hartl 2009). In addition, there is redundancy in the function of folding chaperones such as that found for TF and DnaK in folding newly synthesized proteins (Deuerling, Schulze-Specking et al. 1999; Teter, Houry et al. 1999; Deuerling, Patzelt et al. 2003). It is well established that several chaperone systems can mediate the co-translational folding of newly synthesized nascent chains as they emerge from the ribosome in both prokaryotes and eukaryotes. Such chaperones include the Nascent chain Associated Complex (NAC) (Rospert, Dubaquie et al. 2002; del Alamo, Hogan et al. 2011), the Ribosome-Associated Complex with the Hsp70 homolog Ssb (RAC-Ssb complex) (Gautschi, Lilie et al. 2001; Koplin, Preissler et al. 2010; Chiabudini, Conz et al. 2012; Leidig, Bange et al. 2013) in eukaryotes and TF in prokaryotes (Preissler and Deuerling 2012).

Co-translational degradation is well established in eukaryotes (Schubert, Anton et al. 2000; Turner and Varshavsky 2000). It is thought that both stable and unstable proteins are susceptible to co-translational degradation by the proteasome (Turner and Varshavsky 2000) and that these proteins are ubiquitinated while still on the ribosome (Sato, Ward et al. 1998). Non-stop proteins which escape translational repression have been shown to be removed via co-translational degradation (Chiabudini, Conz et al. 2012). RAC-Ssb and two ribosome-bound E3 ligases, Not4 and Ltn1 in yeast, are involved in this process. In another study, a Ribosome Quality Control Complex (RQC), comprising the Ltn1 E3 ubiquitin ligase, two other proteins Tae2 and Rqc1, and Cdc48 and its cofactors, has been identified to associate with 60S ribosomal subunits containing stalled polypeptides to trigger their degradation (Brandman, Stewart-Ornstein et al. 2012). Hence, in eukaryotes, multiple mechanisms and complexes seem to be involved in facilitating co-translational protein degradation, but such a mechanism has not been identified for prokaryotes.

_E. coli_ TF is a 48 kDa protein containing an N-terminal ribosome binding domain (TF_N).
a central peptidyl-prolyl cis/trans isomerase (PPIase) domain (TF$_p$), and a C-terminal domain with two arm motifs (TF$_c$) (Ferbitz, Maier et al. 2004; Martinez-Hackert and Hendrickson 2009). TF binds to ribosomes in a 1:1 stoichiometry through the ribosomal protein L23 (Kramer, Rauch et al. 2002), which is located at the peptide exit tunnel of the 50S large ribosomal subunit, allowing TF to interact with nascent chains that are emerging from the ribosomal exit tunnel (Lill, Crooke et al. 1988; Valent, Kendall et al. 1995; Hesterkamp and Bukau ; Beck, Wu et al. 2000; Patzelt, Kramer et al. 2002). The $K_d$ for TF binding to non-translating ribosomes is approximately 1.2 µM (Patzelt, Kramer et al. 2002), but the affinity increases up to 30-fold in the presence of nascent chains (Rutkowska, Mayer et al. 2008).

The ATP-dependent caspase-lytic protease complex, ClpXP, is one of the main degradation systems in bacteria. It plays essential roles in protein quality control by removing misfolded, damaged, and regulatory proteins. The proteolytic complex consists of the tetradecameric serine protease ClpP (Yu and Houry 2007) and the hexameric ClpX unfoldase chaperone of the AAA+ superfamily (Neuwald, Aravind et al. 1999; Snider, Thibault et al. 2008). ClpX has two domains, an N-terminal zinc binding domain (ZBD) and a C-terminal AAA+ domain that contains the characteristic Walker A and Walker B nucleotide binding and recognition motifs (Neuwald, Aravind et al. 1999; Dougan, Mogk et al. 2002; Iyer, Leipe et al. 2004; Ammelburg, Frickey et al. 2006; Snider, Thibault et al. 2008). ClpX binds to target substrates and uses energy released from ATP hydrolysis to unfold and transfer the substrates to ClpP. Entrance into the chamber occurs through the two axial pores located on opposite sides of the ClpP cylinder. The proteolytic chamber contains 14 Ser-His-Asp catalytic triads from each protomer. Substrates are generally cleaved into peptides of about 7 to 8 residues (Choi and Licht 2005; Jennings, Lun et al. 2008). We have proposed that these degraded products exit the chamber through equatorial side pores that form due to the dynamics in the ClpP structure (Gribun, Kimber et al. 2005; Sprangers, Gribun et al. 2005; Kimber, Yu et al. 2010; Geiger, Bottcher et al. 2011; Lee, Kim et al. 2011; Zhang, Ye et al. 2011; Gersch, List et al. 2012).

In mapping chaperone interaction networks, we were intrigued to find that tig (gene for trigger factor), clpX, and clpP genes colocalize next to each other in most examined bacteria. Genes that are clustered together on the genome tend to have similar functions, are targeted to the same pathways, or are expressed in the same compartment of the cell (Zampieri, Soranzo et al. 2008). Furthermore, the evolutionary conservation of such clustering is strong evidence for
functional association (Overbeek, Fonstein et al. 1999). The close genomic proximity of *tig* and *clpXP* in many bacteria was the first clue that indicated a potential link between the folding and degradation pathways in prokaryotes. Here we show that TF enhances ClpX ATPase activity, and confirm a physical interaction between TF and ClpX. Unexpectedly, TF was found to enhance the degradation of the λO substrate by ClpXP both *in vitro* and *in vivo*. More significantly, TF was found to enhance the ClpXP-mediated degradation of ribosome-stalled λO-SecM substrate. Our results indicate that the functional association of ribosome-bound TF and ClpXP leads to co-translational protein degradation in *E. coli*. 
2.3 Experimental Procedures

Bioinformatics analysis

Gene neighbourhoods of \textit{tig}, \textit{clpP} and \textit{clpX} in bacteria were downloaded from the Protein Clusters Database (Klimke, Agarwala et al. 2009) at NCBI using the ProtMap tool. The data were assembled with custom Perl scripts. The genome organization of the genes on the bacterial taxonomy was displayed using iTol (Yamada, Letunic et al. 2011). The presence and genome organization of \textit{tig}, \textit{clpP} and \textit{clpX} genes in the different bacteria was then manually confirmed using the ExPASy protein knowledgebase and the NCBI Protein Cluster Database.

Bacterial strains and plasmids

Gene deletions were performed using the \lambda-red recombination system (Datsenko and Wanner 2000) followed by P1 transduction into MC4100 strain using P1 phage as needed (Ikeda and Tomizawa 1965). pBAT-His-Sumo-Barnase-SecM plasmid was a kind gift from Dr. Bernd Bukau (University of Heidelberg, Germany). Barnase was removed and \lambda O was inserted between the AflIII and BamH1 sites of pBAT plasmid to generate pBAT-His-Sumo-\lambda O-SecM. To generate a construct without His-Sumo, the sequence of \lambda O-SecM was PCR amplified from pBAT-His-Sumo-\lambda O-SecM and inserted into pET15b plasmid between the NcoI and HindIII cut sites. For endogenous \lambda O protein turnover experiments, WT, \textit{tig::cat}, and \textit{clpX::cat} MC4100 strains were lysogenized with \lambda cI857Sam7 phage, which was provided by Dr. Alan Davidson (University of Toronto, Canada).

S7 was cloned into p11 vector with an N-terminal His tag followed by a TEV cleavage site for purification of S7 protein. To monitor S7 degradation \textit{in vivo}, untagged S7 was cloned into pET3a vector under the control of the T7 promoter and transformed into WT, \textit{tig::cat}, and \textit{clpX::cat} MC4100 strains. To express S7 in MC4100 strains, the pT7pol26 plasmid expressing T7 RNA polymerase (Mertens, Remaut et al. 1995) was also transformed into these strains. Plasmid pPK823 (pET11a) overexpressing Fnr was from Dr. Patricia K. Kiley (University of Wisconsin, USA). Plasmid pUHE21-2 expressing the FlhDC complex was from Dr. Tomoko Yamamoto (Chiba University, Japan) (Takaya, Erhardt et al. 2012). Plasmid pHF010 expressing \lambda N-His protein was from Dr. Irene Lee (Case Western Reserve University, USA).
**Protein purification**

All recombinant proteins were overexpressed in *E. coli* BL21(DE3) cells. The plasmid pProEXHTa TF expressing trigger factor was from Dr. José M. Barral (University of Texas, USA). TF was induced at midlog phase with 1.0 mM IPTG for 3.5 hours at 37 °C. His-tagged TF was purified on Ni-NTA column (Qiagen) and washed with 50 mM imidazole in buffer A (25 mM TrisHCl, pH 7.5, 300 mM NaCl, 10% glycerol, and 1 mM DTT) to remove ATPase contaminants before elution. Eluted TF fractions were tested and only the ones without ATPase activity were collected. The His-tag was removed using His-tagged Tobacco Etch Virus (TEV) protease during dialysis in buffer B (25 mM TrisHCl, pH 7.5, 200 mM KCl, 10% glycerol, and 1 mM DTT). Purification of TF was performed in one day and the protein was flash frozen in liquid nitrogen after removing His-tagged TEV on a Ni column.

ClpP (Flanagan, Wall et al. 1995), ClpA (Leung, Datti et al. 2011), λO (Roberts, Anzano et al. 1983), λN (Patterson-Ward, Tedesco et al. 2009), MuA (Baker, Mizuuchi et al. 1993), S7 (Wimberly, White et al. 1997), and Fnr (Lazazzera, Bates et al. 1993) were expressed and purified according to published protocols. HtpG was precipitated in 70% ammonium sulfate and further purified using an anion exchange Sepharose Q column followed by gel filtration. FlhDC complex proteins were purified as described (Takaya, Erhardt et al. 2012). Casein-FITC was purchased from Sigma (product number C3777). BSA was purchased from New England Biolabs.

**Degradation assays**

Degradation assays were carried out in the presence and absence of TF. Degradation reaction mixture contained 1.2 μM ClpP, 3.9 μM substrate, 0 or 5.0 μM TF, and an ATP regeneration system (13 units/mL of creatine kinase and 16 mM creatine phosphate) in PD buffer (25 mM HEPES, pH 7.5, 5 mM MgCl2, 5 mM KCl, 0.03% (w/v) Tween 20, and 10% glycerol). All concentrations are those of protomers. Components were incubated at 37 °C for 3 minutes before adding 1.0 μM ClpX to start the reaction. At given time points, aliquots were withdrawn and mixed with 4× Laemmli buffer to stop the reaction. Proteins were then resolved on SDS-PAGE gels and visualized by Coomassie blue staining. Protein bands on the gels were quantified using Quantity One (BioRad). Anti-MuA antisera were kindly provided by Dr. George Chaconas.
GFP-SsrA degradation was monitored by fluorescence using a Fluorolog spectrofluorometer (JobinYvon) with excitation wavelength set at 395 nm and emission wavelength set at 509 nm. Degradation assays using ClpAP were performed using the same concentration of the different components with a high salt ClpA buffer (25 mM HEPES, pH 7.5, 20 mM MgCl₂, 300 mM KCl, 0.03% (w/v) Tween 20, and 10% glycerol).

**FITC-λO labeling and degradation**

Fluorescein isothiocyanate (FITC) was used to label λO and monitor its degradation by fluorescence. Twenty-three mg of λO was conjugated with 2 mg of fluorescein-5-isothiocyanate (Invitrogen) by amine coupling (Fischer 2010) at pH 8.5 for 1 hour. FITC-λO was separated from free FITC on Superdex 200 10/300 GL column using buffer Y (50 mM TrisHCl, pH 7.5, 200 mM KCl, 25 mM MgCl₂, 0.1 mM EDTA, 10% glycerol, and 1 mM DTT). The degradation assay was carried out by preincubating 1.3 units of creatine kinase, 16 mM creatine phosphate, 3 mM ATP, 1 µM ClpX, 1.2 µM ClpP, 0-160 µM TF in PD buffer at 37 °C for 3 min. Subsequently, 4 µM FITC-λO was added and the protein degradation was monitored by fluorescence using a Fluorolog spectrofluorometer (JobinYvon) with the excitation wavelength at 494 nm and the emission wavelength at 521 nm.

**ATPase activity measurement**

The effect of TF on ClpX ATPase activity was measured using a coupled assay (Norby 1988) in a 96-well microtiter plate read using the Molecular Devices SpectraMax 340 PC 384 Microplate Reader. The reaction volume was 150 μL and assays were performed by preincubating 1.0 μM ClpX, various concentrations of TF, 0.2 mM NADH, 3.0 mM phosphoenolpyruvate, 4.7 units of pyruvate kinase, and 7.4 units of lactate dehydrogenase in PD buffer at 37 °C for 3 min. ClpP was added at 1.2 µM if needed. After preincubation for 3 minutes, 5 mM ATP was added to start the reaction and the change in absorbance at 340 nm was measured for 20 min at 37 °C. ATPase assays using ClpA were performed under the same concentration of the different components with a high salt ClpA buffer. ATP hydrolysis was calculated by measuring the disappearance of
NADH at 340 nm, assuming 1:1 correspondence between ADP formation and NADH oxidation. The extinction coefficient used for NADH is 6220 M$^{-1}$cm$^{-1}$ (Norby 1988). Each reaction was performed at least three times.

**Surface plasmon resonance (SPR) assays**

SPR measurements were carried out on a Biacore X from GE Healthcare; all experiments were done at room temperature. TF or TF domain was immobilized on CM5 chip (GE Healthcare) using the Biacore Amine Coupling kit. Analyte was injected at a flow rate of 20 μL/minute in running buffer (10 mM HEPES, pH 7.5, 150 mM NaCl, 3 mM EDTA, and 0.005% (w/v) P20 surfactant), and adequate injection time (90 seconds) was allowed for the interaction to reach steady state if possible. To regenerate the surface between binding experiments, 2 M NaCl was injected for 1 minute and washed with running buffer. From the results of the sensorgrams, response units (RU) at steady state were plotted against the corresponding analyte concentrations to generate the binding curve, and the $K_d$ was calculated using WinCurveFit program assuming one-site Langmuir binding model for most binding experiments. For the binding of MuA to TF, experimental sensorgram curves were fit to rate equations for one-site Langmuir binding model using the BiaEvaluation 4.1 software (GE Healthcare).

**Monitoring λO degradation in cells**

Plasmid pRLM266 expressing untagged λO under the T7 promoter was transformed into WT, tig::cat, and clpX::cat MC4100 strains. The strains were cotransformed with the pT7pol26 plasmid expressing the T7 polymerase (Mertens, Remaut et al. 1995). Cells were grown in LB broth at 37 °C until midlog phase (OD$_{600}$ = 0.6); IPTG was then added at a final concentration of 0.4 mM to induce λO expression for 2 hours. Tetracycline was added at a final concentration of 30 μg/mL to inhibit translation (time 0), and 100 µL of cells were withdrawn after addition of tetracycline at 0, 10, 13, 30, 45, 60, and 90 minutes. Cell density was measured using OD$_{600}$ for each time point. Cells were pelleted and resuspended in 1× SDS loading buffer with equal cell density. Whole cell lysates were subjected to Western blot analysis using rabbit antisera specific for λO, ClpX, ClpP, and TF proteins. About $10^5$ cells were loaded for each time point to monitor the degradation of λO proteins.
For monitoring λO-SecM degradation, pET15b λO-SecM was transformed into MC4100 strains also containing pT7pol26 plasmid, and the same procedures were followed as above except that 10 times more cells (10^6) were loaded for Western blot analysis to monitor λO-SecM proteins due to lower expression level of the fusion protein.

For ribosome fractionation, the experiments were carried out as described above. For each time point, 10 mL of cells were taken, pelleted, and resuspended in buffer R [10 mM TrisHCl, pH 7.5, 10 mM MgCl2, 60 mM NH4Cl, 400 mM potassium acetate, 1:1000 dilution of RNAse inhibitor, 1 U/mL RNAse free DNase I, and EDTA-free protease inhibitor cocktail from Sigma (1 tablet in 50 mL buffer)] to the same cell density. Cells were lysed by sonication and debris was spun down at 20,000g for 30 minutes. Equal amounts of supernatant were loaded on a 3-fold or greater volume of sucrose cushion (0.7 M sucrose, 10 mM TrisHCl, pH 7.5, 10 mM MgCl2, 60 mM NH4Cl, 400 mM KOAc, and 1:1000 dilution of RNAse inhibitor) and centrifuged at 148,379g (average) for 5 hours using Beckman MLS-50 ultracentrifuge. Pelleted ribosomes were resuspended in buffer S (20 mM TrisHCl, pH 7.5, 10 mM KCl, 10 mM MgCl2, and 10% sucrose). Samples containing equal amounts of ribosomal material were separated on SDS-PAGE gels and analyzed by Western blot.

**Monitoring λO degradation in lysogens**

WT and tig::cat MC4100 strains were lysogenized with the heat-inducible prophage λcl857Sam7. The lysogens were grown in M63 minimal media supplemented with 0.2% glucose at 30 °C until midlog phase (OD₆₀₀ = 0.6) and then heat shocked at 42 °C for 8 min to induce phage replication. After heat shock, cells were pulse labeled with 200 µCi of [³⁵S] methionine (PerkinElmer, 1000 Ci/mmol) for 2 minutes, and chased with 200 µg/mL non-radioactive methionine at 37 °C for 30 minutes. At different time points, samples were withdrawn during the chase period, and λO protein was immunoprecipitated using α-λO polyclonal rabbit antiserum and 300 µg Protein A-Acrylic beads from *Staphylococcus aureus* (Sigma).

**Monitoring degradation of newly synthesized proteins**

WT, tig::cat, and tig::cat MC4100 cells complemented with a plasmid expressing tig under its
own promoter were grown at 37 °C in M9 minimal medium supplemented with 0.4% glucose (MG medium) until midlog phase, OD$_{600}$ = 0.6. Cells were then pulse labeled with 60 µCi/mL [$^{35}$S] methionine (PerkinElmer, 1000 Ci/mmol) for 2 min and chased with 0.5 mg/mL non-radioactive methionine for 3 min. To remove excess [$^{35}$S] methionine, cells were washed three times in MG medium containing non-radioactive methionine and resuspended in MG medium with non-radioactive methionine and 30 µg/mL tetracycline to inhibit translation. Aliquots of cells were taken at different time points, and samples were TCA precipitated. To determine the relative amount of degraded proteins, radioactivity in the TCA-soluble fraction, containing small peptides, and TCA-insoluble fraction, containing larger polypeptides, was measured by liquid scintillation counting (Beckman Coulter LS6500 Multipurpose Scintillation Counter). TCA-insoluble fractions were resolubilized in 8 M urea before measuring radioactivity. The percentage of degraded proteins was calculated by taking the ratio of the counts from TCA-soluble fraction over total scintillation counts (TCA-soluble plus TCA-insoluble) (Kirstein, Hoffmann et al. 2009). For heat shock and AZC treatment conditions, the procedures were the same except that cells were heated at 42 °C for 10 minutes or treated with 100 µg/mL AZC before pulse labeling with [$^{35}$S] methionine, respectively.
2.4 Results

Conservation of the close proximity of \( tig \), \( clpP \), and \( clpX \) genes across different bacteria

In *E. coli*, \( clpP \) and \( clpX \) genes form a \( clpPX \) operon, while the \( tig \) gene constitutes its own operon but is located immediately upstream of the \( clpPX \) operon. In some other bacteria, such as *Oenococcus oeni* (Jobin, Garmyn et al. 1999) and *Bacillus subtilis* (Gerth, Wipat et al. 1996), \( clpX \) and \( clpP \) are located at different loci on the chromosome but the \( tig \) gene is found directly upstream of \( clpX \). These initial observations prompted us to perform a detailed sequence analysis of all bacterial genomes available on NCBI to explore the genomic positions of \( tig \), \( clpP \), and \( clpX \). All three genes were found to be highly conserved and present in most bacteria. \( tig \) is present in all bacteria; \( clpP \) and \( clpX \) are only missing in Mollicutes (Figure 9). In *Dehalococcoides sp. BAV1*, \( clpP \) is present without \( clpX \), but \( clpA \) is found instead.

In general, the three genes \( tig \), \( clpP \), and \( clpX \) are almost always located next to each other and in the following order, \( tig-clpP-clpX \). Some bacteria have \( tig \) next to \( clpX \) (\( tig-clpX \)), while others have \( tig \) next to \( clpP \) (\( tig-clpP \)). Only in a small number of bacteria are the three genes located far from each other, namely in *Chlorobium chlororochromatii CaD3*, *Streptococcaceae*, and *Rickettsiaceae* (Figure 9). In *Rhodobactersphaeroides, Bartonellatribocorum CIP 105476*, *Rhizobium leguminosarum bv. trifolii*, and *Brucella abortus S19*, \( clpP \) and \( clpX \) are next to each other while \( tig \) is further away (Figure 9). The highly conserved genomic proximity of \( tig \), \( clpP \), and \( clpX \) led us to further investigate the possibility of functional and physical interaction between TF and ClpXP.
null
Figure 9. The genomic positions of \textit{tig}, \textit{clpP}, \textit{clpX} genes across all bacterial protein clusters.

The information was derived from the NCBI and ExPASy databases (May 2012 version) for all available bacterial genomes with representatives shown in the figure. Red, yellow, and blue boxes represent \textit{tig}, \textit{clpP}, and \textit{clpX} genes. The thin lines where the boxes are located represent the genomes with the 5' to 3' orientation indicated from left to right. The thick lines between \textit{clpP} and \textit{clpX} in \textit{Caulobacter crescentus CB15} and between \textit{tig} and \textit{clpP} in \textit{Bifidobacterium longum DJO10A} indicate the presence of 1 and 2 genes in between, respectively. In \textit{Dehalococcoides sp. BAV1}, \textit{tig} and \textit{clpP} are present but no \textit{clpX} was found; instead, \textit{clpA} is in the genome but far away from \textit{tig} and \textit{ClpP}. 
TF modulates the ClpXP-dependent degradation rates of certain substrates

To examine whether there is a functional relationship between TF and ClpXP, degradation assays were performed in the presence and absence of 5 μM TF (1.0 μM ClpX, 1.2 μM ClpP). Several well-established ClpXP substrates were tested and three scenarios were observed. TF had no effect on the ClpXP-dependent degradation of GFP-SsrA monitored on SDS-PAGE gel (Figure 10A) or by fluorescence (Figure 11A). Similarly, the degradation of FlhDC, a known substrate of ClpXP (Takaya, Erhardt et al. 2012), was not affected by TF (Figure 11B). However, MuA degradation by ClpXP was slower in the presence of TF (Figure 10A). The same inhibitory effect of TF was observed on ClpXP-mediated degradation of S7 (30S ribosomal protein), λN-His (antitermination λ phage protein), and Fnr (fumarate nitrate reduction regulatory protein) (Figure 11B). Surprisingly, the substrate λO phage protein was degraded significantly faster by ClpXP when TF was present (Figure 10A). TF itself is not a substrate for the ClpXP or ClpXΔZBDP degradation system in the presence or absence of other ClpXP substrates (Figure 11C).

If TF strictly acts as a folding or holding chaperone in these assays, we would expect it to either have no effect or to protect the substrate from degradation by ClpXP, hence reducing the degradation rate. The enhancement of λO degradation by ClpXP suggests a functional association between the folding chaperone and the degradation system for specific substrates. To investigate whether other folding chaperones have similar effects on ClpXP-mediated degradation as that of TF, the experiments were repeated using DnaK and HtpG (E. coli Hsp90). DnaK is known to have overlapping substrates and functions with TF (Deuerling, Schulze-Specking et al. 1999; Teter, Houry et al. 1999; Deuerling, Patzelt et al. 2003) and, hence, serves as an ideal control for the effect of TF on ClpXP-mediated degradation. The presence or absence of DnaK or HtpG had no effect on the degradation of all three substrates, GFP-SsrA, λO and MuA (Figure 10B), confirming the unique effect observed with TF.

In order to obtain a quantitative estimate of the effect of TF on λO ClpXP-mediated degradation, the protein was labeled with fluorescein isothiocyanate (FITC). The degradation of FITC-λO was monitored in the presence of increasing concentrations of TF. In the native protein, the fluorescence of the FITC labels is reduced due to collisional quenching as a result of close proximity; when λO is degraded into small peptides, this quenching is relieved causing enhanced
FITC fluorescence. As shown in Figure 10C, the initial degradation rate of FITC-λO was enhanced in a TF concentration-dependent manner. At 15 µM TF, the initial degradation rate of FITC-λO doubled (from 4.2 to 8.1 s⁻¹, 1.0 µM ClpX and 1.2 µM ClpP).
Figure 10. The effect of chaperones on the ClpXP-mediated degradation of model substrates.

(A, B) SDS-PAGE gels, visualized by coomassie staining, are shown for the degradation of GFP-SsrA, MuA, and λO performed using 1.0 µM ClpX, 1.2 µM ClpP, 3.9 µM substrate, and an ATP regeneration system in the absence (-) or presence (+) of 5.0 µM TF, DnaK, or HtpG. Quantification of the substrate bands is given on the right. Shown are examples of at least three repeats; the error bars shown for λO degradation represent the standard deviations from three independent experiments. In the presence of DnaK, MuA degradation was visualized by Western blot using anti-MuA antibodies since MuA and DnaK are close in molecular mass and could not be resolved by SDS-PAGE.

(C) The degradation of FITC-λO by ClpXP in the presence of increasing TF concentration was monitored by fluorescence. Initial degradation rates were calculated using the first 100 seconds of the reaction and plotted as a function of TF concentration. All reactions were background-subtracted and repeated at least 3 times to obtain the averages and standard deviations. The inset shows examples of such degradation reactions.
Figure 11. Effect of TF on ClpXP-mediated degradation.

(A) Effect of TF on ClpXP-mediated GFP-SsrA degradation monitored by fluorescence.

(B) Effect of TF on degradation of FlhDC complex, S7, λN-His, and Fnr proteins. For S7, in addition to the in vitro degradation assay, an in vivo experiment was also carried out as described for λO. To monitor S7 degradation in vivo, about $10^6$ cells were loaded for each time point for Western blot analysis using sheep antisera for S7 (a kind gift from Dr. Knud Nierhaus, Max Planck Institute for Molecular Genetics, Germany).

(C) Gels showing that TF is not degraded by ClpP in the presence of ClpX or ClpXΔZBD.
Direct binding of TF to ClpX

Since the degradation efficiency of ClpXP was affected by TF, we investigated whether TF or the TF domains (Figure 12A) exert this effect by directly binding to ClpX, ClpXΔZBD, ZBD, ClpP, and substrates using surface plasmon resonance (SPR) experiments. In these assays, TF or the individual TF domains were immobilized on the sensor chip and different proteins were injected over the chip. TF was found to interact with ClpX but not ClpP (Figure 12B, C). The interaction between TF and ClpX seemed to be mediated mainly by the AAA+ domain of ClpX (ClpXΔZBD), which exhibited about 6 fold stronger interaction with TF than ZBD (K_d of 2 μM vs. 12.8 μM, Figure 12B, C). However, it should be noted that ClpX, in the absence of nucleotides and ClpP, exists as a heterogeneous mixture of many oligomeric states including monomers, dimers, trimers, and hexamers (Grimaud, Kessel et al. 1998). Therefore, the K_d obtained is an apparent dissociation constant, but it provides a qualitative assessment of the binding between TF and ClpX.

TF did not interact with GFP-SsrA (Figure 12B,C), consistent with the observation that TF has no effect on GFP-SsrA degradation by ClpXP (Figure 10A). The interaction between TF and MuA was found to be about 150 fold tighter than that between TF and ClpX (60 nM vs. 9.3 μM, Figure 12B, C). The affinity of MuA monomer for ClpXP has been estimated to be about 10 μM (Abdelhakim, Oakes et al. 2008). Therefore, the high affinity between TF and MuA is expected to sequester MuA away from ClpXP, resulting in slower ClpXP-dependent degradation in the presence of TF, which is what we observed (Figure 10A). Finally, TF bound λO with similar affinity as ClpX (3.3 μM vs. 9.3 μM, Figure 12B, C), suggesting that λO can partition or be transferred between the two systems.

The individual domains of TF bound ClpXΔZBD with 7-700 higher fold affinity than ZBD, confirming that the AAA+ domain of ClpX is the major binding site for TF (Figure 12C and Figure 13). Amongst the individual TF domains, TF_P has the highest affinity for ClpX and ClpXΔZBD (Figure 12C). TF_N and TF_C have similar affinities to ClpX (K_d 10.2 μM and 12.5 μM, respectively), while the K_d for TF_N binding to ClpXΔZBD (1.4 μM) is lower than that of TF_C (32.0 μM) but close to that of TF_P (0.6 μM). The binding between TF domains and λO was also investigated, and only the λO-TF_C interaction showed a saturation binding curve (Figure 13B) with a K_d of 10.8 μM, close to that of λO-TF interaction of 3.3 μM (Figure 12C). The
binding of TN and TP to λO was unspecific as RU increased linearly with increasing ligand concentration (not shown).

The SPR experiments indicate that all three domains of TF are involved in ClpX binding (Figure 12C, top row of numbers), and that the N and P domains of TF have high affinity for the AAA+ domain of the chaperone (Figure 12C, second row of numbers from top). On the other hand, only the C domain of TF binds the substrate λO.

Based on the above results, it is reasonable to expect that TF might modulate the ATPase activity of ClpX. It is well established that the ATPase activity of the chaperone is affected by substrate and ClpP binding (Wojtyra, Thibault et al. 2003). The presence of 5 μM TF stimulated the ATPase of 1 μM ClpX by about 20% compared to ClpX alone, while the control protein, BSA, had no significant effect on ClpX ATPase (Figure 14A). In titration experiments, TF enhanced the ATPase activity of ClpX, ClpXP, ClpXΔZBD, and ClpXΔZBDP (Figure 14B). At maximum TF concentration (80 μM), ClpX and ClpXΔZBD hydrolyzed ATP 2.9 and 2.4 fold faster compared to their intrinsic ATPase activity (784 compared to 270 pmol min⁻¹ μg⁻¹ and 413 compared to 175 pmol min⁻¹ μg⁻¹, respectively). Importantly, TF did not affect the ATPase activity of the related AAA+ protein ClpA (Figure 14B).

The TF single-domain (TN, TP, and TC) and two-domain constructs (TFNP, TFNC, and TFPC) (Kramer, Rutkowska et al. 2004) did not affect the ClpXP-dependent degradation of λO (Figure 15). Furthermore, they did not enhance the ATPase activity of ClpX, ClpXΔZBD, or the control ClpA (Figure 14C), indicating the full length TF is required for such an effect on ClpX ATPase.
A. ClpX

<table>
<thead>
<tr>
<th>ClpX</th>
<th>TF</th>
<th>TFN</th>
<th>TF_P</th>
<th>TFC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9.3</td>
<td>10.2</td>
<td>0.2</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>(1.8)</td>
<td>(3.1)</td>
<td>(0.1)</td>
<td>(2.0)</td>
</tr>
</tbody>
</table>

| ClpXΔZBD | 2.0  | 1.4  | 0.6  | 32.0 |
|          | (0.5) | (0.1) | (0.1) | (3.6) |

| ZBD     | 12.8 | 157.7 | 417.6 | 208.0 |
|          | (2.9) | (23.2) | (104.2) | (43.3) |

| ClpP     | no interaction |
|          |                |

| GFP-SsrA | no interaction |
|          |                |

| λO       | 3.3  | non-specific | 10.8 |
|          | (1.4) |              | (1.8) |

| MuA      | 0.06 | N/A          |
|          | (1.4) |              |

B. ClpX

\[ \text{RU}_{\text{max}} = 157 (17) \]
\[ K_d = 0.3 (1.8) \mu M \]

\[ \text{RU}_{\text{max}} = 146 (12) \]
\[ K_d = 2.0 (0.5) \mu M \]

\[ \text{RU}_{\text{max}} = 51 (9) \]
\[ K_d = 12.8 (2.9) \mu M \]

\[ \text{RU}_{\text{max}} = 2.3 \mu M \]
\[ \text{RU}_{\text{max}} = 30.7 \mu M \]

\[ \text{GFP-SsrA} \]
\[ \text{(no binding)} \]
\[ 0.4 \mu M \]
\[ 3.7 \mu M \]

\[ \text{λO} \]
\[ \text{(no binding)} \]
\[ 2747 (487) \]
\[ K_d = 3.8 (1.4) \mu M \]

C. MuA

\[ K_d = 60 \text{ nM} \]
\[ \chi^2 = 10.2 \]
Figure 12. SPR analysis of the binding between TF, ClpX and substrates.

(A) Schematic of the domain organization of TF and ClpX.

(B) Left panels show the sensorgrams of the binding of different proteins to full length TF, which was coupled on the chip. Right panels show the binding curves as response units (RU) vs. the respective protein concentration at steady state and fitted to a one-site Langmuir binding model. The $K_d$ between TF and MuA was determined using kinetic fitting. For ClpP and GFP-SsrA, injections did not generate any response over time even at high concentrations (30.7 µM and 3.7 µM, respectively) indicating no binding.

(C) The table lists the apparent $K_d$ values obtained from the SPR experiments. The numbers in brackets indicate standard errors.
Figure 13. Binding of TF domains to ClpX and λO.

Shown are sensorgrams (left) and binding curves (right) for the interaction of TF\textsubscript{N} (A), TF\textsubscript{C} (B), and TF\textsubscript{P} (C) to ClpX, ClpX\textsubscript{ΔZBD}, ZBD, and λO. The TF domains were coupled to the chip. The binding of TF\textsubscript{N} and TF\textsubscript{P} to λO was unspecific and is not shown.
Figure 14. The effect of TF on the ATPase activity of ClpX

(A) Shown are the ATPase activities of ClpX in the presence and absence of TF or BSA control. TF and BSA alone have no ATPase contaminants. All reactions were background-subtracted and repeated at least 3 times to obtain the averages and standard deviations.

(B) The ATPase activities of ClpX and ClpXΔZBD in the presence or absence of ClpP and increasing concentration of TF. The ATPase activity of ClpA is not affected by TF.

(C) The ATPase activities of ClpX, ClpXΔZBD and ClpA were measured in the presence of increasing concentration of different TF domains.
Figure 15. Effect of TF domains on ClpXP-mediated λO degradation.

Coomassie stained gels showing the ClpXP-dependent degradation of λO in the presence of different TF domains. For TF<sub>NP</sub> and TF<sub>PC</sub>, the degradation was monitored by Western blot analysis since TF<sub>NP</sub> / TF<sub>PC</sub> have a similar mass as λO.
**λO degradation is compromised in tig deletion strain**

Having demonstrated that TF enhances λO degradation by ClpXP *in vitro*, we carried out experiments to assess whether such an enhancement also occurs *in vivo* in *E. coli* cells. The degradation of λO expressed from a plasmid was measured in WT and trigger factor knockout (Δ*tig*) cells. λO protein expression was induced by IPTG in both strains. By OD$_{600}$ ~ 1.5, protein levels of ClpX, ClpP, TF, and overexpressed λO in WT strain were determined by quantitative western blot analysis as follows: 16000, 20000, 44000, and 390000 molecules/cell, respectively. To measure the degradation of λO *in vivo*, λO was first induced for 2 hours then translation was inhibited by 30 µg/mL of tetracycline. Whole cell lysates obtained at different time points after tetracycline addition were separated on SDS-PAGE gels and proteins were visualized by Western blot analysis. Degradation of λO occurred at a faster rate in WT cells compared to Δ*tig* cells (Figure 16A), which is consistent with the *in vitro* results that TF enhances λO degradation (Figure 10A,C). ClpXP levels remained unchanged during the time course. No significant degradation of λO was observed in Δ*clpX* cells, clearly demonstrating that ClpXP is the main proteolytic system responsible for λO degradation (Gottesman, Clark et al. 1993). Furthermore, it is important to note that λO steady state protein level was significantly higher in Δ*tig* strain compared to WT strain (zero time point in Figure 16A) demonstrating the rapid turnover of this protein by ClpXP. On the contrary, the 30S ribosomal protein S7 was degraded more slowly in the WT strain compared to Δ*tig* strain (Figure 11B), which mimics the *in vitro* results for S7 (Figure 11B). The slower S7 degradation caused by TF suggests that TF may be acting as a chaperone for S7 under these conditions to protect it from ClpXP degradation.

In parallel experiments, λO stability was also analyzed in WT, Δ*tig*, and Δ*clpX* lysogens carrying λcI857Sam7 prophage (Gottesman, Clark et al. 1993). Strains infected with λ phage were grown at 30 °C until OD$_{600}$ ~ 0.4. Lysogens were heat shocked at 42 °C for 8 minutes to induce phage replication and λO protein expression; cells were then returned to 37 °C and pulse labeled with $[^{35}S]$ methionine for 2 minutes and chased with unlabeled methionine for 30 minutes. λO protein level was monitored by immunoprecipitating the protein with anti-λO antibody from samples at the indicated time points during the chase period and visualized by autoradiography. Consistent with the results for λO overexpressed from a plasmid, λO levels in the lysogen were lower in WT compared to Δ*tig* lysogens at the early time points (Figure 16B). λO was not degraded in the Δ*clpX* lysogen (Figure 16B).
It is known that the N-terminus of λO is required for the protein to be recognized by the ClpXP system (Gonciarz-Swiatek, Wawrzynow et al. 1999). Consistent with this, λO expressed from a plasmid with an N-terminal His tag followed by a tobacco etch virus (TEV) cut site (HV-tag) is stable against degradation by ClpXP *in vivo* in the presence or absence of TF (Figure 16C). As a result, TF cannot induce the ClpXP-dependent degradation of a substrate not exposing a ClpX recognition motif. Furthermore, in contrast to what was observed for λO, the levels of HV-λO at the zero time point were the same in WT and Δtig strains, which strongly suggests that TF promotes the rapid degradation of newly synthesized λO protein.

**TF facilitates the degradation of newly synthesized proteins in *E. coli***

In order to determine whether TF exhibits a global effect on the degradation of newly synthesized proteins in *E. coli*, pulse chase experiments were performed in WT and Δtig cells under normal growth conditions (37 °C in LB media). As shown in Figure 16D left panel, the overall proteolysis of newly synthesized proteins was slower in the Δtig strain compared to WT. One hour after the chase, about 7.7% and 6.0% of newly synthesized proteins were degraded in WT and in Δtig cells, respectively. The lower degradation capacity in Δtig strain was complemented by placing the *tig* gene under its own promoter on a plasmid back into Δtig cells (Δtig + ptig, Figure 16D left panel). Evidently, TF is responsible for facilitating the degradation of about 2% of newly synthesized proteins under regular growth conditions.

Under heat shock conditions (42 °C), there was an overall increase in the total amount of proteins degraded compared to 37 °C but the difference between WT and Δtig at 42 °C was similar to that at 37 °C (12.0% in WT and 10.0% in Δtig cells after 1 hr, Figure 16D middle panel). This could possibly be related to the fact that TF is cold shock induced rather than heat shock induced (Kandror and Goldberg 1997). When protein misfolding is induced in cells by adding the proline analogue azetidine-2-carboxylate (AZC), the overall proteolysis of newly synthesized proteins after one hour of chase was 15.6% in WT and 11.8% in Δtig cells (Figure 16D right panel), much higher than that for normal growth conditions. Hence, under this condition, TF promoted the degradation of twice the number of newly synthesized proteins (about 4%) compared to the regular growth condition.
A

<table>
<thead>
<tr>
<th>(min)</th>
<th>WT + λO</th>
<th>Δtg + λO</th>
<th>ΔclpX + λO</th>
</tr>
</thead>
<tbody>
<tr>
<td>λO</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>TF</td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>ClpX</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
</tr>
<tr>
<td>ClpP</td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>(min)</th>
<th>WT + λcI857Sam7</th>
<th>Δtg + λcI857Sam7</th>
<th>ΔclpX + λcI857Sam7</th>
</tr>
</thead>
<tbody>
<tr>
<td>λO</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th>(min)</th>
<th>WT + HV-λO</th>
<th>Δtg + HV-λO</th>
<th>ΔclpP + HV-λO</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV-λO</td>
<td><img src="image16" alt="Image" /></td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
</tr>
<tr>
<td>TF</td>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
<td><img src="image21" alt="Image" /></td>
</tr>
<tr>
<td>ClpX</td>
<td><img src="image22" alt="Image" /></td>
<td><img src="image23" alt="Image" /></td>
<td><img src="image24" alt="Image" /></td>
</tr>
<tr>
<td>ClpP</td>
<td><img src="image25" alt="Image" /></td>
<td><img src="image26" alt="Image" /></td>
<td><img src="image27" alt="Image" /></td>
</tr>
</tbody>
</table>

D

- WT
- Δtg
- Δtg + pTg

- WT at 42°C
- Δtg at 42°C

- WT + AZC
- Δtg + AZC
Figure 16. The role of TF in the cellular degradation of λO.

(A) Western blot analysis of the degradation of overexpressed λO as a function of time after inhibiting protein synthesis in WT, Δtig, and ΔclpX cells. About 10^5 cells were loaded for each time point. Reactions separated by dashed lines are on the same gel, while those separated by solid lines are on different gels. Curves on the right show the quantification of the bands. The error bars are calculated based on three independent experimental repeats.

(B) MC4100 WT and Δtig lysogens carrying heat-inducible prophage λcI857Sam7 were heated at 42 °C for 8 minutes to induce phage protein expression. A pulse chase experiment followed by immunoprecipitation of λO protein was performed to monitor endogenous λO degradation over time. Quantification of the λO band as a function of time is shown on the right.

(C) Same as A, but for cells that overexpress HV-λO from plasmid.

(D) Pulse chase experiments were carried out using [35S] methionine to assess the degradation of newly synthesized proteins in different strain backgrounds at midlog phase at 37 °C (left), or after heat shock at 42 °C for 10 minutes (middle), or after growth in the presence of 100 µg/mL AZC for 30 min at 37 °C (right).
Degradation of ribosome-stalled λO is enhanced by TF

The data from the above experiments suggest that TF promotes the degradation of newly translated proteins either co-translationally or shortly after release from the ribosome. Since TF is a ribosome-associated chaperone, we considered the possibility that the interaction between TF and ClpXP leads to co-translational protein degradation. To this end, the C-terminus of λO was fused to the SecM motif composed of 17 amino acids having a conserved sequence (FxxxxWIxxxxGIRAGP), which has been found to interact with the ribosomal polypeptide exit tunnel and to cause translational arrest (Rutkowska, Beerbaum et al. 2009; Bhushan, Hoffmann et al. 2011). The degradation of the fusion protein was visualized by Western blot analysis after inhibiting translation by tetracycline. Unlike the case for λO, the levels of λO-SecM were similar in WT, Δtig, and ΔclpX cells at the zero time point (Figure 17A versus Figure 16A,B). However, like λO, the degradation of λO-SecM was much slower in the absence of TF compared to WT (Figure 17A).

To ensure that the degradation of λO-SecM occurred while the protein was attached to the ribosome, cell lysates at different time points from the experiment were separated on sucrose cushions (see Methods) and the ribosomal fractions were blotted for λO-SecM. As shown in Figure 17B, ribosome bound λO-SecM was degraded at a slower rate in Δtig compared to WT strain. These results indicate that the ClpXP-dependent degradation of λO-SecM is predominantly co-translational, and is facilitated by TF.
Figure 17. Degradation of ribosome-stalled λO-SecM in vivo

(A) Western blot analysis showing λO-SecM degradation. λO-SecM was induced using pET15b plasmid for 2 hrs by IPTG in different strain backgrounds after inhibition of translation using 30 μg/mL tetracycline. 0− and 0+ refer to samples taken before and directly after addition of tetracycline. Quantification of the λO-SecM band as a function of time is shown on the right based on three experimental repeats.

(B) Same experiment as in A, but shown is the amount of λO-SecM in the ribosomal fraction. Samples with equal OD260 were loaded to ensure that each point contained equal amount of ribosomal material.

(C) Model for TF-ClpXP-dependent degradation of newly translated proteins. Refer to the text for details.
2.5 Discussion

This study provides for the first time evidence that the ribosome-associated folding chaperone TF cooperates with the ClpXP degradation system to promote co-translational protein degradation in prokaryotes. Initial clues for such a possibility came from the broad conservation of the placement of the tig gene next to clpX/clpP genes across bacteria (Figure 9). As shown in Figure 10A, TF can either enhance or retard the ClpXP-mediated degradation of substrates, or have no effect. For substrates whose degradation is not affected by TF, such as GFP-SsrA, it is most likely that TF does not interact with them, which is supported by the SPR analysis showing no binding between TF and GFP-SsrA (Figure 12B,C). For those substrates whose degradation is slowed down by TF, e.g. MuA (Figure 10A), a higher affinity of TF for those substrates (Figure 12B,C) compared to ClpXP would sequester the substrates away from the protease. Finally, for substrates whose degradation is enhanced by TF, such as λO, a similar binding affinity to TF and ClpX was observed. We propose that TF could act as an adaptor which transfers such substrates to ClpX.

The in vitro findings were further supported by the in vivo experiments, which showed that λO degradation, whether from an overexpression plasmid or from phage, is enhanced in WT strain versus a tig deletion strain (Figure 16A,B). Importantly, by arresting λO on ribosomes using a λO-SecM fusion protein, we showed that this degradation occurs co-translationally (Figure 17A,B). To our knowledge, this is the first demonstration of co-translational protein degradation in prokaryotes mediated by TF and ClpXP. We estimate that about 2% of E. coli proteins are degraded co-translationally under normal growth conditions and that this percentage increases under stress conditions that cause global protein misfolding (Figure 16D).

Our findings can explain the observation made by Martinez-Hackert and Hendrickson (Martinez-Hackert and Hendrickson 2009) who found that overexpressing TF in E. coli alleviates aggregation and restores cell viability in cells deficient of TF, DnaK and DnaJ chaperones (ΔtigΔdnaKdnaJ) at 37 °C. In that study, TF was unexpectedly found to have a greater effect on cell viability than on protein aggregation. TF was able to keep protein aggregation levels low in vivo however the soluble fraction did not increase, thus showing that the low levels of aggregation were not the result of the resolubilization of the aggregates. Consistent with our findings, the observations made by Martinez-Hackert and Hendrickson suggest that this is due to
TF directing aggregation-prone proteins to ClpXP for degradation.

Our results set the stage for investigating the mechanistic details of co-translational protein degradation in prokaryotes. Based on the binding studies of Figure 12, the in vivo results of Figure 16 and Figure 17A,B, and the published literature on TF, we propose that TF acts as a bridge between the ribosome and ClpX (Figure 17C). It is possible that the N-terminal domain of TF (TF_N) binds the ribosome, while its P domain (TF_P) binds the AAA+ domain of ClpX. A nascent polypeptide chain coming out of the ribosomal exit tunnel would first interact with the C (arm) domain of TF (TF_C) and then be transferred to the ZBD or AAA+ domain of ClpX (Figure 17C, left panel). The transfer of nascent chain from TF to ClpX would involve the enhancement of ClpX ATPase activity (Figure 14A,B). This model assumes a static interaction of TF with the ribosome. Alternatively, since TF is also known to interact with ribosomes and nascent chains in a dynamic functional cycle (Kaiser, Chang et al. 2006), TF might dissociate from the ribosome while the polypeptide chain is still being made, and then TF_N and TF_P would interact mainly with the AAA+ domain of ClpX while TF_C transfers the bound substrate to ClpX (Figure 17C, middle panel). Finally, if TF dissociates from the ribosome and interacts with ClpXP only upon completion of translation, this would facilitate the rapid degradation of full length newly made proteins (Figure 17C, right panel). Future studies will further elucidate the details of these different models.
3 Structural and Theoretical Studies Indicate that the Cylindrical Protease ClpP Samples Extended and Compact Conformations


*co-first authors

Supplemental movies are not included in this thesis but can be viewed at:

Data Attribution: As co-first authors, Dr. Kimber, Dr. Borg and I conducted the majority of the experiments in this chapter. I performed the following experiments: purification of all the proteins in this study including the mutant (ClpPA153C) for crystallization, biophysical characterization of ClpPA153C using circular dichroism, gel filtration, and sedimentation velocity analysis (Table 2, Figure 19 and Figure 21). The binding between ClpPA153C and chaperone ClpX, and activity of the mutant were also determined by me (Table 3 and Figure 22).
3.1 Summary

The highly conserved ClpP protease consists of two heptameric rings that interact by the interdigitation of an α-helix β strand handle domain motif to form a tetradecameric cylinder. We previously proposed that protease dynamics results in the temporary unstructuring of interacting pairs of handle domains, opening transient equatorial side pores that allow for peptide egress. Here, we report the structure of an *Escherichia coli* ClpP mutant in which each opposing pair of protomers is linked by a disulfide bond. This structure resembles the compact structures of *Streptococcus pneumoniae*, *Mycobacterium tuberculosis*, and *Plasmodium falciparum* ClpPs, rather than the active, extended structures that have previously been determined for *E. coli* ClpPs. The structural data, along with normal mode analysis, support a model whereby the ClpP cylinder switches dynamically between an active extended state required for substrate degradation and an inactive compact state allowing peptide product release (Figure 18).

![Figure 18. Graphical representation of the dynamics of ClpP cylinder.](image)

ClpP samples an extended and a compact state. The extended state is proposed to be required for substrate degradation and the compact state is proposed to allow for peptide-product exit.
3.2 Introduction

The ClpP protease is a highly conserved serine protease found in bacteria and most eukaryotes. The *E. coli clpP* gene codes for a 207 amino acids protein (Figure 19A) where the first 14 residues form a pro sequence that is rapidly cleaved off autocatalytically to form the mature, active protease (Maurizi, Clark et al. 1990a; Maurizi, Clark et al. 1990b). The mature ClpP protomer can be structurally divided into the axial loop (residues 15 – 31 in *E. coli*), a head domain composed of residues 32 – 138 and 172 – 207, and a handle region (β-strand 6 and α-helix E, residues 139 – 171) (Wang, Hartling et al. 1997) (Figure 19A,B). ClpP assembles into a tetradecamer, with the 14 subunits in the ClpP oligomer arranged into two heptameric rings, forming a D7 symmetric cylinder approximately 100 Å in both height and diameter (Wang, Hartling et al. 1997). A large interior chamber (approximately 50 Å in diameter) contains the 14 active sites, whose catalytic triad is composed of Ser111, His136 and Asp185 (*E. coli* ClpP SwissProt numbering). The intra-ring contacts are mediated by the head domain, and are predominantly hydrophobic in nature while the interaction between the two ClpP rings is mediated mainly by the handle region. Surprisingly, truncations in the handle region do not result in the dissociation of the two rings, implying a high degree of plasticity of this region (Gribun, Kimber et al. 2005). It was proposed that the interaction between the two ClpP rings is stabilized by charge-charge interactions contributed by residues of the head domains (Maurizi, Singh et al. 1998; Gribun, Kimber et al. 2005).

Entry of protein substrates into the ClpP lumen occurs through the two axial pores. These pores are lined by flexible axial loops, which in some conformations form β hairpins that start from inside the ClpP cylinder, extend above the cylinder and then double back to pack between symmetry related copies of helix αA on the surface of the heptameric ring (Figure 19B and Figure 20A). These loops are observed in human (Kang, Maurizi et al. 2004), *S. pneumoniae* (Gribun, Kimber et al. 2005), and in later *E. coli* ClpP structures (Bewley, Graziano et al. 2006; Szyk and Maurizi 2006). These axial loops gate the entry of substrates into the ClpP proteolytic chamber (Gribun, Kimber et al. 2005; Bewley, Graziano et al. 2006; Jennings, Bohon et al. 2008). The axial loops mediate the interaction of ClpP with its cognate ATPases (Gribun, Kimber et al. 2005; Martin, Baker et al. 2008), and, hence, are important for ClpP function.
Figure 19. The sequence and secondary structure of *E. coli* ClpP

(A) The axial loop is colored green, the equatorial handle region cyan, and the head domain yellow. Ser111, His136, and Asp185 residues of the catalytic triad are underlined. Ala153 is boxed. (B) Secondary structure elements of ClpP (1YG6), colored as described in (A). The side chain of Ala153 is shown in ball and stick representation.
Figure 20. The extended and compact structures of ClpP.

(A) The structure of *E. coli* ClpP (1YG6) showing the ClpP tetradecamer in the extended state (Bewley, Graziano et al. 2006).

(B) Structure of *P. falciparum* ClpP (2F6I) showing ClpP in the compact state (Vedadi, Lew et al. 2007). The handle region is predominantly disordered in *P. falciparum* ClpP, and the two rings are approximately 10 Å closer together. The axial loops in this structure are predominantly disordered, though this is not true of all structures in the compact configuration (e.g. *S. pneumoniae* ClpP (Gribun, Kimber et al. 2005).
The structures of ClpP from six organisms have been experimentally determined and are found to be composed of highly similar protomers (Yu and Houry 2007). However, we noted that they exhibit differences in oligomeric organization and can be grouped into two distinct structural states. The first state, observed for ClpP from *E. coli* (Wang, Hartling et al. 1997; Bewley, Graziano et al. 2006; Szyk and Maurizi 2006), *H. sapiens* mitochondria (Kang, Maurizi et al. 2004), and *H. pylori* (Kim and Kim 2008), we term the extended state. In this state, the distance from the apical surface of one subunit in one ring to the apical surface of the opposing subunit in the second ring (excluding the N-terminal loops) is approximately 100 Å (Figure 20A). In this state residues 139 to 150 are ordered and form β6 and the N-terminal end of αE of the handle domain; these residues from opposing rings interlock to form a continuous, unbroken surface at the ClpP equator. The catalytic triad in the extended state is generally organized in a catalytically competent, or near competent configuration. A second state, observed in the ClpP structures from *S. pneumoniae* (Gribun, Kimber et al. 2005), *P. falciparum* (Vedadi, Lew et al. 2007), and *M. tuberculosis* (Ingvarsson, Mate et al. 2007) we term the compact state (Figure 20B). In this state, residues 139 to 150 are typically disordered, with residual electron density generally indicating that these residues occupy the lumen of the ClpP cylinder. The opposing rings are shifted closer together by a screw-like motion which slides the surfaces of opposing αE helices over one another, resulting in the opposing apical surfaces being approximately 10 Å closer. The component residues of the catalytic triad in these structures are disorganized, which, coupled with the disordered β6 guide strand, indicates that this structure likely corresponds to a catalytically inactive state of ClpP. It is worth noting that, to date, there is no firm experimental data showing that both states are accessible to a given ClpP from a specific species.

ClpP on its own can efficiently degrade small peptides (Thompson and Maurizi 1994; Thompson, Singh et al. 1994) as well as very slowly degrade poorly folded proteins (Jennings, Lun et al. 2008; Bewley, Graziano et al. 2009). The enzyme processively cleaves substrate proteins into peptides of seven to eight residues, which are then released from the chamber (Choi and Licht 2005; Licht and Lee 2008). The release mechanism is still under investigation. One model proposes that degraded products exit the ClpP proteolytic chamber by the same axial pores that allow polypeptide entry by passive diffusion (Thompson and Maurizi 1994; Thompson, Singh et al. 1994). A second model, which we put forward, proposes that peptides are released through side pores that transiently form at the interface between the two heptameric
ClpP rings by localized unfolding of pairs of interacting handle domains (Gribun, Kimber et al. 2005; Sprangers, Gribun et al. 2005).

In this study, we determined the structure of a cross-linked *E. coli* ClpP in which the two heptameric rings of ClpP are held together by disulfide bonds. Surprisingly, while all *E. coli* ClpP structures solved to date are in the extended state, the cross-linked ClpP structure was found to be in the compact state. Based on structural analysis and theoretical considerations, we propose that the ClpP compact state is naturally sampled by the protease in the course of its dynamics and that this state is functionally important, possibly for promoting degraded product exit from the protease catalytic chamber in the presence of the cognate chaperone.
3.3 Experimental Procedures

Protein purification

*E. coli* WT ClpP was purified as previously described (Wojtyra et al., 2003). *E. coli* ClpP(A153C) mutant was generated from WT ClpP using the QuikChange kit (Stratagene) following manufacturer’s protocol. The construct was then subcloned into pET9a vector and transformed into SG1146(DE3) ΔclpP cells to express untagged ClpP(A153C). A single transformant colony was grown overnight in 10 mL of Luria-Broth (LB) media containing 100 μg/mL ampicillin at 37°C. The overnight culture was inoculated into 1 L LB with 100 μg/mL ampicillin and grown at 37°C. When OD600 reached 0.6, isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added at a final concentration of 1 mM to induce protein expression for 3 hours at 37°C. Cells were then collected by centrifugation and lysed by French press. Cell debris was removed by centrifugation. The supernatant containing ClpP(A153C) was partially purified by ammonium sulphate precipitation at 30-60% saturation. Precipitated ClpP(A153C) was resolubilized and dialyzed in buffer A (50 mM TrisHCl, pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol). The protein was further purified using an anion-exchange column (Q Sepharose). The high salt buffer (buffer B) contained 50 mM TrisHCl, pH 7.5, 1 M KCl, 1 mM DTT, and 10% glycerol. ClpPSS eluted at 250-300 mM KCl. Subsequently, the protein was further purified by size exclusion chromatography using Superdex 200 column equilibrated with buffer A. The protein was then stored in buffer A, but in the absence of DTT. The yield was typically 60 mg of purified ClpPSS from 1 L culture. Human ClpP was expressed from the plasmid pDT1668 LhclpP, which encodes carboxyterminal 6xHis tagged human mitochondrial ClpP (a generous gift from Dr. David Dougan, La Trobe University, Australia). pDT1668 LhclpP was transformed into SG1146(DE3) ΔclpP strain. The expressed protein was purified on Ni-NTA-agarose (Qiagen) according to manufacturer’s protocols.

Protein biochemical characterization

To visualize the mobility of WT ClpP and ClpP(A153C) on reducing and non-reducing SDS-PAGE gels (Figure 21A), the proteins were incubated for 30 minutes at 37°C in the presence or absence of 10 mM DTT and then incubated with 55 mM iodoacetamide (Sigma) for 30 minutes at 37°C to block free cysteines. Size exclusion chromatography was carried out using a Superdex
200 HR 10/30 (GE Healthcare) column attached to an AKTA FPLC (GE Healthcare). The column was equilibrated with buffer C (50 mM TrisHCl, pH 7.5, 200 mM KCl) in the presence or absence of 1 mM DTT. Molecular mass standards (Sigma) used were: thyroglobulin (669 kDa), apoferritin (443 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome c (12.4 kDa). All experiments were performed at 4°C. Proteins were detected using absorbance measured at 280 nm. CD spectra were measured at 25°C in a 1 mm path length cuvette using a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control device. Prior to the measurement, the protein was incubated at 4°C overnight in buffer C with or without 1 mM DTT. The final protein concentration was 10 µM. Figure 21B shows the average of three recorded spectra.

**Sedimentation velocity analytical ultracentrifugation**

Sedimentation velocity experiments were carried out at the Ultracentrifugation Service Facility at the Department of Biochemistry, University of Toronto. Proteins were buffer exchanged into buffer C in the presence or absence of 1 mM DTT using Bio-Spin 6 column (Bio-Rad). Samples were then incubated at 37°C for 2 hours prior to the experiment. Samples were spun at 25,000 rpm (45,500 g at cell centre) at 20°C in a Beckman Optima Model XL-A analytical ultracentrifuge equipped with An-60 Ti rotor. The density (ρ) and viscosity (η) of buffer C were calculated to be 1.00909 g/mL and 0.010143 poise at 20°C, respectively, using Sednterp (Laue 1992). Partial specific volumes were calculated to be 0.7387 mL/g for WT ClpP, 0.7382 mL/g for ClpP(A153C), and 0.7407 mL/g for human ClpP using Sednterp based on the amino acid composition of each protein. The sedimentation data were fit to a continuous distribution model c(s) using SEDFIT (Schuck 2010). The sedimentation coefficients obtained from the fitting were corrected to the density and viscosity of water at 20°C to get $s_{20, w}$.

**Peptidase and proteinase assays**

The peptidase activity of WT ClpP and ClpP(A153C) under reducing and non-reducing conditions was determined by measuring the rate of cleavage of the fluorogenic peptide Nsuccinyl-Leu-Tyr-7-amino-4-methyl-coumarin (Suc-LY-AMC) purchased from Sigma. 0.5 µM
ClpP was equilibrated in buffer C at 37°C for 2 minutes in the presence or absence of 1 mM DTT and then 0.1 to 3.0 mM Suc-LY-AMC was added to start the reaction. The fluorescence change resulting from the release of AMC upon cleavage of the Y-AMC bond was monitored for 10 minutes on a Fluorolog 3-222 spectrofluorometer (Jobin Yvon) using an excitation wavelength of 360 nm and an emission wavelength of 440 nm. A standard curve for AMC fluorescence was constructed to convert fluorescence into concentration units in order to determine the amount of AMC generated. $K_M$ and $k_{cat}$ were obtained from double reciprocal plots.

Degradation of casein by ClpAP was carried out by first pre-incubating 1.5 μM of α-casein with 1.2 μM ClpP [WT or ClpP(A153C)], ATP regenerating system (0.32 mg/mL creatine phosphokinase, 16 mM creatine phosphate, and 3 mM ATP), plus or minus 10 mM DTT in buffer D (25 mM HEPES, pH 7.5, 20 mM MgCl$_2$, 300 mM KCl, 0.03% Tween 20, and 10% glycerol) for 3 minute at 37°C. All concentrations are those of monomers. To start the degradation reaction, 1.0 μM ClpX was added to the mixture incubated at 37°C. Samples were withdrawn at indicated time points and immediately mixed with 4x Laemmli buffer to stop the degradation reaction. Proteins samples isolated at different degradation time points were separated on SDS-PAGE gels and visualized by Coomassie staining. The degradation of fluorescein isothiocyanate labeled casein (Sigma) was monitored by fluorescence using a Fluorolog 3-222 spectrophotometer (Horiba Jobin Yvon). Excitation wavelength was set at 490 nm and emission wavelength at 525 nm.

**ATPase assays**

ATPase activity of ClpX was measured using the ATP/NADH coupled assay (Norby 1988). Various concentrations of ClpP or ClpP(A153C), 1.0 μM ClpX, and NADH mixture (0.2 mM NADH, 3.0 mM phosphoenolpyruvate, 0.05 mg/ml pyruvate kinase, 0.025 mg/mL lactate dehydrogenase) were pre-incubated at 37°C for 3 minutes in ClpX buffer (25 mM HEPES, pH 7.5, 5 mM MgCl$_2$, 5 mM KCl, 0.03% Tween 20, and 10% glycerol) plus or minus 10 mM DTT.

All concentrations are those of monomers. To start the reaction, 5 mM ATP was added and the change in absorbance at 340 nm was measured for 20 min at 37°C. Each reaction was repeated at least three times. The data were fit to a simple single-site Langmuir binding isotherm.
**ClpP^{SS} structure**

ClpP^{SS} was crystallized by the sitting drop method with vapor diffusion using 1 μL of 23 mg/mL protein solution mixed with 1 μL of buffer (1 M 1,6-hexanediol, 0.1 M sodium acetate, pH 4.6, 10 mM CoCl_{2}, and 100 mM CaCl_{2}). The protein was equilibrated against the same buffer. Crystals grew as hexagonal bipyramids up to 1.2 mm in length. Crystals were gradually dehydrated by exposure to air prior to freezing in a 100 K nitrogen stream. X-ray diffraction data were collected at Advanced Photon Source (APS), Argonne National Laboratory (ANL), Chicago, Illinois, USA, on beamline 19ID and processed in HKL3000 (Minor, Cymborowski et al. 2006). Diffraction data conformed to the trigonal space group P3_{2}2_{1} with cell dimensions a = b = 182.30 Å and c = 476.86 Å. Despite the long c-axis, low mosaicity (0.47°) resulted in individual reflections being sharp and well resolved. Crystals diffracted X-rays to a maximal resolution of 3.2 Å. Data collection and model refinement statistics are reported in Table 4. Data collection and structure refinement statistics.

The structure was solved using Molrep in CCP4 (Collaborative 1994), searching with one heptameric ring from 2FZS (Szyk and Maurizi 2006). Four such heptameric rings were found arranged as two tetradecamers, and individual domains were subjected to rigid body refinement. Despite the high solvent content (72%), no further heptameric rings could be found. Given the contiguous crystal lattice contacts in all directions and good refinement statistics for the final model, the structure is deemed to be complete. Refinement was performed using Refmac in CCP4 with each monomer refined as a single TLS (translation/libration/screw) group. Structure rebuilding was done in XtalView (McRee 1999) and Coot (Emsley and Cowtan 2004). NCS restraints were found to degrade the R_{free} and were not used in the refinement. Axial loop residues 15 – 31 were disordered in all structures. Most residues in the handle domain were disordered, typically residues 138 – 151, though the first disordered residue ranged from 137 to 141, and the last disordered residue between 147 and 152 (Table 5).

The coordinates and structure factors for ClpP^{SS} were deposited in the Protein Data Bank (PDB) with accession number 3HLN.
3.4 Results and Discussion

Biochemical characterization of disulfide cross-linked ClpP

In a mutant variant of ClpP where A153 in helix αE of ClpP (Figure 19B) is mutated to cysteine, the introduced cysteines of two ClpP subunits from opposite rings readily oxidize, in the absence of reducing agent, to form a disulfide cross-linked ClpP-A153C (Figure 21A). The overall secondary structure content of ClpP-A153C is very similar to that of WT ClpP under reducing and non-reducing conditions as deduced from circular dichroism (CD) spectroscopy at 4°C (Figure 21B). Furthermore, reduced and non-reduced ClpP-A153C migrate at around 300 kDa as observed by size exclusion chromatography at 4°C (Figure 21C), which is close to the expected size of the tetradecamer. A more careful analysis of the oligomerization of the protease at 20°C by sedimentation velocity analytical ultracentrifugation (AUC) indicates the presence of multiple oligomeric solution states for WT and mutant ClpP under reducing and non-reducing conditions (Figure 21D). However, the main species observed have a measured molecular weight of 287 – 346 kDa (Table 2), which is close to the theoretical molecular weight of ClpP tetradecamer of 302 kDa. The sedimentation coefficients obtained, $s_{20,w}$, for the main species are also consistent with a tetradecameric oligomer (Table 2). Hence, it can be concluded that the main oligomeric form of ClpP and ClpP-A153C under reducing and non-reducing solution conditions is the tetradecamer. Human mitochondrial ClpP, which is known to form a single heptameric ring in solution (Kang, Dimitrova et al. 2005), was used as a control in the AUC experiments (Figure 21D and Table 2).
A

![Image of gel electrophoresis](image1)

**Reducing**
- WT A153C
- ClpP(A153C)

**Non-reducing**
- ClpP
- WT ClpP

B

**CD spectra of WT ClpP and ClpP(A153C)**

![Graph showing CD spectra](image2)

C

**Relative absorbance change at 280 nm**

![Graph showing relative absorbance change](image3)

D

**c(s) (OD at 260 nm)**

![Graph showing c(s) values](image4)

- WT ClpP non-reduced
- WT ClpP reduced
- ClpP(A153C) non-reduced
- ClpP(A153C) reduced
- Human ClpP reduced
Figure 21. Biophysical characterization of ClpP(A153C)

(A) Mobility of WT ClpP and ClpP(A153C) in denaturing reducing and nonreducing gels. (B) Far-UV CD spectra at 25°C of ClpP and ClpP(A153C) under reducing (+DTT) and nonreducing (no DTT) conditions. (C) ClpP(A153C) under reducing and nonreducing conditions has similar elution profiles as observed by size exclusion chromatography using a calibrated Superdex 200 HR 10/30 column equilibrated ± DTT at 4°C. (D) Profiles obtained from fitting the sedimentation velocity analytical ultracentrifugation data for WT and A153C *E. coli* ClpP and human mitochondrial ClpP to a continuous distribution model c(s) versus sedimentation coefficients in Svedberg units (S).
### Table 2. Results of the sedimentation velocity analytical ultracentrifugation experiments.

<table>
<thead>
<tr>
<th></th>
<th>$s_{20,w}$ (S)</th>
<th>$MW$ (kDa)</th>
<th>Percent of total for major species</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT ClpP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reduced</td>
<td>11.6</td>
<td>287</td>
<td>71</td>
</tr>
<tr>
<td>non-reduced</td>
<td>11.7</td>
<td>288</td>
<td>74</td>
</tr>
<tr>
<td>ClpP(A153C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reduced</td>
<td>12.0</td>
<td>346</td>
<td>55</td>
</tr>
<tr>
<td>non-reduced</td>
<td>12.2</td>
<td>308</td>
<td>56</td>
</tr>
<tr>
<td>human ClpP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reduced</td>
<td>6.3</td>
<td>139</td>
<td>90</td>
</tr>
</tbody>
</table>

The table lists the solvent corrected sedimentation coefficients, $s_{20,w}$, at maximum c(s) as well as the corresponding apparent molecular weight of the major species observed obtained from fitting the sedimentation velocity analytical ultracentrifugation data (Figure 21D). The percent of total for the major species observed is also given for each of the samples.
Under reducing conditions, in which the disulfide bonds are not formed, ClpP(A153C) peptidase activity against the fluorogenic peptide N-succinyl-Leu-Tyr-7-amino-4-methylcoumarin (Suc-LY-AMC) has a similar $K_M$ but a 32-fold lower $k_{cat}$ compared to that of WT ClpP (Table 3). This seems to indicate that the A153C mutation does not affect the substrate binding site, but does affect the conformation of the active site of the protease. Alternatively, ClpP might be in equilibrium between an active and inactive conformation, and the mutation pushes the equilibrium towards the inactive state (see below). In the absence of DTT, the $K_M$ for WT ClpP is unaffected, while the $k_{cat}$ is reduced by about 1.6-fold compared to the values obtained in the presence of DTT (Table 3). The enhancement in WT ClpP activity by DTT seems to be caused by a non-thiol mediated side effect of DTT (Alliegro 2000) (data not shown). The disulfide cross-linked ClpP(A153C) formed in the absence of DTT has no detectable peptidolytic activity (Table 3). This could be either due to a local rearrangement of the active site, which is located at the base of the handle region, or to a more global, although subtle, effect on ClpP structure. Consistent with the above observations, in the presence of the cognate chaperone ClpX, disulfide cross-linked ClpP(A153C) does not degrade GFP-SsrA model substrate (data not shown; see also (Sprangers, Gribun et al. 2005)). Furthermore, in the presence of ClpA, the rate of degradation of the poorly structured model substrate casein by ClpP(A153C) is abolished under non-reducing conditions and is lower than that of WT ClpP under reducing conditions (Figure 22A, B). Note that, as observed for the peptidase activity (Table 3), the proteinase activity of WT ClpP is enhanced in the presence of DTT (Figure 22A,B).
Table 3. Peptidase activity of WT ClpP and Clp(A153C).

<table>
<thead>
<tr>
<th></th>
<th>$K_M$ (mM)</th>
<th>$k_{cat}$ (min$^{-1}$) per protomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT ClpP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reducing</td>
<td>1.0 ±0.1</td>
<td>15.8 ±1.0</td>
</tr>
<tr>
<td>non-reducing</td>
<td>0.8 ±0.1</td>
<td>9.6 ±0.3</td>
</tr>
<tr>
<td>ClpP(A153C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reducing</td>
<td>1.1 ±0.1</td>
<td>0.5 ±0.1</td>
</tr>
<tr>
<td>non-reducing</td>
<td>inactive</td>
<td>inactive</td>
</tr>
</tbody>
</table>

$K_M$ and $k_{cat}$ for the cleavage of Suc-LY-AMC fluorogenic peptide were measured after incubation at 4°C overnight under reducing or non-reducing conditions.
Figure 22. Binding of ClpP(A153C) to the AAA+ Chaperones.

(A) The degradation of FITC-labeled casein by WT ClpP or ClpP(A153C) under reducing and nonreducing conditions in the presence of ClpA is shown as monitored by fluorescence. (B) The degradation of α-casein by WT ClpP or ClpP(A153C) under reducing and nonreducing conditions in the presence of ClpA is shown on SDS-PAGE gels. The autodegradation of ClpA under reducing conditions can be observed at later time points. (C) The inhibition of ClpX ATPase activity by WT ClpP or ClpP(A153C) under reducing (open circles) and nonreducing (closed circles) conditions is shown. Data points are averages of three repeats. The data were fit to a single-site Langmuir binding isotherm (solid lines) to obtain apparent binding constants. Numbers in parenthesis refer to standard deviations of at least three experiments. The measured ATPase activity of ClpX in the absence of the protease is 233 (±10) pmol min⁻¹ μg⁻¹ under reducing and nonreducing conditions.
The interaction of ClpP(A153C) with ClpX was assessed by measuring the ATPase activity of ClpX in the presence of increasing amounts of ClpP. It is known that ClpP inhibits ClpX ATPase activity (Wojtyra, Thibault et al. 2003; Joshi, Hersch et al. 2004). Fitting the data to a simple single-site Langmuir binding isotherm indicates that binding of ClpP(A153C) to ClpX under reducing conditions is similar to that of WT ClpP under reducing and non-reducing conditions (Figure 22C). However, ClpP(A153C) does not bind (or very weakly binds) ClpX under non-reducing conditions (Figure 22C).

The results of Figure 21 and Figure 22 suggest that ClpP(A153C) under non-reducing conditions is an inactive protease that does not bind its cognate chaperone.

**X-ray structure of disulfide cross-linked ClpP**

In order to gain further understanding of the conformational changes in ClpP, we determined the X-ray crystal structure of ClpP(A153C) under non-reducing conditions. We will refer to disulfide cross-linked *E. coli* ClpP(A153C) as ClpP\textsuperscript{SS}.

The structure of ClpP\textsuperscript{SS} was determined at 3.2 Å resolution. Table 4 lists the data collection statistics and the final refinement statistics. There are 28 chains, comprising two complete tetradecamers, in the asymmetric unit, along with fourteen well-ordered Ca\textsuperscript{2+} ions that predominantly mediate crystal contacts. At 3.2 Å resolution, some details are difficult to discern; nevertheless, the broad features of the structure are readily apparent. The structure interpretation is helped by the availability of high-resolution model structures to guide the correct placement of the side chains even where the electron density is ambiguous. Furthermore, given the high solvent content of this crystal form (72%), there are four reflections per atom, a ratio more typical of a 2.7 Å structure. Finally, having 28 independent copies in two tetradecamers allows multiple independent views of each detail, greatly improving the reliability of the interpretation. At minimum, residues 32 to 137 and 153 to 207 are built for all monomers, though details vary by chain (Table 5).
Table 4. Data collection and structure refinement statistics.

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>100 K</td>
</tr>
<tr>
<td>Space group</td>
<td>P3_21</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a = b = 182.30 Å</td>
</tr>
<tr>
<td></td>
<td>c = 476.86 Å</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.0000</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>40.0 – 3.2</td>
</tr>
<tr>
<td>Total observations</td>
<td>437604</td>
</tr>
<tr>
<td>Unique observations</td>
<td>148306</td>
</tr>
<tr>
<td>Completeness (last shell)^a</td>
<td>0.975 (0.917)</td>
</tr>
<tr>
<td>&lt;I/σ(I)&gt; (last shell)^a</td>
<td>14.2 (2.1)</td>
</tr>
<tr>
<td>R_sym (last shell)^a</td>
<td>0.045 (0.394)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>40 - 3.2</td>
</tr>
<tr>
<td>Asymmetric unit contents</td>
<td></td>
</tr>
<tr>
<td>Chains</td>
<td>28</td>
</tr>
<tr>
<td>Ca^{2+}</td>
<td>14</td>
</tr>
<tr>
<td>R_cryst</td>
<td>0.213</td>
</tr>
<tr>
<td>R_free</td>
<td>0.253</td>
</tr>
<tr>
<td>r.m.s.d. bond lengths (Å)</td>
<td>0.007</td>
</tr>
<tr>
<td>r.m.s.d. bond angles (°)</td>
<td>0.99</td>
</tr>
<tr>
<td>Ramachandran plot:</td>
<td></td>
</tr>
<tr>
<td>Most favorable (%)</td>
<td>94.6</td>
</tr>
<tr>
<td>Additionally allowed (%)</td>
<td>4.9</td>
</tr>
</tbody>
</table>

^aThe last shell includes all reflections between 3.31 and 3.20 Å.

^bR_free calculated using 5% of the data which were chosen randomly
Table 5. Disordered residues in the ClpP<sup>SS</sup> structure

<table>
<thead>
<tr>
<th>Tetradecamer 1 (A-N)</th>
<th>Tetradecamer 2 (O-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain</td>
<td>Disordered residues</td>
</tr>
<tr>
<td>A</td>
<td>15-31, 138-151</td>
</tr>
<tr>
<td>B</td>
<td>15-31, 139-152</td>
</tr>
<tr>
<td>D</td>
<td>15-31, 141-151</td>
</tr>
<tr>
<td>F</td>
<td>15-31, 138-147</td>
</tr>
<tr>
<td>I</td>
<td>15-31, 139-151</td>
</tr>
<tr>
<td>J</td>
<td>15-31, 139-151</td>
</tr>
<tr>
<td>K</td>
<td>15-31, 139-152</td>
</tr>
<tr>
<td>L</td>
<td>15-31, 138-151</td>
</tr>
</tbody>
</table>
Disulfide cross-linking appears to minimally perturb the protomer structure, and individual monomers of ClpP<sup>SS</sup> superpose well with the protomer structures previously determined for WT <i>E. coli</i> ClpP (~0.4 Å r.m.s.d., Figure 23A, protomers on left), although there is some structural heterogeneity among different ClpP<sup>SS</sup> protomers (Figure 24A). Among the more notable local changes is a significant bending of αE, as well as the disorder of residues Pro138 to Cys153, though some protomers have a few more residues ordered. The electron density map clearly shows that Cys153 in each monomer forms a disulfide bridge with the corresponding residue in the opposite ring (Figure 23B). Cross-linking appears to predominantly induce significant changes within the oligomeric structure at two levels. First, packing within the ring is subtly changed, with each protomer translated inward towards the sevenfold axis by as much as 1 Å (Figure 23A, protomers on right). Note that comparisons of all <i>E. coli</i> ClpP ring structures shows that some variations occur between the structures, but ClpP<sup>SS</sup> represents an extreme case (Figure 24B). Second, all <i>E. coli</i> ClpP structures solved to date exhibit the extended form; the compact conformation has never before been observed for the <i>E. coli</i> protease. Here, the structure of ClpP<sup>SS</sup> is clearly in the compact form, adopting a conformation closely resembling the previously determined structures of <i>S. pneumoniae</i> and <i>P. falciparum</i> ClpP (Figure 23C). (The <i>M. tuberculosis</i> ClpP is also in a compact conformation, but the protomers within a ring are rotated outward, making detailed comparisons difficult.) The similarity of these three compact structures is interesting as they involve three different sequences with different modifications: A153C oxidized, A153P (using <i>E. coli</i> SwissProt sequence numbering), and WT for <i>E. coli</i>, <i>S. pneumoniae</i>, and <i>P. falciparum</i>, respectively. The structural similarities suggest that this is a naturally occurring, low energy structural state for ClpP. For ClpP<sup>SS</sup> in particular, the concerted structural and functional changes observed as the protein is switched from reducing to oxidizing conditions, argue that the <i>E. coli</i> ClpP protein must have an innate ability to switch between these states.
Figure 23. The overall structure of ClpP<sup>SS</sup> and comparison with WT ClpP.

(A) The protomers on the left show the superposition of a single subunit of ClpP<sup>SS</sup> (blue) on WT ClpP (yellow, 1YG6). On the right are protomers from the opposite side of the heptameric ring. Arrow here and arrows in (B) show the direction of protomer motion from the extended to the compact conformations. (B) Structure of <i>E. coli</i> ClpP<sup>SS</sup> showing four monomers as cartoon traces with Cys153 represented as spheres. A surface trace for the entire tetradecamer is also shown. In ClpP<sup>SS</sup>, β6 and two turns of the αE helix are completely disordered. For comparison, WT <i>E. coli</i> ClpP (1YG6) is also shown with Ala153 drawn as spheres. The magnified area depicts the 2mFo-DFc electron density map, contoured at 1.0 σ, showing the handle region viewed from inside the ring. Electron density corresponding to the disulfide bond is clear in all 14 interring disulfide bonds (not shown). (C) Superposition of the <i>E. coli</i> ClpP<sup>SS</sup> tetradecamer (cyan cartoon with residue 153 in sphere representation) on the ClpP structures from <i>S. pneumoniae</i> (gray, 1Y7O) and <i>P. falciparum</i> (yellow, 2F6I). Only the lower ring was used for superposition. In general, single rings in all three structures superpose well, but there are minor differences in the degree to which the rings approach one another, with ClpP<sup>SS</sup> being the least compact. (D) WT <i>E. coli</i> ClpP is shown (2FZS) as a gray surface, showing yellow where it is sliced through in the plane of the page at the level of the Val20 side chain. The corresponding footprint of the ClpP<sup>SS</sup> surface in the same plane is overlain as a semitransparent pale blue surface. Cartoon and stick representations for selected residues that protrude above this surface are in yellow (WT) and pale blue (ClpP<sup>SS</sup>). Residues 17–22 of WT ClpP, in pink, bind in the groove between adjacent copies of the helix αA. The domain shifts in ClpP<sup>SS</sup> pinch closed this groove, disrupting the binding surface of the axial loop that anchors the loop to the protease apical surface resulting in a disordered loop. See also Figure 20, Figure 24, Figure 25, and Table 5.
Figure 24. Overlay of *E. coli* ClpP structures.

(A) Overlay of all 28 protomers of the ClpP<sup>SS</sup> structure. In general, all monomers superpose well, except in the region of the handle domain. αG, which interacts with the αG of a second monomer across the ring, also shows some structural variability.

(B) Stereoview of the overlay of the tetradecamers from the four independently determined *E. coli* ClpP structures. ClpP<sup>SS</sup> is in cyan, 1YG6 is in yellow, 1YG8 is in blue, and 2FZS is in magenta. ClpP<sup>SS</sup> clearly differs from all of the other structures, which closely resemble one another except in the conformation of the axial loops.
Figure 25. Overlay of the catalytic sites in ClpP.

(A) Stereoview of the overlay of the 28 independent Ser111-His136-Asp185 catalytic triads from the two Clp<sup>SS</sup> tetradecamers in the asymmetric unit. The active site shows a high degree of disorder. Ser111 mostly shows differences in rotamer, while His136 and Asp185 show a high degree of variability in both backbone placement and side chain rotamer. None of the individual catalytic sites are properly positioned for catalysis.

(B) Stereoview of the overlay of the 14 independent active sites of WT E. coli ClpP (1YG6). The active sites are all very similar, with the exception of Ser111, which samples two preferred rotamers. Aside from the mispositioning of Ser111, these active sites are appropriately pre-organized for catalysis.
The structure of ClpP<sup>SS</sup> contrasts markedly with the extended tetradecameric structure seen in all previous <i>E. coli</i> ClpP structures (Figure 20A), as well as the similarly extended <i>Helicobacter pylori</i> and human mitochondrial ClpP structures. In the ClpP<sup>SS</sup> tetradecamer structure, the two opposite rings are shifted approximately 4 Å closer relative to WT <i>E. coli</i> ClpP, and one ring is rotated approximately 5° around the 7-fold NCS axis. Collectively, this motion can be visualized as a screw-like motion generated by the sliding of the equatorial helices αE approximately one full turn along their length towards the opposite subunit (Figure 23B). These rearrangements also result in the catalytic site being disorganized similar to what is also observed for <i>S. pneumoniae</i> ClpP(A153P), <i>M. tuberculosis</i> ClpP, and <i>P. falciparum</i> ClpP, while WT <i>E. coli</i> ClpP structures show that the active site is well organized (Figure 25). This reinforces the idea that the compact conformation is accompanied by the formation of an inactive arrangement in the catalytic site. This is the first instance in which the extended and compact forms are observed for ClpP from the same organism.

As mentioned earlier, the N-terminal axial loop is comprised of residues 15 – 31 (Figure 19A,B). Residues 22 – 31 are hydrophilic and form a loop that extends out from the pore above the heptameric ring surface; these extended hydrophilic loops appear to be only weakly structured, being either too disordered to trace in electron density maps, or stabilized in a variety of conformations by crystal packing and other interactions (Gribun, Kimber et al. 2005; Bewley, Graziano et al. 2006; Szyk and Maurizi 2006). Residues 15 – 21 are hydrophobic and anchor the N-terminus of the axial loop to the head domain mainly by the placement of the side chains of Pro18 and Val20 in a shallow hydrophobic groove formed predominantly by the residues at the N-terminus of αA from two adjacent protomers (Figure 23D). Examination of the electron density map clearly indicates that all axial loops are wholly disordered in ClpP<sup>SS</sup>, with the last ordered residue being Asp32. The absence of clear electron density in this region could, in principle, be due to the 3.2 Å resolution of the electron density maps combined with the intrinsically high temperature factors for these residues as observed in other ClpP structures.

While the axial loops have some plasticity in all ClpP structures solved to date, adopting different conformations in different protomers even when ordered, this plasticity is found N-terminal to Phe31. Phe31 itself in WT ClpP structures is typically held in place by packing the phenyl ring against the side chains of Tyr34 and Ser35 of helix αA of the same subunit, as well
as Met19 and Ile21 of the neighboring subunit (Figure 23D). B-factors for Phe31 in published *E. coli* WT ClpP structures are similar to the overall chain average. Inspection of all 28 copies of ClpPSS subunits in the asymmetric unit revealed no electron density consistent with Phe31 packing in this fashion (Figure 23D). Density for this residue in ClpPSS is very weak, with strong indications of disorder, but some monomers show hints that this residue may form a continuation of helix αA, in a manner analogous to the *M. tuberculosis* (2CE3) (Ingvarsson, Mate et al. 2007) and *P. falciparum* ClpP (2F6I) (Vedadi, Lew et al. 2007) structures. Since both of these structures are in the compact state, this suggests that the heptameric ring in the compact conformation may be at least partially coupled to structural changes in the axial loops. In ClpPSS, the disorder in the N-terminal loops appears to correlate with shifts of the protomers within each heptameric ring, one effect of which is to bring the adjacent αA helices slightly closer together. This pinches closed the hydrophobic grooves in which the very N-terminal residues (typically Pro18 and Val20) of the axial loops bind, leading to the disordering of these residues (Figure 23D). Protomer rotations within the ring provide a mechanism to couple the disulfide cross-linking of axial helices to disorder of the axial loops, and this would result in the loss or reduction of the interaction between ClpP and its cognate ATPase chaperones, which is consistent with the binding data of Figure 22C. This suggests a possible general mechanism by which ClpP coordinates events at the axial pores and the equatorial helices.

**Normal mode analysis suggests that the structure of ClpPSS corresponds to a naturally sampled conformation of WT ClpP**

In order to further analyze the compact ClpPSS structure and its relation to the extended ClpP structures, we performed normal mode analysis (NMA) using the elastic network model (ENM) (Tirion 1996; Bahar, Atilgan et al. 1997; Hinsen 1998). In brief, ENM is a framework for calculating collective motions in biomolecules by assuming a simple network of springs between atoms closer than a specified cut-off distance. The normal modes are then calculated as the eigenvectors of the Hessian matrix that corresponds to perturbations of the input structure. This methodology is widely used for analyzing collective motions, residue fluctuations, and conformational changes of proteins (Case 1994; Ma 2005; Yang, Song et al. 2008). The structures that we used for this analysis are: PDB codes 1TYF (Wang, Hartling et al. 1997), 1YG6 (Bewley, Graziano et al. 2006), 1YG8 (V20A mutant) (Bewley, Graziano et al. 2006) and
2FZS (ClpP containing a covalently bound inhibitor) (Szyk and Maurizi 2006). Multiple structures were selected to ensure that the results obtained are not specific to one ClpP structure in the PDB, but are rather generally applicable. We will collectively refer to these previously solved E. coli ClpP structures as the extended ClpP structures, ClpPex. Normal mode analysis was not carried out on the ClpPSS structure since missing residues in the unstructured equatorial region resulted in unrealistic normal modes.

We calculated the 100 lowest frequency eigenmodes of ClpPex (1TYF, 1YG6, 1YG8 and 2FZS with inhibitor excluded) using the NOMAD web server (Lindahl, Azuara et al. 2006), with the default settings of a distance weight parameter (Hinsen 1998) of 5 Å and a cut-off of 10 Å, and including all atoms in the structures. The six slowest eigenmodes correspond to rigid translations and rotations were discarded. In order to compare the solved structures of ClpPex to the current ClpPSS structure, we calculated the overlap between each eigenmode of ClpPex and the difference vectors, Δq, between the atoms in the experimental structures of ClpPSS (using the tetradecamer with chains O-Z and 1-2; similar results were obtained when using the other tetradecamer in the unit cell, chains A-N) and ClpPex. The overlap, Ij, was calculated as in (Marques and Sanejouand 1995):

\[ I_j = \frac{\sum_i |u_{ij}| \Delta q_i}{\sqrt{\sum_i u_{ij}^2 \sum_i \Delta q_i^2}} \]  

(1)

Where \( u_{ij} \) is the displacement vector of atom \( i \) for eigenmode \( j \). The overlap is a measure of whether a given eigenmode can drive the two structures to a lower r.m.s.d. distance. The result from this calculation, shown in Figure 26A, shows that in all cases, the highest overlap has an \( I_j \) of about 40%. The highest overlap mode is the fifth slowest mode, mode 11, which, for the different structures considered, corresponds to very similar but not identical motions. For all four structures, mode 11 is a breathing motion accompanied with a twisting motion. The twisting motion is more accentuated for 1YG8 and 2FZS. The ClpP dynamics based on mode 11 for 1YG6 is predominantly a breathing motion of the two heptameric rings along the longitudinal axis of the full structure in conjunction with a slight twist. The motion perturbs the equatorial region and suggests the possibility of pores forming. Most of the displacements in the structure appear to result from the subunits moving relative to one another.
Figure 26. ENM normal mode analysis of ClpP.

(A) Shown are the overlaps $I_i$ (top panels) and the eigenmode weights $w_j$ (bottom panels) for each eigenmode of the extended ClpP structures 1TYF, 1YG6, 1YG8, and 2FZS. (B) Scatter plots of the overlap of the eigenvectors between the extended ClpP structures. Each scatter plot shows the overlaps of the 20 slowest eigenvectors (modes 7–26) between the two structures indicated below the plot. The first six eigenvectors that correspond to rigid translations and rotations are excluded. The definition of overlap is given in the text. The color of a pixel in row $k$ and column $l$ represents the overlap between eigenvector $k$ of the first structure and eigenvector $l$ of the second structure. The row and column numbering is in matrix format, so that the top left corner corresponds to the slowest mode under consideration. The top row of the first scatter plot thus contains the overlap of mode 7 of 1TYF with modes 7, 8, … 26 of 1YG6. An overlap of 1 corresponds to identical eigenvectors. In some cases, similar eigenvectors have different eigenvalue ranks, which results in high intensity off-diagonal elements. For example, the fourth slowest mode of 1TYF is similar to the third slowest mode of 1YG8 which is visible as a bright pixel at row 4 column 3 in the second scatter plot (1TYF-1YG8).
We also compared the eigenmodes from the different ClpP<sup>CS</sup> structures by calculating the overlaps between the eigenmode vectors, \(|u_k v_l|/|u_k||v_l|\) where \(u_k\) (\(v_l\)) is the \(k^{\text{th}}\) (\(l^{\text{th}}\)) eigenvector of structure \(u\) (\(v\)). For these calculations, we used the Cα displacement vectors. In Figure 26B, the eigenvector overlaps are shown in the form of matrix scatter plots, where the color of each pixel with matrix index \((k, l)\) represents the overlap between modes \(k\) and \(l\). The results show a very high degree of similarity between the eigenmodes of the different structures, which is to be expected, as the structures are very similar. In particular, the eigenvectors of mode 11 from 1TYF and 1YG6 are virtually identical (overlaps ~0.99). Furthermore, mode 11 of 1TYF and 1YG6 is very similar to mode 12 of 1YG8 and 2FZS (overlaps ~0.8), but there is also appreciable overlap with mode 11 (~0.5 – 0.6) of 1YG8 and 2FZS.

Even though the elastic network model has been shown to be surprisingly accurate despite its simplicity (Bahar, Atilgan et al. 1997; Atilgan, Durell et al. 2001; Bahar and Rader 2005), it is important to note that the collective motions and frequencies rely on a harmonic approximation of the forces. The amplitudes of the motions depend not only on the eigenfrequencies, but also on temperature, initial conditions, and on the effects of crystal structure and solvent (Hinsen 2008). In our case, the normal mode analysis is further complicated by the fact that some amino acids are missing in the experimental structures, which can result in wrong eigenfrequencies and spurious eigenmodes. Some of the motions found by the analysis may be unrealistic due to steric hindrance, which is not taken into account in the NMA calculations. In order to get an indication of the importance of the individual modes, we used the information contained in the \(B\)-factors of the experimental crystal structures. We fit the amplitudes of the displacements individually in the form of a weight, \(w_j\), attributed to each mode, to the experimental temperature factors. The temperature factors, \(B_i\), are given as:

\[
B_i = 8\pi^2 \langle \Delta r_i^2 \rangle \quad (2)
\]

Where \(\langle \Delta r_i^2 \rangle\) is the mean square displacement of atom \(i\), which is related to the eigenmode displacements through (Brooks, Janezic et al. 1995):

\[
\langle \Delta r_i^2 \rangle = \sum_j w_j |u_{i,j}|^2 \quad (3)
\]

where \(u_{i,j}\) is the eigenvector of mode \(j\). In this formulation, the set of eigenmodes is viewed as a
basis set, and the weights $w_j$ are then fit to achieve the best agreement with the experimental temperature factors using non-negative least square minimization (Lawson 1974) of $\left\{ B_{exp} - B_{calc} \right\}^2$. The result of this calculation shows that most of the thermal motions can be attributed to the slowest 16 eigenmodes (modes 7 – 22), with the fifth slowest mode, mode 11, making the highest contribution (Figure 26A), except for 2FZS where it is second highest. The Pearson’s correlation coefficients between the experimental and fitted $B$-factors were 0.69, 0.66, 0.78, and 0.69 for 1TYF, 1YG6, 1YG8 and 2FZS, respectively. Our results can be compared to the average correlation of ~0.6 obtained using the Gaussian Network Model in a comparison between $B_{exp}$ and $B_{calc}$ for 113 protein structures (Kundu, Melton et al. 2002), where the authors scaled the amplitudes with the inverse of the eigenvalues (reciprocal of frequency squared) and fit the global amplitude scale to the experimental data (also see below).

The result of these calculations shows that in all cases except for 2FZS, the mode with the highest overlap is also the mode with the highest contribution to the temperature factors, providing key support to our hypothesis that the ClpP<sup>SS</sup> structure can be viewed as a not-so-infrequent snapshot of the ClpP<sup>ex</sup> structure undergoing thermal motions. The 2FZS structure has a covalently bound inhibitor at the active sites, which may influence the dynamics. The inhibitor was not included in the NMA calculation.

In order to further investigate the relationship between the WT and mutant ClpP structures, we calculated the all-atom r.m.s.d. distance between ClpP<sup>SS</sup> and the normal-mode-perturbed conformations of ClpP<sup>ex</sup>, for the different modes. The atomic positions, $r_i$, of the perturbed conformations were calculated as described previously (Marques and Sanejouand 1995):

$$r_i = r_{0,i} + a_j u_{i,j} \sin\theta \quad (4)$$

Where $r_{0,i}$ is the experimental position of the atom, $a_j$ and $u_{i,j}$ are the amplitude and eigenmode vector of mode $j$, respectively, and $\theta$ is the phase factor, which describes the oscillations around the unperturbed structure. We then calculated the perturbed structures for each mode for different amplitudes and compared each of these structures to ClpP<sup>SS</sup> structure using r.m.s.d. as a measure of similarity. The result of this calculation confirmed the result from the overlaps (Figure 26A): the motions of the modes with the highest overlap cause the structures of ClpP<sup>ex</sup> to become more
similar to that of ClpPSS. The experimental ClpPSS and the ClpPex structures of 1TYF, 1YG6, 1YG8, and 2FZS differ by r.m.s.d. values of 3.61, 3.48, 3.06 and 3.35 Å, respectively. The optimal single normal mode perturbation decreased this value to a minimum of 3.00, 2.70, 2.09, and 2.18 Å, respectively.

As the amplitudes required for these motions were found numerically, we also checked whether the amplitudes are realistic. This was carried out by fitting the overall scale, \( W \), of the motions to the experimental \( B \)-factors using the expression (Brooks, Janezic et al. 1995):

\[
B_i = W \sum_j u_{ij}^2 / \rho_j^2 \quad (5)
\]

where \( \rho_j \) is the frequency of eigenmode \( j \). The relative contributions from the different modes to the \( B \)-factors are inversely proportional to the square of the eigenfrequencies. This would be the appropriate formulation consistent with equipartition of kinetic energy among eigenmodes if we could compute the physical eigenmodes of the structures. The value of \( W \) depends to some extent on the number of modes included in the fit. Here, we included all 94 non-trivial eigenmodes in order to get a rough estimate of \( W \), even though the high frequency modes actually worsened the correlation between the calculated and the experimental \( B \)-factors. The result from this fit showed that the amplitudes of mode 11 were a factor of about 20 to 40 smaller than the optimal amplitudes. Using only the 16 slowest eigenmodes, they were a factor of 10 smaller. It should be noted that amplitudes fit to experimental \( B \)-factors were obtained from crystallized samples at \( \sim 100 \) K, thus the corresponding amplitudes describing ClpP dynamics at \( \sim 300 \) K should be larger. Some of the eigenmodes may also be artifacts due to missing residues in the structures. Nonetheless, even taking these considerations into account, the amplitudes required for fitting the conformational difference between ClpPex and ClpPSS are significantly larger than the amplitudes obtained from fitting the \( B \)-factors (Figure 26A). This implies that if thermal vibrational motions embodied by mode 11 do bridge ClpPex and ClpPSS as we propose here, the vibrational motion cannot be purely harmonic because the energy cost involved would be too high. Most likely the motion should pass over a barrier separating the ClpPex and ClpPSS conformations which are situated at two different free energy minima.

This physical picture of ClpP dynamics is reminiscent of that provided by a straightforward anharmonic interpolation approach to model the ‘switching’ barrier of an
allosteric process (Miyashita, Onuchic et al. 2003). In fact, intuitively, the average absolute values of atomic displacement corresponding to the optimal amplitudes are not large. For 1TYF, 1YG6, 1YG8, and 2FZS, the average of \( a_j \|u_i,j\| \) over \( i \) for mode \( j = 11 \) (see eq. 4) for the optimal \( a_j \) are 1.77, 1.76, 1.90, and 2.15 Å, respectively. In light of the insights gained here, it will be extremely interesting for future investigations to study possible conformational and energetic pathways (Kim, Jernigan et al. 2002; Miyashita, Onuchic et al. 2003) between the extended and compact states of ClpP.

Thus, in summary, the normal mode analysis of four extended ClpP structures shows that a few normal modes have appreciable overlap with the difference vectors to the compact ClpP\textsuperscript{SS} structure. These high overlap modes coincide with the modes that best describe the experimental temperature factors of the extended structures, suggesting that the high overlap modes are likely the dominating large-scale motions. In this view, the ClpP\textsuperscript{SS} structure is stipulated as a probable snapshot of ClpP\textsuperscript{ex} undergoing thermal motion.

In conclusion, our current analysis of the ClpP\textsuperscript{SS} structure provides new insights into the dynamics of this system. Our data support the view that ClpP undergoes several dynamic conformational changes that we propose are part of its functional cycle. Recently published biochemical data also support the presence of such conformational changes in ClpP (Maglica, Kolygo et al. 2009). Our X-ray structure of ClpP\textsuperscript{SS} and normal mode analysis suggest that one dominant motion that ClpP might undergo is a compression motion that is accompanied with a slight rotation in which the two ClpP rings of the tetradecamer are brought closer together. This should result in the reduction of the size of the catalytic chamber of ClpP. Since we had earlier proposed that degraded products are released from ClpP through the formation of transient equatorial side pores (Gribun, Kimber et al. 2005; Sprangers, Gribun et al. 2005), it might be reasonable to speculate that the compression motions depicted by mode 11 might actually allow the exit of the peptides generated after substrate degradation from the ClpP catalytic chamber. Importantly, if the ClpP\textsuperscript{SS} structure represents a conformational state sampled not infrequently by the WT ClpP, then ClpP might undergo switching between an active extended state required for substrate degradation and an inactive compact state supporting product release. A bound ClpX or ClpA chaperone might modulate this conformational switching of ClpP using energy derived from ATP hydrolysis. It is interesting to note that the breathing motion accompanied by slight twisting observed for ClpP is similar to the motions of the slowest normal modes found for the
GroEL-GroES complex (Ma, Sigler et al. 2000; Keskin, Bahar et al. 2002). Other conformational states that, for example, allow substrate translocation into the ClpP chamber, and probably involve motions of the axial loops, must also be sampled by ClpP dynamics. Understanding the dynamics of this system is an important undertaking for gaining insights into its mechanism of function. Future studies will have to combine novel biophysical techniques including NMR spectroscopy and single molecule studies with theoretical considerations in order to properly describe the motions of the ClpP tetradecameric cylinder at the molecular level.
4 Conclusions and future directions
4.1 Conclusions

The work presented in this thesis has given new insights of the biological roles of TF folding chaperone in degradation and also provided a foundation of understanding the dynamics of the ClpP protease. In Chapter 2, I demonstrated that TF physically interacts with ClpX, and that they are functionally associated to enhance phage replication protein λO degradation by ClpXP both in vitro and in vivo. In addition to λO, I found that TF enhances the degradation of about 2% of newly synthesized *E. coli* proteins. We proposed that the TF enhanced degradation is co-translational since TF is a ribosome-associated chaperone. By using a ribosome-stalled λO, I showed that TF assists in λO degradation by ClpXP while being on the ribosome. Experiments suggest that TF transfers λO to ClpX, and λO is subsequently degraded by the ClpP serine protease, demonstrating the existence of co-translational protein degradation in *E. coli*.

Another aspect of my study is on the mechanism of product release of ClpP protease after degradation. In Chapter 3, it is shown that the solved crystal structures of ClpP WT and ClpPSS mutant exhibit different conformations; the WT ClpP has an extended form while ClpPSS has a compact state. We proposed that both conformations are naturally sampled and that ClpP switches dynamically between an active extended state required for substrate degradation and an inactive compact state allowing peptide product release. Our hypothesis is supported by structural data and normal mode analysis.

In both Chapters 2 and 3, there are unanswered questions, and this section proposes a number of experiments aimed to understand the ClpXP system in terms of its dynamics and its interaction with other chaperones.
4.2 Future directions

Identifying ClpXP substrates whose degradation is TF-modulated

In Chapter 2, I found that TF enhances the degradation of about 2% of newly synthesized *E. coli* proteins (Figure 16). To identify the proteins whose ClpXP-mediated degradation was enhanced by TF, I have performed experiments using an inactive His-tagged ClpP to trap endogenous substrates in ∆clpP and ∆clpPΔtig cells (shown in Appendices). The inactive ClpP<sup>trap</sup> was expressed from a plasmid with IPTG in ∆clpP and ∆clpPΔtig cells. Substrates that are translocated into the ClpP<sup>trap</sup>-His chamber are trapped inside the protease chamber but not degraded. The trapped substrates inside ClpP<sup>trap</sup>-His chamber were isolated by Ni purification and then identified by mass spectrometry. If a substrate is trapped in the presence of TF (∆clpP) but not in its absence (∆clpPΔtig), this would indicate that substrate degradation is TF-dependent. I compared the substrates identified from the ∆clpP and ∆clpPΔtig strains, and the substrates that were trapped at least two fold more in ∆clpP than ∆clpPΔtig cells were selected for further confirmation. These proteins can be cloned and purified. Degradation assays can be performed on these substrates in the presence and absence of TF to examine the effect of TF on their degradation by ClpXP. Proteins identified by the *in vivo* trapping experiment and tested using the *in vitro* degradation assay are likely true substrates of ClpXP whose degradation is dependent on TF.

Mapping the interactions between TF and ClpX

My binding studies described in Figure 12 in Chapter 2 indicated that both TF<sub>N</sub> and TF<sub>P</sub> interact with ClpX (Figure 12 and Figure 13). Together with the *in vivo* results which showed that TF enhanced the ClpXP-mediated λO-SecM degradation co-translationally (Figure 16 and Figure 17A,B), we propose that TF acts as a bridge between the ribosome and ClpX (Figure 17C). The tip of N-terminus of TF (GFRxGxxP) is known to interact with the ribosome, and we propose that the P domain binds to ClpX and the C domain interacts with the nascent chain exiting the ribosome peptidyl-tRNA tunnel. In addition, since TF is known to interact with ribosomes in a dynamic cycle, TF might dissociate from the ribosome, and the N and P domains would be both interacting with ClpX. However, the exact binding site is not clear yet. TF is shaped as an elongated molecule where TF<sub>N</sub> interacts with the ribosome at the tip of the N terminus
(GFRxGxxP), we propose that $\text{T}_P$ might interact with ClpX at the tip of the P domain as well. When TF is dissociated from the ribosome, the ribosome binding site at $\text{T}_N$ (GFRxGxxP) may dock onto ClpX. To examine the binding site, the residues on the tip of $\text{T}_P$ (FPEEYHAEN) and $\text{T}_N$ (GFRxGxxP) can be mutated (Figure 27), and the binding between TF mutants and ClpX should be examined by measuring the ATPase activity of ClpX in the presence of the TF mutants. TF mutants that fail to enhance ClpX ATPase activity can be further studied using SPR analysis to verify whether the mutations have abolished the binding between TF and ClpX.
Figure 27. TF structure showing potential binding sites to ClpX.
TF has strong PPIase activity \textit{in vitro} and we speculate that TF binds to the prolines in the PXP motif of ClpXΔZBD. It has been demonstrated that the ZBDs in ClpX undergo large block movement towards ClpP and into the AAA+ ring (Thibault, Tsitrin et al. 2006). The PXP motif in the ClpX was suggested to serve as a hinge region that facilitates such movements. We propose that TF might be acting on the prolines of the PXP motif, which further accelerates the movements of ZBD to AAA+ and thus enhances the translocation degradation rate of substrates. Mutations of ClpX in the PXP motif can be generated where the prolines are mutated to alanines. To examine whether TF binds to the first or/and second proline, three PXP variants should be made: AXP, AXA, PXA. ATPase assay of these ClpX PXP mutants should be performed to ensure that the mutations did not alter their ATP hydrolysis function. The ability of ClpX PXP mutants to bind ClpP and to degrade substrates should also be tested using a standard degradation assay with GFP-SsrA as substrate. If the ClpX PXP mutants behave like the wild type ClpX, the ATPase activity of ClpX mutants can be tested in the presence of TF. The ClpX mutant whose ATPase activity is not stimulated in the presence of TF suggests that the mutant does not interact with TF. Further SPR analysis can be carried out by immobilizing TF on the chip and flowing different ClpX PXP mutants over and identify the interaction between TF and ClpX mutants.

\textbf{Investigating if TF collaborates with other proteases including HslUV, Lon protease, and FtsH to promote co-translational degradation}

I have shown that TF functionally associated with the ClpXP degradation system in Chapter 2 and this association resulted in co-translational degradation of λO protein. In addition, TF was found to enhance the degradation of about 2% of newly synthesized \textit{E. coli} proteins. However, whether these 2% of proteins were all degraded by ClpXP was not clear. TF may interact with other proteases in the cell to promote degradation. The major ATP-dependent protease systems in \textit{E. coli} include ClpXP, ClpAP, HslUV, Lon protease and FtsH. I have demonstrated that TF does not physically interact with ClpA by ATPase assay (Figure 14). To examine if TF functionally associates with other proteases including HslUV, Lon and FtsH, these proteases can be purified and SPR analysis can be performed to examine whether any of these proteases interacts with TF. Since HslU, Lon and FtsH all exhibit ATPase activity, ATPase assay can be performed in the presence and absence of TF to investigate whether TF exerts any effect on their
ATPase activities. The proteases that interact with TF and are affected by TF can be seen as potential targets that TF collaborate with. *In vitro* degradation assays of the substrates of these proteases can be performed with and without TF, and the substrates whose degradation is enhanced should be further studied by *in vivo* degradation assay. If a substrate is identified to be degraded faster in the presence of TF both *in vitro* and *in vivo*, whether the TF-assisted degradation occurs co-translationally should be investigated. Methods described in Chapter 2 can be applied to study systems involved in co-translational degradation.

**Understanding the dynamics of ClpP**

Due to the dynamic nature of ClpP, there are different conformations that ClpP samples. Even though it is not possible to synchronize the conformational dynamics of a population of ClpP molecules required for standard biochemical approaches, single molecule techniques are able to resolve conformational heterogeneities. Förster resonance energy transfer (FRET) can be used to monitor the global dynamics in ClpP. The head domain and the equatorial domain of ClpP should be labeled by a FRET pair. The FRET efficiency is a function if the inter-dye distance, therefore this technique allows us to map the time trajectory of the FRET efficiency to the time trajectory of the inter-dye distance and therefore to the conformational dynamics. The distance between the head and equatorial domains is \( \sim 50 \text{ Å} \). Therefore a FRET pair with a comparable forster radius \( R_0 \) can provide the optimal sensitivity to the changes in the inter-dye distance.

However, there is a major challenge to study ClpP homo-oligomer using single molecule approach. ClpP contains 14 protomers, but only 1 protomer should be labeled with the FRET pair to reduce non-specific background. In order to do so, a fusion protein with 14 ClpP protomers can be made with only one protomer suitable to have the residues for labeling. In the wild type ClpP, there are two cysteines in the head domain which are buried and thus not accessible to dye. A mutant ClpP A153C where the cysteine is at the handle region is accessible to dye (data not shown). Therefore the cysteine mutation A153C can be introduced in the one of the protomers, and a tryptophan mutation can be generated on the surface of the head domain in the same protomer (WT ClpP has no Trp). The tryptophan residue can be used as donor and the cysteine can be conjugated with AEDANS as an acceptor. If a global ClpP dynamics existed, it will modulate the FRET signal as ClpP continues to compress and expand. ClpP is a complicated
system; hence we expect to see complex kinetics. The fundamental understanding of the two different states may be useful for further investigating the molecular mechanism of substrate translocation and product release.
References


Appendices

Identification of ClpXP substrates whose degradation may be enhanced by TF using in vivo ClpP\textsuperscript{trap}

Data Attribution: Mr. Adedeji Ologbenla and I performed the pull down experiments. Dr. Mohan Babu identified the proteins by mass spectrometry.

To identify more substrates whose ClpXP degradation is modulated by TF, a ClpP\textsuperscript{trap} mutant was constructed by mutating the active site serine to alanine. A C-terminal His tag was added to ClpP\textsuperscript{trap} for easy isolation. ClpP\textsuperscript{trap}-His was expressed in ΔclpP and ΔclpPΔtig strains. ClpP\textsuperscript{trap}-His can still interact with ClpX (Flynn, Neher et al. 2003). Hence, substrates that are translocated into the ClpP\textsuperscript{trap}-His chamber are trapped inside the protease chamber but not degraded. After Ni-column purification, the ClpP\textsuperscript{trap}-His and the trapped substrates were isolated and subsequently identified by mass spectrometry. I compared the substrates identified from the ΔclpP and ΔclpPΔtig strains. The substrates that were trapped at least two fold more in ΔclpP than ΔclpPΔtig cells are listed in the table below.

<table>
<thead>
<tr>
<th>gene name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>adhE</td>
<td>fused acetaldehyde-CoA dehydrogenase and iron-dependent alcohol dehydrogenase and pyruvate-formate lyase deactivase</td>
</tr>
<tr>
<td>cbpA</td>
<td>curved DNA-binding protein, DnaJ homologue that functions as a co-chaperone of DnaK</td>
</tr>
<tr>
<td>cheA</td>
<td>fused chemotactic sensory histidine kinase (soluble) in two-component regulatory system with CheB and CheY</td>
</tr>
<tr>
<td>clpX</td>
<td>ATPase and specificity subunit of ClpX-ClpP ATP-dependent serine protease</td>
</tr>
<tr>
<td>dnaK</td>
<td>chaperone Hsp70, co-chaperone with DnaJ</td>
</tr>
<tr>
<td>dps</td>
<td>Fe-binding and storage protein</td>
</tr>
<tr>
<td>fbaA</td>
<td>fructose-bisphosphate aldolase, class II</td>
</tr>
<tr>
<td>frdA</td>
<td>fumarate reductase (anaerobic) catalytic and NAD/flavoprotein subunit</td>
</tr>
<tr>
<td>gapA</td>
<td>glyceraldehyde-3-phosphate dehydrogenase A</td>
</tr>
<tr>
<td>hypB</td>
<td>GTP hydrolase involved in nickel liganding into hydrogenases</td>
</tr>
<tr>
<td>iscS</td>
<td>cysteine desulfurase (tRNA sulfurtransferase), PLP-dependent</td>
</tr>
<tr>
<td>lexA</td>
<td>DNA-binding transcriptional repressor</td>
</tr>
<tr>
<td>pepT</td>
<td>peptidase T</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>poxB</td>
<td>pyruvate dehydrogenase (pyruvate oxidase), thiamin-dependent, FAD-binding</td>
</tr>
<tr>
<td>nirD</td>
<td>nitrite reductase, NAD(P)H-binding, small subunit</td>
</tr>
<tr>
<td>rpoC</td>
<td>RNA polymerase, beta prime subunit</td>
</tr>
<tr>
<td>rpoS</td>
<td>RNA polymerase, sigma S (sigma 38) factor</td>
</tr>
<tr>
<td>wrbA</td>
<td>predicted flavoprotein in Trp regulation</td>
</tr>
<tr>
<td>yqhD</td>
<td>alcohol dehydrogenase, NAD(P)-dependent</td>
</tr>
</tbody>
</table>