Biochemical Studies of the Essential Clp protease in Cyanobacteria and its Associated Adaptor Proteins

Anders Tryggvesson

GÖTEBORGS UNIVERSITET

FACULTY OF SCIENCE
DEPARTMENT OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES

Akademisk avhandling för filosofie doktorsexamen i Naturvetenskap med inriktning Biologi, som med tillstånd från Naturvetenskapliga fakulteten kommer att offentligt försvaras fredagen den 24 oktober 2014 kl. 10.00 i Hörsalen, Institutionen för biologi och miljövetenskap, Carl Skottsbergs gata 22B, Göteborg.

Examinator: Professor Cornelia Spetea Wiklund, Institutionen för biologi och miljövetenskap, Göteborgs Universitet
Fakultetsopponent: Professor Christiane Funk, Kemiska Institutionen, Umeå Universitet.

ISBN 978-91-85529-70-4

ISBN: 978-91-85529-70-4

Tryck: Kompendiet AB

Epublicerad: http://hdl.handle.net/2077/36799

Till Eva och Sten

Biochemical Studies of the Essential Clp protease in Cyanobacteria and its Associated Adaptor Proteins

Anders Tryggvesson
Gothenburg university, Department of Biological and Enviromental Sciences Box 461, SE-405 30 Gotheburg Sweden.

Abstract

Proteins are an essential part of all organisms and are involved in many cellular processes. To regulate the function of proteins and facilitate their removal when damaged or otherwise compromised, sophisticated control systems have evolved that include molecular chaperones and proteases. These regulatory proteins attempt to repair damaged polypeptides and if necessary degrade them before they can interfere with cellular activities. Clp/Hsp100 proteins are a family of chaperones that belong to the broader family of AAA+ proteins (ATPases associated with diverse cellular activities) that are present in a wide range of organisms. Many of these AAA+ Clp proteins function as the chaperone partner within Clp proteases, conferring substrate specificity and transferring the unfolded protein substrate to the proteolytic component for degradation. The Clp chaperones form single hexameric rings that associate to a proteolytic complex consisting of two opposing heptameric rings comprised typically of a single type of subunit, ClpP. The catalytic sites of these ClpP subunits are sequestered within the tetradecamer to avoid inadvertent protein degradation.The Clp protease in *E. coli* is the best studied Clp proteases to date, with two distinct types depending on if the chaperone partner is ClpA or ClpX. Also present are adaptor proteins that modify the substrate specificity of the chaperone component, such as ClpS that redirects the ClpAP protease to degrade N-end rule substrates. Although Clp proteins are found in a wide range of organisms, those in photosynthetic organisms such as cyanobacteria and vascular plants are by far the most numerous and diverse. In the cyanobacterium *Synechococcus elongatus* (*Synechococcus*) two types of mixed Clp proteolytic cores exist; ClpP3/R and ClpP1/P2. The ClpP3/R core associates to the chaperone ClpC to produce a protease that is essential for phototrophic growth, whereas ClpP1/P2 binds to ClpX to form a second Clp protease whose activity is non-essential. This thesis work has examined the structure and function of the mixed proteolytic core within the essential ClpCP3/R protease and its associated adapter proteins ClpS1 and ClpS2. This has been done using molecular and biochemical methods to purify recombinant versions of each Clp protein or complex and analyzing them *in vitro*. In **Paper I**, the ClpP3/R complex was over-expressed in *E. coli* and purified by column chromatography. The proteolytic core was shown to consist of two identical heptameric rings, each with three ClpP3 and four ClpR subunits in an alternating configuration. The ClpR subunit is catalytically inactive but its inclusion within the ClpP3/R core did not appear rate-limiting for the activity of the ClpCP3/R protease. The general architecture of ClpP3/R mirrored that of the proteolytic core within the eukaryotic 26S proteasome, with three active and four inactive subunits in the central heptameric rings. A model of ClpP3/R was also presented in this paper, along with the finding that the ClpS1 adaptor protein binds to ClpC and modifies its substrate specificity. In **Paper II**, two N-terminal regions in ClpR (the Tyr- and Pro motifs) and one in ClpP3 (the MPIG motif) were shown to be important for the interaction with ClpC and correct assembly of the ClpP3/R tetradecamer. We also identified a motif in the C-terminal region of ClpC (the R-motif) that confers the specific association to the ClpP3/R core. In **Paper III**, we investigated the essential adaptor protein ClpS2 that is so far unique to cyanobacteria. A recombinant version of ClpS2 was purified and its activity compared to that of ClpS1. Like ClpS1, ClpS2 binds to ClpC and alters its substrate specificity. However, ClpS1 and ClpS2 recognize different destabilizing residues and thus target different N-end rule substrates for degradation by the ClpCP3/R protease. Overall, this thesis provides new insights into the structure and function of the essential ClpCP3/R protease in cyanobacteria and how its substrate specificity is modified by the ClpS adaptor proteins.

ISBN: 978-91-85529-70-4

List of publications

This thesis is based on the following papers which are referred to by their Roman numerals in the text:

I. Andersson F. I., **Tryggvesson A.**, Sharon M., Diemand A. V, Classen M., Best C., Schmidt R., Schelin J., Stanne T.M., Bukau B., Robinson C.V., Witt S., Mogk A., Clarke A. K. (2009). Structure and function of a novel type of ATP-dependent Clp protease. *J. Biol Chem. 284*(20), 13519–32.*

II. **Tryggvesson A**.[1], Ståhlberg F.M.[1], Mogk A., Zeth K. and Clarke A.K. (2012). Interaction specificity between the chaperone and proteolytic components of the cyanobacterial Clp protease. Biochem. J. 446(2):311-20.**

III. **Tryggvesson A**., Ståhlberg F.M., Töpel M., Tanabe N., Mogk A. and Clarke A.K. Characterization of ClpS2, an essential adaptor protein for the cyanobacterium *Synechococcus elongatus. Manuscript*

1 Both authours contributed equally to this work

* Reprinted with permission from The Journal of Biological Chemistry

** Reprinted with permission from Biochemical Journal

TABLE OF CONTENTS

1.Introduction.....

1.1 Molecular chaperones.....

1.2 Proteases.....

1.3 AAA+ proteins and associated proteases .....

1.4 Clp/Hsp100 proteins.....

1.4.1 ClpA .....

1.4.2 ClpB.....

1.4.3 ClpC.....

1.4.4 ClpD.....

1.4.5 Clp E.....

1.4.6 ClpL and ClpV.....

1.4.7 ClpX.....

1.4.8 ClpY.....

1.5 ATP-dependent proteases .....

1.5.1 26S proteasome.....

1.5.2 N-end rule degradation.....

1.5.3 Lon protease.....11

1.5.4 FtsH protease 12

1.5.5 HsIUV protease 13

1.5.6 Clp-protease.....13

1.6 Clp-proteases and associated adaptor proteins in *E.coli*.....15

1.6.1 ClpAP.....15

1.6.2 ClpS.....16

1.6.3 ClpXP.....18

1.6.4 Adaptor proteins for ClpX.....20

1.7 Cyanobacteria 20

1.7.1 Cyanobacterial Clp proteins 21

1.8 Clp proteases in Plants.....22

2.Aim of the thesis 22

3.Results and discussion 23

3.1 Composition of the ClpP3/R complex.....23

3.2 Inactivation of ClpP3 and reactivation of ClpR.....24

3.3 Modeling of ClpP3 and ClpR.....26

3.4 Analysis of the N-terminal region of ClpP3 and ClpR.....27

3.5 Core specificity factor in ClpC.....29

3.6 The ClpS adaptor proteins.....30
3.6.1 Identification of ClpS2.....31
3.6.2 Quantification of ClpS2 proteins *in vivo*.....32
3.6.3 Substrate specificity of ClpS1 and ClpS2.....32
3.6.4 Conserved amino acids regions in ClpS2.....33
3.6.5 Potential substrates for ClpS2.....34
4 Future prospects and remaining questions.....35
4.1 Structure of ClpP3/R and the ClpS adaptor proteins.....35
4.2 *In vivo* substrates for ClpP3/R.....36
4.3 *In vivo* substrates for ClpS1 and ClpS2.....36
4.4 Is ClpS2 degraded by the ClpXP!/P2 protease *in vivo*?.....37
4.5 Which regions of ClpS1 and ClpS2 are important for substrate recognition. .37
5.References 38
6.Populärvetenskaplig sammanfattning.....54
7.Acknowledgements.....57

1. Introduction

Proteins are essential for the functioning of all living organisms. They can be considered the machinery of the cell, participating in almost all metabolic and regulatory processes. Their importance to cell homeostasis and viability makes it necessary to closely monitor protein integrity, and thus protein turnover is tightly controlled throughout the lifetime of the cell. An intricate surveillance mechanism has evolved to help maintain the cellular protein environment and detect inactive proteins arising from synthetic errors, chemical damages, or protein misfolding and aggregation. Changes in the environment can also have a significant impact on the functioning of many proteins. Factors such as temperature extremes, increased salinity, desiccation, exposure to toxic pollutants are all well known to compromise protein integrity and activity and as such the ability to respond to such fluctuating growth conditions is crucial for cell survival. Two important components of this protein surveillance and maintenance system are molecular chaperones and proteases.

1.1. Molecular chaperones

The cell matrix is not simply an aqueous solution but more of a viscous “soup”, a crowded environment where all parts of the cellular machinery must function. Concentrations of RNA and proteins within cells are thought to be as high as 340 mg/ml (Zimmerman and Minton 1993). As a consequence, it is crucial proteins that have lost activity or are otherwise damaged are rapidly detected and efficiently removed to prevent them from interacting with non-specific targets which could threaten cell homeostasis. Molecular chaperones are a group of proteins that recognize polypeptides that have been damaged or have some problem with their native structure. By binding to these impaired proteins, chaperones prevent their denaturation and potential aggregation, thereby minimizing their interference on functional enzymes and regulatory proteins. Chaperones are crucial for protein folding as well, helping many newly synthesized proteins acquire their correct tertiary structure. (Ellis. 2006). They also facilitate the assembly of certain multi-subunit protein complexes, a role for which they were first identified in promoting the oligomerization of the Rubisco large and small subunits. Certain chaperones can also work in concert with proteases. This means that if the chaperone fails to stabilize the damaged protein and return it to its functional state, it can instead unfold the compromised protein and deliver it to the protease for degradation.

1.2. Proteases

Proteases are enzymes that degrade other proteins by cleaving (hydrolyzing) the peptide bonds within the polypeptide chain. They are an integral part of cellular protein turnover and their function is vital in maintaining proteostasis within all organisms. Proteolysis is required for removing those proteins that have reached the end of their useful lifespan, or that have become irreversibly damaged by chemical or structural modifications. Proteases also target key proteins within different regulatory systems such as certain transcription factors that modulate gene expression.

Proteases are generally divided into two groups: exopeptidases and endopeptidases. Exopeptidases cleave the terminal peptide bond at the end of the polypeptide chain, while endopeptidases hydrolyze the internal peptide linkages within the protein (Beynon and Bond 2001). Proteases are further classified by their mode of action and active sites, with the six main types being serine, threonine, cysteine, aspartic acid, glutamic acid and metalloproteases (reviewed by Lopez-Otín and Bond 2008). Another defining characteristic of different proteases is their reliance on energy. Some are dependent on the hydrolysis of ATP for their proteolytic activity, whereas others such as Deg and SppA function independent of ATP. Of the energy-dependent proteases, the best characterized are the 26S proteasome in eukaryotes and the bacterial FtsH, Lon and Clp proteases.

Protein degradation is a tightly regulated process that would have dire consequences for a cell if it somehow malfunctioned. If proteases degraded any polypeptide they came into contact with, many cellular processes would be detrimentally affected and likely result in cell death. As a consequence, the activity of most proteases is controlled to avoid inadvertent proteolysis. One of the best examples of such regulated proteases is the AAA+ (ATPases associated with diverse cellular activities) family that include a chaperone activity along with proteolytic one. The chaperone component of the AAA+ protease unfolds the target protein by hydrolysis of ATP and then degrades it into smaller fragments for recycling (Wickner et al. 1999, Sauer et al. 2004).

1.3 AAA+ proteins and associated proteases

AAA+ proteins are associated with a vast number of processes in the cell, and members of this family have been found in virtually all organisms studied to date (Neuwalt et al. 1999). The largest number of AAA+ proteins found so far is from the plant *Arabidopsis thaliana*, with about 140 members. AAA+ proteins all share a common region of ca. 220 amino acids called the AAA region or nucleotide-binding region (NBR). This region contains the Walker A and B motifs that bind and hydrolyze ATP, respectively (Dougan et al. 2002A). There are also domains that are specific for the different groups of AAA+

proteins depending on their exact function, such as substrate specificity. AAA+ proteins are involved in processes such as DNA replication, heat stress adaptation, membrane transport and protein turnover (Sanchez et al. 1990, Tomoyasu et al. 1995, Schirmer et al. 1996, Chaney et al. 2001). A number of human diseases are connected to malfunctions in different AAA+-proteins, illustrating their importance for cell homeostasis and function (reviewed by Hanson and Whiteheart 2005).

AAA+ proteins are sometimes referred to as molecular motors. Their mode of action uses ATP hydrolysis to drive conformational changes within proteins, which in turn promotes the functional process of the ATPase. Often this process is folding or unfolding of other proteins. The mechanism for protein unfolding used by the Clp family (referred to as translocation coupled unfolding; Sauer et al. 2004, Baker et al. 2006) is one of the best studied of all AAA+ proteins and will be described later in more detail in section **1.6.3**

AAA+ proteases are degradative enzymes that incorporate the unfoldase activity of AAA+ proteins. They can be divided into two groups depending on whether the proteolytic and unfoldase activities are separated to different subunits or located as domains within the same polypeptide. The Clp protease is one of the best characterized of the former group and those in *E.coli* are the most extensively studied (Reviewed by Gottesman 1996). Since the unfoldase component of the Clp protease are also a recognized family of molecular chaperones (Hsp100), a brief description of the different Clp-AAA+ proteins will first be given.

1.4. Clp/Hsp100 proteins

The Clp/Hsp100 family of molecular chaperones plays an important role in many different organisms. They are divided into two classes. Class I includes the subtypes ClpA to ClpE and ClpL and ClpV, and all have two AAA regions designated D1 and D2 that each contain a Walker A and B motif. In contrast, the Class II subtypes ClpX and ClpY contain only one AAA region and are therefore considerably smaller than the Class I proteins.

1.4.1. ClpA

The ClpA protein is present only in Gram-negative bacteria, such as *E. coli* in which it was first discovered. *E. coli* ClpA is a protein of 84 kDa. Its functional state is a hexamer that requires binding of ATP for oligomerization (Maurizi 1991, Singh and Maurizi, 1994, Kessel et al. 1995). ClpA can function separately as a chaperone as demonstrated by its ability to refold RepA into its active form (Pak and Wickner 1997). In *E. coli*, ClpA is not essential for normal growth, and inactivation of the *clpA* gene produces no obvious phenotypic changes (Katamaya et al. 1988). ClpA also has a conserved motif within the

C-terminal region known as the P-loop that mediates binding to the ClpP peptidase, a role that will be discussed further in a later section (see section **1.6.1**).

1.4.2. ClpB

ClpB is a heat-shock inducible chaperone in both bacteria and most eukaryotes that functions to dissolve aggregated proteins during heat stress (Weibezhan et al. 2004). It lacks the P-loop motif necessary for ClpP association and is thus only active as a chaperone (Kim et al. 2001). In most eubacteria including the cyanobacterium *Synechococcus*, two forms of ClpB are produced from a single gene due to a second translation start just upstream of the first AAA domain (Eriksson and Clarke 1996). *Synechococcus* also has a second *clpB* gene that codes for an unusual type of ClpB protein, one that is not heat shock inducible but whose function is essential for constitutive growth and cell viability (Eriksson et al 2001). Plants also possess multiple ClpB proteins, such as in *Arabidopsis* which has three paralogs localized in the chloroplast, mitochondria or cytosol.

1.4.3. ClpC

ClpC is the counterpart to ClpA in Gram-positive bacteria, cyanobacteria, algae and plants. It has been extensively characterized in *Bacillus subtilis*, in which it is important for acquired thermotolerance but not necessary for constitutive growth. *Bacillus* ClpC (BsClpC) has chaperone activity *in vitro* and can refold denatured polypeptides, while also mediating proteolysis in association with ClpP (Krüger et al. 1994, Kirstein et al. 2006). The functioning of BsClpC, however, requires the adaptor protein MecA for oligomerization and its interaction with the ClpP proteolytic partner (Turgay et al. 1997). In comparison, ClpC in *Synechococcus* (SyClpC) is a constitutively expressed protein that is essential for cell viability (Clarke and Eriksson 1996). Earlier studies by our group demonstrated that SyClpC displays chaperone activity *in vitro* by the refolding and reactivation of heat-aggregated proteins (Andersson et al. 2006). Vascular plants such as *Arabidopsis* commonly have two ClpC paralogs (ClpC1 and ClpC2) localized in the chloroplast. The two ClpC proteins are almost identical to each other and ca 90% similar to SyClpC (Zheng et al. 2002). The combined activity of ClpC1 and ClpC2 in *Arabidopsis* is essential for plant viability. Deletion of the more abundant ClpC1 causes a chlorotic leaf phenotype and growth retardation (Sjögren et al. 2004, Constan et al. 2004) whereas loss of ClpC2 produces no visible phenotypic changes (Park and Rodelmel 2004). Besides being primarily a stromal protein, *Arabidopsis* ClpC is also bound to the inner envelope membrane in association with the protein import system . It is thought that ClpC functions as the motor protein that drives the translocation of preproteins across the inner envelope membrane (Flores-Peréz and Jarvis 2013,

Schwenkert et al. 2011), although it has been more recently suggested that it might have additional roles related to proteolysis (Sjögren et al. 2014).

1.4.4. ClpD

ClpD is a closely related variant of ClpC that is only found in chloroplasts of vascular plants. Its exact function is still largely unknown although it does possess all the functional regions of an AAA+ protein, including the P-loop motif for ClpP interaction (Fig 1). It was originally identified as the dehydration-inducible protein ERD1 (Weaver et al. 1999) and is also upregulated by other stresses such as salinity and cold temperature, as well as during senescence (Zheng et al. 2002). Purified recombinant *Arabidopsis* ClpD exhibits chaperone activity by the refolding of aggregated luciferase *in vitro* (Rosano et al. 2011) but native protein substrates have yet to be identified. More recently, the amount of ClpD was shown to increase during leaf development compared to that of ClpC1 and ClpC2, suggesting that these different Hsp100 chaperones might bind different types of substrates (Sjögren et al. 2014).

1.4.5. ClpE

ClpE is a class I Hsp100 that is only present in certain Gram-positive bacteria, including many known pathogens. Besides the two AAA domains, ClpE also contains a N-terminal zinc finger motif that are essential for ATPase activity. In *B. subtilis*, ClpE is involved in protein quality control and promotes the degradation of the repressor protein CtsR, which regulates the expression of *clp* genes (Derré et al. 1999, Miethke et al. 2006).

1.4.6. ClpL and ClpV

ClpL and ClpV are both present in pathogenic bacteria (Pietrosiuk et al. 2011). ClpV is most similar to ClpB and is important for protein secretion in the type VI pathway but lacks the ability to resolubilize aggregated proteins *in vitro* (Pietrosiuk et al. 2011). ClpL is widely distributed in Gram-positive bacteria but absent from Gram-negative species. It is involved in various cellular functions including stress tolerance and virulence in pathogenic bacteria. As a chaperone, ClpL is essential for the correct folding of CtsR and prevents protein aggregation (Tau and Biswas 2013). The activity of ClpL has also been demonstrated to increase tolerance to penicillin in *Streptococcus pneumoniae* by affecting cell wall enzymes (Tran et al. 2011).

1.4.7. ClpX

ClpX is a Clp ATPase containing one AAA-motif, which more closely resembles the AAA-2 domain of ClpA. Since it has only one AAA domain ClpX is smaller (48 kDa) than class I

Hsp100 proteins. ClpX also has a zinc-binding motif at the N-terminus that is crucial for several key functions (see section **1.6.3)**(Banecki et al. 2001, Wojtyra et al. 2003). ClpX acts as an independent chaperone (reviewed by Burton and Baker. 2005) and is capable of refolding aggregated proteins formed during heat stress (Wawrzynow et al. 1995). *Synechococcus* ClpX is essential for phototrophic growth (Schelin et al 2002), whereas ClpX also appears to play an important role in mitochondria of mammals and vascular plants (Table).

1.4.8. ClpY

ClpY (also referred to as HslU) is an AAA+ protein that is part of the HslUV protease, generally considered the prokaryotic equivalent of the 26S proteasome in certain bacteria including *E. coli* and *B. subtilis*. It contains a single AAA domain and oligomerizes into a hexamer with defined chaperone activities (Kessel et al. 1996, Rohrwild et al. 1996, Seong et al. 2000).

1.5. ATP-dependent proteases

Most of the Clp ATPases described above are regulatory components of AAA+ proteases, which have been extensively studied and classified in *E. coli* (Gottesman 1996) These enzymes can be divided into two major groups depending on their architecture (Fig 1), whether the proteolytic and ATPase activities are separated to distinct polypeptides or as two domains within the one protein. Examples of the latter type are the Lon and FtsH proteases (Goldberg et al. 1994), while the Clp protease and 26S proteasome are well-known examples of the former **(**Gottesman et al. 1998, Porankiewicz et al. 1999). Of all these different proteases, it is the 26S proteasome in eukaryotes that is arguably the most important (Lupas et al. 1997).

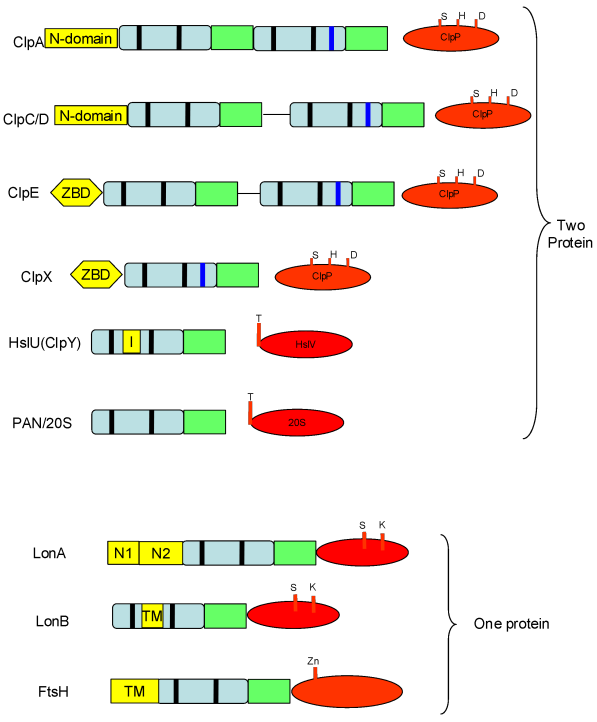


Figure 1. Types of AAA+ proteases. The two major groups of AAA+ proteases are those in which the proteolytic and ATPase activities are present on separate subunits (two protein) or separated into unique domains within the one polypeptide (one protein). Marked are regions of interest: Greyish green is the large AAA+ domain and the bright green is the small AAA+ domain. Also marked are Walker A and B motifs (black lines), the P-loop for ClpP association (blue line), and the proteolytic component/domain (red) with the active site amino acids indicated in single letter code (or chemical symbol in the case of zinc). Also shown in yellow are specific regions within certain Hsp100 chaperone partners; conserved N-terminal domain (ClpA, C-D), zinc-binding domain (ZBD, ClpE and –X), transmembrane domain (LonB, FtsH) Accessory domain in HsIUV (I domain) and the N1 and N2 domains in Lon A (adapted from Sauer and Baker 2011and Gyr et al. 2013).

1.5.1. 26S Proteasome

The 26S proteasome is a large proteolytic complex present in the nucleus and cytosol of all eukaryotes. It is considered the most important machinery for protein degradation and performs many crucial regulatory roles. It performs both housekeeping turnover of proteins and the more specific degradation of key enzymes and regulatory polypeptides (Glickman et al. 1998, reviewed by Baumester et al 1998, reviwed by Finley 2009). Proteins degraded by the proteasome are first marked by the addition of at least four

monomers of ubiquitin (Ub), a 76 amino acid protein that is conjugated to the N-terminal of the substrate. Ubiquitination is dependent on a series of enzymes named E1, E2 and E3 (Fig2 B). E1 is an ubiquitin-activating enzyme that hydrolyzes ATP to activate Ub and then transfer it to the E2 ubiquitin-conjugating enzyme. It is the E3 Ub ligases that identify and bind the protein targets, and then transfer the Ub from E2 to the N-terminus of the bound protein. E3 Ub ligases often have a narrow substrate-specificity, which is why more than a hundred different paralogs typically exist in most eukaryotes (reviewed by Ravid and Hochstrasser 2008). It should also be mentioned that some studies suggest certain proteins can be degraded by the 26S proteasome directly without the need for ubiquitination (Baugth et al. 2009).

The 26S proteasome consists of two parts, 19S and 20S (Fig. 2A). The 19S component is the regulatory particle (RP), often referred to as the "cap" since it controls access to the proteolytic active sites located in the 20S particle. The role of the 19S particle is to recognize ubiquitinated proteins, unfold and de-ubiquitinate them (allowing the Ub tag to be recycled) and then transfer the substrates to the 20S particle for degradation. The 19S RP is comprised of 19 proteins and is divided into a base that associates to the 20S component and a lid (Glickman et al. 1998). The base consists of ten subunits, six of which are ATPases required for the unfolding of substrates and their transfer into the proteolytic cavity by interaction with residues in the α-ring (Smith et al. 2007). The lid is thought to recognise substrates and remove the Ub tags before the substrate is fed into the degradation chamber (Fig. 2A). The pore of the 20S part is narrow requiring that substrates are unfolded by the ATPase part before they can be degraded.

The 20S particle consists of two double rings each of seven subunits consisting of α and β subunits. The subunits are arranged in a $\alpha\beta\beta\alpha$ pattern with α-subunits in contact with the 19S particle and the β subunits containing the proteolytically active sites in the centre. The proteolytic activity of the active β-subunits are of the threonine type. Of the seven β-subunits only three contain proteolytically active sites (Fig. 2). In eukaryotes, the α and β rings consist of seven different subunits (Groll et al. 1997), while the 20S particles in archaea generally contain α and β rings consisting of identical subunits (Zwickl et al. 1998, Löwe et al. 1995). These bacteria lack the 19S particle, but some have complexes that have similarities with the ATPase subunits of the base of the 19S particle which might fill the role of unfolding substrate (Forouzan et al. 2012). The ATPase partner for the prokaryotic 20S proteasome is called PAN (Proteasome Activating Nuclease) and docking of PAN to the 20S proteasome opens up the pore (Smith et al. 2006, 2007). Some analogues to the ubiquitination pathway have been found in bacteria. A prokaryotic Ub-like protein (PuP) exists in *Mycobacterium tuberculosis* where it targets substrate to the prokaryotic proteasome (Darwin 2009,

Burns and Darwin 2010). Another protein that mimics the function of Ub is SAMP (Ub Small Archeal Modifying Proteins), which has been found in several archael species (Humbard et al. 2010, Miranda et al. 2011, Hepowit et al. 2012). The process of ubiquitination of substrates is part of the degradation pathway termed the N-end rule degradation pathway.

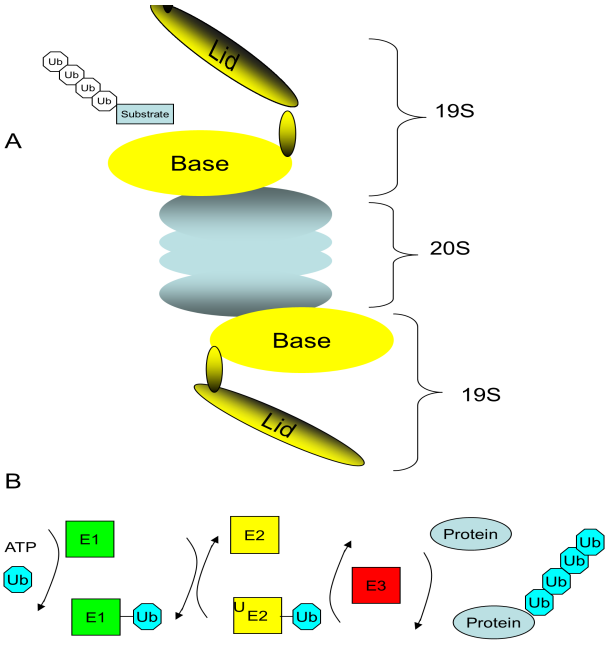


Figure 2. The 26S proteasome. A. Overall layout of the 26S proteasome. The regulatory particle (19S) recognizes Ub-tagged proteins that are unfolded and translocated into the proteolytic part (20S) (adapted from Mogk et al. 2007). **B.** Ubiquitination in eukaryotes. Ub is attached to E1 (Ub-activating enzyme) by the hydrolysis of ATP and then transferred to the Ub-conjugating enzyme E2. Ub is then transferred to the protein substrate *via* the E3 Ub-ligase which recognizes specific residues in the target polypeptide (adapted from Varshavsky 2011).

1.5.2. N-end rule degradation

The N-end rule degradation pathway was first discovered in yeast when it was observed that the *in vivo* stability of proteins was dependent on the type of amino acid at the N-terminus (Bachmar et al. 1986). It was shown that some proteins *in vivo* were rapidly turned over while others were stable for long periods of time. It was found that this pathway was one of the regulators for ubiquitination of protein substrates for the

proteasome. N-end rule degradation has now been identified in a wide range of organisms from *E. coli* to higher eukaryotes (Gonda et al. 1989, Tobias et al. 1991, Bachmar et al. 1993). The amino acids responsible for destabilizing proteins based on the N-end rule principle are referred to as N-degrons (Varshavsky 2003). Certain amino acids are recognized directly by the adaptors that promote their degradation and are referred to as "primary destabilizing amino acids". Other amino acids can be recognized by modifying enzymes that add an amino acid that functions as a primary degron to the N-terminus; these are referred to as secondary destabilizing amino acids. The modification involves conjugation of an amino acid by an amino acid transferase. In eukaryotes, some residues can also function as tertiary degrons (Fig. 3A).

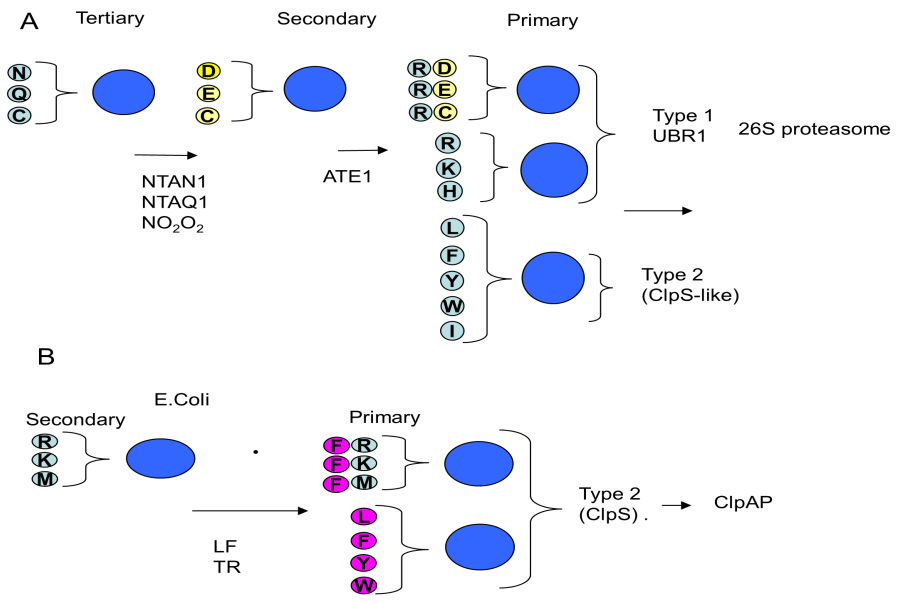


Figure 3. Principle of N-end rule degradation in *E. coli* and one of the mechanisms in eukaryotes. A In eukaryotes, the amino acids N, Q and C can function as tertiary destabilizing residues. They can be converted to secondary destabilizing residues by N-terminal amidases (NTAN1 or NTAQ) or by NO_2/O_2. A primary destabilizing residue is added to these by ATE1 (arginyal-tRNA-transferease). Ultimate recognition by the binding pocket is dependent on which type of residue that the degron carries. **B** In *E.coli*, secondary N-degrons are modified by L/F-tRNA transferase to primary N-recognins by addition of Phe. Recognition of all N-end rule proteins in *E.coli* is by the adaptor ClpS, which alters the substrate specificity of the ClpAP protease to recognize N-end rule substrates. (adapted from Dougan et al. 2010**)**

The amino acids that functions as primary degrons in eukaryotes are Arg, Lys, His, Leu, Phe, Tyr, Trp. These residues are directly recognized by E3 ligases, ubiquitinated and then degraded by the 26S proteasome. This pathway of ubiquitination is termed the Arg/N-end rule pathway. A more recently discovered pathway known as Ac/N-end rule involves the Nt-acetylation of the N-terminal residue of the protein that then functions as a recognition signal for ubiquitination (Hwang et al. 2010). There are two types of recognition sites in the N-recognins. Type 2 recognizes bulky hydrophobic amino acids (Ile, Leu, Phe, Tyr and Trp) and is referred to as the bacterial ClpS domain. Type 1 recognizes basic destabilizing residues (Arg, Lys, and His) and is termed the UBR-box (Tasaki et al. 2005, Dougan et al. 2012). A detailed review of N-end rule degradation in both prokaryotes and eukaryotes has been given by Varshavsky (2011).

In prokaryotes lacking the ubiquitination system, N-end rule degradation is regulated by the adaptor protein ClpS which directs substrates to degradation by the ClpAP protease in *E.coli* (Dougan et al. 2002b). Most bacterial ClpS proteins recognize the primary degrons Leu, Phe, Tyr and Trp, whereas Arg and Lys function as secondary degrons that can be modified by addition of Phe or Leu by L/F-transferease (Ninnis et al. 2009, Humbard et al. 2013). The function of ClpS will be described in more detail in section **1.6.2**.

1.5.3. Lon protease

Lon is one of the first ATP-dependent peptidase discovered in bacteria (Charette et al. 1981) and is now known to degrade many specific polypeptides as well as contribute to the general quality control of cellular proteins. *E. coli* Lon is an 87 kDa monomer that oligomerizes into a functional hexamer (Botos et al. 2004, Cha et al. 2010). The complex forms a chamber within which the Serine-type proteolytically active sites (a Ser-Lys dyad) are located. The AAA+ part of Lon unfolds bound substrates and feeds them into the proteolytic chamber for degradation. Lon functions as an endopeptidase and produces fragments of 3-20 amino acids in length upon digestion of substrates.

Lon protease is involved in the heat-shock response. Although characterized mostly in *E. coli*, Lon is also present in almost all bacteria and many eukaryotes. In eukaryotes, Lon orthologs have been found in mitochondria, peroxisomes and plastids (Ostersetzer et al. 2007). Two forms of the protease exist, called Lon A and Lon B. Lon A is present in most eubacteria, while Lon B can be found in archaea. (Fig. 1) Lon B has a transmembrane insertion in the AAA-region, and also lacks the N-terminal domain of Lon A. The transmembrane insertion allows Lon B to associate with the membrane and participate in the degradation of membrane-bound proteins (Rotanova et al. 2004, Cha et al. 2010). Some bacteria have both types of Lon, including *B. subtilis*. Lon B in *B.*

subtilis probably fulfils the function of FtsH in degradation of membrane-bound proteins, since FtsH is lacking from this species. Lon is responsible for most of the degradation of misfolded proteins in bacteria and in mitochondria (Tsilibaris et al. 2006). The protease recognizes short hydrophobic regions that are exposed in misfolded or mistranslated polypeptides but not in correctly folded proteins (Gur and Sauer 2008). Several more specific substrates of Lon are known, among them SulA, a regulator protein that blocks cell division upon DNA damage, and UmuD, a protein involved in the SOS response (Sonezaki et al. 1995, Gonzalez et al. 1998). Despite the existence of Lon orthologs in many different organisms, they are absent in most strains of cyanobacteria although remnants of a former *lon* gene can be found in some such as *Synechococcus*.

1.5.4. FtsH protease

FtsH is a membrane-bound 71 kDa ATP-dependent metallo-protease. It is considered to be part of the AAA+ family of proteins since it has both Walker A and Walker B motifs (Tomoyasu et al 1993). The N-terminal region has a membrane-spanning domain, while the C-terminus has the proteolytic domain containing the zinc-binding proteolytic active site (H.E.X.X.H, Fig. 1) (Tomoyasu et al. 1995). Like most AAA+ proteases, the functional unit of FtsH is a hexamer. FtsH is the only protease that is essential for cell viability in *E. coli*, in which it was first discovered and characterized (Herman et al. 1993). Apart from playing an important role in the quality control of membrane proteins (Shimohata et al. 2002), FtsH also helps facilitate protein integration into the membrane (Akiyama et al. 1994). The eukaryotic FtsH protease can be found in the mitochondria and chloroplasts. In mitochondria, there are two types of FtsH, i-AAA and m-AAA. The transmembrane region of both types is anchored to the inner membrane, but the active domains are exposed to either the intermembrane space (i-AAA) or matrix (m-AAA). Loss of m-AAA activity can cause neurodegenerative diseases in humans (Rugarli and Langer 2006). In comparison to other organisms, photosynthetic species have a greater number of putative *ftsH* genes within their genome. Cyanobacteria can have up to four *ftsH* genes while the model plant species *Arabidopsis* has 12 (Garcia-Lorenzo et al. 2006). There are also five genes coding for FtsH-like proteins but these lack a recognizable zinc-binding motif and thus are likely to be inactive as proteases (Sokolenko et al. 2002). Nine of the active FtsH proteases and all of the inactive ones are localized or predicted to be localized in chloroplasts. Several substrates have been identified for FtsH, with the heat shock sigma factor σ^{32} being one of the first (Tomoyasu et al. 1995). SecY, a component of the secretory pathway is another substrate for FtsH in *E. coli* (Kihara et al. 1995). FtsH also degrades to some extent mistranslated polypeptides that are tagged with SsrA (Hermen et al. 1998). It is

also involved in the regulation of lipid biosynthesis by controlling the levels of LpxC and KdtA, each controlling different steps in the lipid synthesis (Ogura et al. 1999, Fuhrer et al. 2006, 2007, Katz and Ron 2008). FtsH has been demonstrated to release substrates when it encounters a tightly folded motif, which suggests it might also function in the activation of certain proteins; a process that is referred to as protein processing (Herman et al. 2003, Koodathingal et al. 2009). In plant chloroplasts FtsH is responsible for the degradation of the photosystem II (PSII) reaction center protein D1 together with the ATP-independent Deg protease during the PSII repair cycle (Kato et al. 2009, 2012). It is thought that FtsH might also function as a chaperone but the details of such a role are as yet unclear (Zheng et al. 2010).

1.5.5. HslUV protease

The threonine-type protease referred to as HslUV (Heat shock locus UV) consists of AAA+ protein ClpY and the proteolytic partner ClpQ. HslUV was the first AAA+ protease to be fully crystallized and structurally resolved (Bocther et al. 2000). It is sometimes called a "hybrid" protease since the proteolytic component (ClpQ) is similar in sequence and structure to the prokaryotic 20S proteasome β-subunits, while the unfoldase partner (ClpY) is more similar to ClpX (Chuang et al. 1993, Rohrwild et al. 1996). In this complex, both the chaperone and proteolytic subunits form separate hexameric rings. Known substrates for HslUV in *E. coli* are σ^{32} and the RcsA protein (Kuo et al. 2004).

1.5.6. Clp protease

Like many of the bacterial proteases, the Clp enzyme was first discovered and characterized in *E. coli* (Mauritzi et al. 1990). It has since been found in a wide range of organisms, including eubacteria, apicomplexa, plants and mammals (Adam et al. 2001, Roos et al. 2002). The *E. coli* ClpP is synthesized as a 207 amino acid polypeptide that is then auto-proteolytically processed to the mature 193 amino acid protein of 21.5 kDa (Maurizi et al. 1990). ClpP is a serine-type protease, with the active site consisting of the catalytic triad of S, H and D residues. In *E.coli*, ClpP is not essential for cell viability and its loss produces no detectable phenotype during exponential growth (Maurizi et al 1990) but does slightly impair cell survival under starvation conditions (Damerau and St John 1993, Weichart et al. 2003).

Mature ClpP monomers form heptameric rings that assemble into a tetradecameric complex (Shin et al. 1996, Flanagan et al. 1995). The back-to-back stacking of both heptameric rings creates a cavity within which the proteolytically active site of each ClpP subunit is positioned. The chamber is ca. 90 Å long with a

diameter of 51 Å, and it is accessible through narrow pores (10 Å) on either side of the proteolytic complex (Wang et al. 1997). The structure of the ClpP protein can be divided into three functional regions; the head and handle regions plus an N-terminal loop that appears to be highly flexible. These flexible loops appear to flank the side of the central pore in the ClpP tetradecamer. The crystal structure of the ClpP proteolytic core has now been solved from twelve other organisms, among them *B. subtilis* (Lee et al. 2011) *Helicobacter pylori* (Kim and Kim 2008) and *Mycobacterium turbeculosis* (Ingvarsson et al 2007). In many of these structures, the N-terminal region of ClpP is disordered and is now thought to be arranged in two different positions, "up" and "down" (Bewley et al. 2006). When in the up-conformation the N-terminus of six of the seven ClpP subunits extends out of the pore, blocking the entrance. When in the "down" conformation, the N-termini are contained within the access pore, and the pore is no longer blocked. This conformational change occurs when an unfoldase partner binds to the proteolytic core. The N-terminal region of ClpP is also important for association to the chaperone partner (Kang et al. 2004, Gribun et al. 2005, Bewley et al. 2006, Jennings et al. 2008a).

The ClpP complex can degrade peptides shorter than five amino acids without assistance of a chaperone. Peptides of this length can enter the proteolytic chamber through the narrow pore unassisted, probably by diffusion. Larger substrates cannot access the proteolytically active cavity due to the steric restriction of the narrow entrance (reviewed by Gottesman and Maurizi 1992, Thompson and Maurizi 1994, Kessel et al. 1995). As a consequence, degradation of longer peptides or folded proteins requires that they are first unfolded by the associated chaperone partner and then translocated into the proteolytic chamber of ClpP (Fig. 4) (Ortega et al. 2000, 2002). The fact that the proteolytic chamber is inaccessible to folded polypeptides is almost certainly a regulatory mechanism to prevent nonspecific protein degradation. More recent studies suggest that ClpP *in vitro* can degrade larger unfolded protein substrates without hydrolysis of ATP but at a very slow rate, although whether this activity is relevant *in vivo* is uncertain (Jennings et al 2008b). ClpP can also degrade larger protein substrates if it binds to a group of substances called acyldepeptidases (AEDPs). These types of molecules bind to the side of the ClpP tetradecamer and open up the entrance pore without the involvement of any interacting unfoldase. This causes uncontrolled degradation of proteins that eventually leads to cell death, making these types of substances interesting as potential candidates for antibiotics (Nagpal et al. 2013).

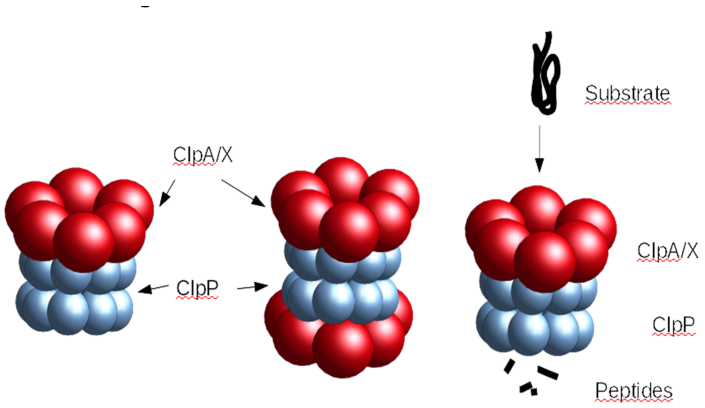


Figure 4. Clp proteases in *E. coli*. ClpA or ClpX hexamers can complex with the ClpP tetradecamer at one end or both *in vitro*, although it remains unclear if both possibilities also occur *in vivo*. The Clp/Hsp100 partner unfolds large protein substrates and translocates them into the Clp core for degradation.

1.6. Clp proteases and associated adaptor proteins in *E. coli*

1.6.1. ClpAP

ClpAP in *E. coli* was the first Clp protease characterized. ClpA forms hexamers in the presence of ATP that associate to the ClpP tetradecamer (Maurizi et al. 1991). A ClpA hexamer can associate to ClpP at either one end (1:1) or both (2:1), with the relative amount of ClpA to ClpP determining which type of complex is favoured (Maurizi et al. 1994: Kessel et al. 1995). ClpA acts as a gatekeeper for the protease, unfolding the protein substrate once bound and then threading it through the narrow entrance of ClpP into the degradative chamber. Several motifs important for ClpA function have been identified. The N-terminal region of ClpA is involved in substrate recognition (Lo et al. 2001, Erbse et al. 2008) as well as the binding of the adaptor protein ClpS (Dougan et al. 2002b, Guo et al. 2002 Zeth et al. 2002). Both of the AAA domains in ClpA are crucial for substrate processing and degradation by the ClpAP machinery (Kress et al. 2006). The D1 domain is important for oligomerization of the ClpA hexamer, while the D2 domain is necessary for ATP hydrolysis (Singh and Maurizi 1994). Binding of ClpA appears to cause the N-termini of ClpP to change conformation from “down” (blocking the pore) to “up” in which the pore is accessible (Bewley et al. 2006, Effantin et al. 2010). Another critical motif in ClpA is the P-loop (IGF/L) that is required for association to the ClpP proteolytic core where it binds to a hydrophobic region on the ClpP

subunits; this motif is also present in ClpX (Kim et al. 2001).

One of the first substrates found for ClpAP was ClpA itself, which is probably part of an autoregulatory mechanism controlling the level of the protease (Gottesman et al. 1990). RepA (bacteriophage plasmid P1 replication initiator) is another ClpAP substrate (Wickner et al. 1994). ClpAP can also degrade SsrA-tagged polypeptides *in vitro* (Farell et al. 2005), although this role *in vivo* is more likely performed by the ClpXP protease (see section 1.6.3.). Instead, much of the ClpAP protease appears to have ClpS attached, which redirects the specificity of ClpA to N-end rule substrates.

1.6.2. ClpS

ClpS is an adaptor that mediates N-end rule protein degradation in prokaryotes. *E. coli* ClpS has a molecular mass of ca 12 kDa, and its gene is located in an operon with *clpA* (Dougan et al. 2002b). When characterized it was discovered that this protein binds to *E. coli* ClpAP and changes its substrate preferences. ClpS was found to promote the degradation of two heat-aggregated proteins *in vitro* (Dougan et al. 2002b). It inhibits the degradation of SsrA-tagged proteins by ClpAP and also blocks the auto-degradation of ClpA (Dougan et al. 2002b). It promotes degradation of N-end rule substrates by binding to the N-terminus of ClpA and modifying its substrate specificity (Dougan et al. 2002b). ClpS is homologous to a domain in the eukaryotic E3 ligase that binds type 2 substrates according to the N-end rule pathway (Kwon et al. 1998, Lupas et al. 2003) ClpS recognizes the primary destabilizing N-degrons Leu, Phe, Tyr and Trp in *E. coli* and binds to the substrate and ClpAP (Erbse et al. 2006). There was debate if ClpS was essential for N-end rule degradation to occur or if it just modified the activity of ClpAP, but later data supported that ClpS is essential for N-end rule degradation in *E. coli* (Erbse et al. 2006, Schmidt et al. 2009). Removal of the first 17 amino acids in ClpS compromised its ability to block degradation of SsrA-tagged proteins (Hou et al. 2008), and it has since been proposed that the first 25 amino acids of ClpS form a flexible N-terminal extension (NTE) that is vital for degradation of substrates (Román-Hernández et al. 2011). Figure 5 illustrates the current model of how ClpS operates. The formation of a high affinity complex is dependent on the residue His66 in ClpS, which mediates the contact between the N-degron-binding region on ClpA and the substrate. The NTE binds to residues inside the ClpA pore. The unfolding machinery can then pull on ClpS and bring the N-degron of the substrate close to ClpA and binding can occur, after which ClpS is released. This model explains why the N-terminal regions of ClpS are vital for degradation but still allow binding to the substrate. From functional studies it is known that the N-terminal region of ClpS is crucial for degradation of N-end rule substrates (Hou et al. 2008). Structurally, however, the N-terminal region of ClpS is highly flexible and as such it has been poorly resolved in crystal structures to date.

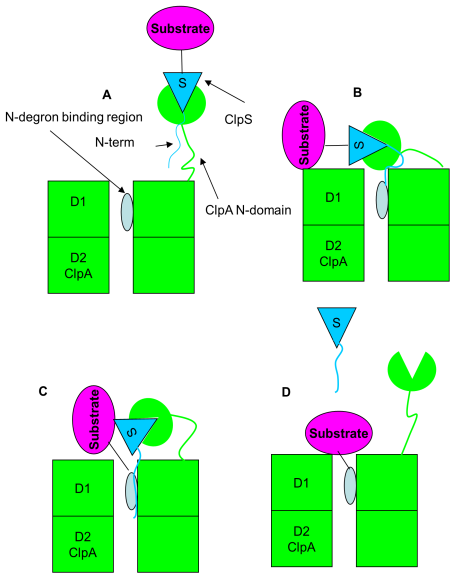


Figure 5. Model for ClpS function. ClpS binds to a substrate with the N-degron (**A**) and then binds to the D1 region of ClpA (**B**). The NTE region of ClpS binds to residues in the ClpA pore and is used by ClpA to pull the substrate into contact with the N-degron-binding region *via* a power stroke (**C**). ClpS is then released and degradation of the substrate proceeds and a new cycle begins (**D**) (adapted from Román-Hernández et al. 2011).

Until recently, only a few substrates of ClpAPS were known from *E. coli* and these included Dps and PATase (Schmit et al. 2009). Dps (DNA-binding protein from starved cells) is a protein that protects DNA during starvation, while PATase is a putrescine aminotransferase. Up to 100 putative substrates have now been identified for ClpS in *E. coli*, with many being modified either by addition of a primary degron or by processing to create primary degrons (Humbard et al. 2013). Orthologs for ClpS exist in a diverse range of organisms from bacteria to higher plants. Deletion of ClpS does not cause any visible phenotype in *E. coli* or *Arabidopsis* (Nishumyra 2013). The ClpS in *Arabidopsis* are localized in the chloroplast, where it is thought to be involved in N-end rule degradation mechanism (Nishumyra et al 2013), although there appears to be few chloroplast proteins with recognizable N-degrons (Apel et al 2010).

1.6.3. ClpXP

The other Clp protease that exists in *E.coli* is ClpXP. The mechanism of binding and action between the ClpX chaperone and ClpP proteolytic core has been extensively studied. ClpX forms stable dimers, which later assemble into hexamers upon binding of ATP. The Zn-binding domain in the N terminus is essential for dimerization of ClpX and its chaperone function (Wojtyra et al. 2003). ClpXP is best known for its central role in the degradation of SsrA-tagged substrates in *E. coli* (Flynn et al 2003). The SsrA degradation tag in *E. coli* is eleven amino acids long, with the sequence AANDENYALAA (Tu et al. 1995, Keiler et al. 1996). SsrA is commonly added to the C-terminus of proteins that are mistranslated prior to their release from the ribosome. Proteins thus marked are then rapidly degraded by ClpXP, with the possibility that ClpAP and FstH might also be involved (Lies and Maurizi 2008, Farrell et al. 2005).

Three conserved regions in ClpX have been identified as important for substrate unfolding and translocation into the proteolytic core. These are named the RKH-loop, the GYVG-loop (sometimes also called the pore 1-loop) and the pore 2-loop ((Siddiqui et al. 2004, Farrell et al. 2007, Martin et al. 2007, 2008a, 2008b) (Fig. 6). A maximum of four of the six ClpX monomers with the hexameric ring appears to bind ATP during the process (Martin et al. 2005). For translocation of protein substrates through the ClpX oligomer, the loops within ClpX transmit conformal changes that occur during ATP hydrolysis to the substrate protein (Glynn et al. 2009). The three different loops move up and down the central pore during ATP hydrolysis, which “pulls” the substrate through the pore and into the ClpP proteolytic chamber for degradation. The pore 2-loop also interacts with the N-terminal loop of ClpP in a highly flexible manner (Gribun et al. 2005, Bewley et al. 2006, Martin et al. 2007). Mutation in any of these loops causes decreased recognition and degradation of SsrA-tagged proteins.

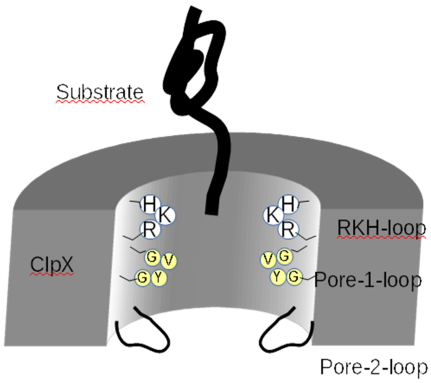


Figure 6. Principle mechanism for substrate threading through the ClpX pore. The three different loops inside the ClpX pore (RKH, Pore 1- and Pore 2) undergo conformational changes upon ATP hydrolysis that causes them to move up and down, "pulling" the substrate through and into the ClpP proteolytic core (adapted from Gur et al. 2013).

Another important motif in ClpX is the so-called P-loop (IGF/L). These residues are necessary for association to the ClpP core where it binds to hydrophobic pockets within each ClpP subunit (Kim et al. 2001, Singh et al. 2001, Joshi et al. 2004). This loop is conserved in all types of Clp/Hsp100 protein that function within a Clp protease. Inactivation of one of the six subunits within the ClpX hexamer decreases the degradation rate of the ClpXP protease, while mutation of two subunits abolishes all proteolytic activity (Baker et al. 2007). The general model of how substrate proteins are unfolded and translocated through the ClpX oligomer is almost certainly similar for other Clp proteases as well as for other AAA+ proteases (Flynn et al. 2003). Nearly 100 substrates have been identified for ClpXP using an inactivated ClpP core that traps substrates inside (Flynn et al. 2003, Nehrer et al. 2006). From these studies, five substrate recognition signals were found, two in the C-terminal region of the protein substrates and three at the N-terminal region (Flynn et al. 2003). Examples of such substrates include RecN, a damage response protein, which carries a C-terminal signal for degradation by ClpXP, and the MuA transposase and bacteriophage replication factor λO that possess intrinsic recognition sites (Gottesman et al. 1993, Flynn et al. 2003).

1.6.4. Adaptor proteins for ClpXP

The ClpXP protease has several known adaptor proteins that affect its substrate specificity and activity. Three such adaptor proteins have been identified in *E.coli*. SspB is the best characterized of these to date, and is a ribosome-associated protein that enhances the affinity of ClpXP for SsrA-tagged polypeptides (Bolon et al. 2004, Flynn et al. 2004). Another adaptor protein is RssB, which mediates the interaction between ClpX and σ^S. The transcription factor σ^S regulates the expression of genes important during various stress conditions such as heat, cold, osmotic stress, and oxidative stress, and for the transition to stationary growth phase (Loewen and Hengge-Aronis, 1994, reviewed by Hengge-Aronis 2000). Upon phosphorylation, RssB binds to σ^S and directs it to ClpXP for degradation *in vivo*. During exponential growth, RssB is phosphorylated but is then dephosphorylated upon the transition to the stationary phase, thereby reducing its affinity for σ^S and the sigma factor's susceptibility to ClpXP degradation. The third adaptor, UmuD, participates in the fast repair of DNA when damage. The active form of UmuD is referred to as UmuD´, which has the first 24 amino acids at the N-terminus removed by RecA (Shinagawa et al. 1988, Neher et al. 2003). UmuD' is degraded by the ClpXP protease but only if it forms a dimer with UmuD, although UmuD itself is not degraded during this process.

1.7. Cyanobacteria

Since the work in this thesis has been to characterize a Clp protease and its ClpS adaptors in the cyanobacterium *Synechococcus elongatus* a brief introduction to cyanobacteria as a model organism is in place. Cyanobacteria are the oldest known organisms that perform oxygenic photosynthesis, with a fossil record going back at least two billion years. Their photosynthetic activity is responsible for the oxygenation of the Earth's atmosphere in a process referred to as the "great oxidation event". These early cyanobacteria are now regarded as the ancestor of plastids in algae and plants *via* an endosymbiotic event in which the cyanobacterial progenitor was engulfed by the pre-eukaryotic cell (Martin et al. 1998).

The relative simplicity of cyanobacteria makes them an ideal model system for many chloroplast functions. The model cyanobacterial strain used in this study is *Synechococcus elongatus* PCC 7942 (hereafter referred to *Synechococcus*), which is a freshwater obligate photoautotroph that originates from ponds in California. Being naturally competent, it is readily amenable to genetic manipulations such as the creation of specific gene knockout lines (**van der Plas 1990**). As our standard growth condition, *Synechococcus* was grown in batches at 37°C under a photon flux density of 70 µmol. photons m^{-2} s^{-1}. The culturing medium was BG-11 and cells were bubbled with

5% CO_2 in air to produce fast-growing and reproducible cultures with a consistent pigment composition.

Table of Clp-proteins in some organisms.

	E.coli	*Synechococcus*	*H.sapiens*	*Arabidopsis*
Proteolytic Core components	ClpP	ClpP1-P3 ClpR	ClpP[1]	ClpP1, P3-6 ClpR1-R4 P2[1]
ATPase with P-loop	ClpA ClpX	ClpC ClpX	ClpX[1]	ClpX1[1], X2[1], X3[1] ClpC1, C2 ClpD
ATP-ase missing P-loop	ClpB	ClpB1, B2		ClpB1 (cytosolic) ClpB2[1] ClpB3[2]
Adaptors	ClpS	ClpS1, S2		ClpS ClpT1, T2

1= Mitochondria
2 = Chloroplast

1.7.1. Cyanobacterial Clp proteases

In comparison to other eubacteria, cyanobacteria have a more diverse range of Clp proteins (Table). *Synechococcus* has four Clp/Hsp100 proteins: ClpB1, ClpB2, ClpC and ClpX. Of the proteolytic subunits, there are three ClpP paralogs (ClpP1-P3) as well as a ClpP variant known as ClpR that lacks the catalytic triad (Clarke et al. 1999). There are also two ClpS paralogs, ClpS1 and ClpS2 (Stanne et al. 2007). Previous work by our group has demonstrated the existence of two distinct Clp proteolytic cores in *Synechococcus*, one containing ClpP1 and ClpP2 and the other ClpP3 and ClpR, with the likely chaperone partners being ClpX and ClpC, respectively (Stanne et al 2007). The structure and function of the ClpP3/R core complex and its interaction with ClpC has been examined in detail in Papers I and II. Of the adaptor proteins, ClpS1 was also shown to associate to ClpC in earlier studies (Andersson et al. 2006, Stanne et al. 2007), with the function of ClpS1 further examined in Papers I and III. The properties of the second adaptor ClpS2 are also detailed in Paper III. A third Clp proteolytic core consisting of ClpP1/R attached to the membrane was also proposed by our group (Stanne et al. 2007) but subsequent work suggests that such a Clp protease does not exist *in vivo* (Tryggvesson unpublished data).

1.8. Clp proteins in plants

Although cyanobacteria have many diverse Clp proteins, those in photosynthetic eukaryotes are far more numerous and complex. *Arabidopsis* has up to 22 different Clp proteins, most of which are localized in chloroplasts including four Hsp100 proteins (ClpB3, ClpC1, ClpC2, ClpD), five ClpP (ClpP1, ClpP3-6), four ClpR (ClpR1-4), one ClpS and two accessory proteins (ClpT1 and ClpT2) that have sequence similarity to the N-terminal region of ClpC. Several Clp proteins are also present in mitochondria, three ClpX paralogs (ClpX1-X3) and one ClpP (ClpP2) (Adam et al. 2001, Peltier et al. 2004). Despite the many Clp proteins inside the chloroplast, only a single proteolytic core complex exists (Peltier et al. 2004, Sjögren et al. 2006). This core complex consists of two distinct heterogeneous heptameric rings, one containing ClpP3-P6 (P-ring) and the other ClpP1 and ClpR1-R4 (R-ring) (Sjögren et al. 2006). The accessory proteins ClpT1 and ClpT2 associate to only the P-ring and appear to facilitate the assembly of the tetradecameric complex (Sjögren et al. 2011). ClpC1, ClpC2 and ClpD all possess the C-terminal P-loop and are therefore likely chaperone partners for the Clp proteolytic core, although ClpC1 is by far the most abundant of the three throughout vegetative growth (Sjögren et al. 2006, Sjögren et al. 2014). Many putative *in vivo* substrates for the chloroplast Clp protease have been identified in *Arabidopsis*, the functions of which suggest that Clp acts primarily as a housekeeping protease in chloroplasts (Sjögren et al. 2006, Stanne et al. 2009). Loss of the Clp proteolytic activity is seedling lethal, highlighting the importance of this enzyme for chloroplast function. Although mainly localized in the stroma, the Clp protease has also recently been found attached to the envelop membranes (Sjögren et al. 2014), potentially broadening its range of protein substrates and thereby its overall importance.

2. **Aims of the thesis.**

At the commencement of my studies, no Clp protease with a heterologous proteolytic core had been characterized and as such the main focus of my work has been the ClpP3/R core from *Synechococcus*. Earlier work in our group had attempted to purify ClpR and ClpP3 separately but neither protein was proteolytically active *in vitro*. Since the *clpR* and *clpP3* genes are arranged within a bicistronic operon in *Synechococcus* (Schelin et al. 2002) and most other cyanobacteria we considered the possibility that these proteins oligomerized together to form a single proteolytic core. We later confirmed that ClpP3/R did indeed form such a heterologous core *in vivo* (Stanne et al. 2007). Within this complex, we also wanted to address the role of ClpR and why it is present in only photosynthetic organisms. Sequence comparisons revealed that ClpR lacked the active site amino acids of a Ser-type peptidase but it remained unclear if it

was indeed proteolytically inactive or possessed another type of proteolytic activity.

Besides characterizing the basic structure and function of the ClpP3/R core, the role of the N-termini of the ClpP and ClpR subunits was also examined. It was known that the N-terminus of *E. coli* ClpP is very important for the function of the proteolytic complex and its association to the chaperone partner. We were interested in how this would function in a complex consisting of more than one type of subunit. Also of interest were the sequences that defined the recognition and interaction between the proteolytic core and chaperone partner and the specificity of this association.

Another aim was to investigate the function of the ClpS adaptor proteins in cyanobacteria. It is known that EcClpS plays an important role in regulating the Clp protease in *E.coli*, changing the substrate specificity of ClpA. In an earlier study, it was shown that *Synechococcus* ClpS1 associates to ClpC *in vivo* (Stanne et al. 2007) but does it alter the specificity of ClpC in a similar way to that by ClpS of ClpA in *E. coli*? Cyanobacteria also have a second ClpS protein, ClpS2, but it function remains unknown. Of particular interest is whether the two ClpS adaptors recognize different sets of protein substrates for the Clp protease and if so how are these different substrates identified?

3. Results and Discussion

3.1. Composition of the ClpP3/R complex.

Given that *Synechococcus* ClpP3 and ClpR were shown to form a single complex (Stanne et al. 2007) and that their genes are co-expressed *in vivo* (Schelin et al. 2002), we attempted to purify recombinant ClpP3 and ClpR by co-expressing them in *E. coli*. The two *Synechococcus* genes were cloned into the pACYC Duet expression vector (Invitrogen), with *clpP3* containing additional sequence at the 3´end coding for a His_6-tag to facilitate purification. If the two different proteins oligomerize into a complex then ClpR should co-purify with ClpP3. This construct yielded large amounts of recombinant protein when over expressed in *E. coli (***Paper I).** Purification by sequential Ni^{2+} affinity and gel filtration column chromatography yielded a highly pure protein preparation without any visible contamination. The fact that ClpR did co-purify with ClpP3 indicated that the two proteins oligomerized together in a single complex *in vitro*.

Having obtained the recombinant ClpP3 and ClpR proteins, we then examined the composition of the oligomer relative those Clp proteolytic cores characterized from *E. coli* and other bacteria. When separated by native-PAGE, recombinant ClpP3 and ClpR were present in a single complex of ca. 270 kDa, matching the size of the ClpP3/R core *in vivo* as previously described (Stanne et al. 2007, **Paper I**). Different microscopy techniques were then used to reveal that the recombinant ClpP3/R complex formed a

symmetrical barrel-shaped structure similar to that formed by ClpP in *E. coli* (Schnider et al. 2005). The ClpP3/R complex consisted of two identical heptameric rings, each with a stoichiometry of three ClpP3 and four ClpR subunits. Within each ring structure, the ClpP3 and ClpR subunits were arranged in the alternating configuration of R/P3/R/P3/R/P3R/R (**Paper I**).

To determine if the recombinant ClpP3/R complex was functional, we first tested its ability to degrade small peptides *in vitro*. In contrast to EcClpP, however, ClpP3/R exhibited no peptidase activity against several different peptides. The ability of ClpP3/R to degrade the model substrate α-casein was then examined, combining ClpP3/R with its known chaperone partner ClpC along with an ATP-regeneration system. The proteolytic assay revealed that ClpP3/R with ClpC could degrade α-casein but at a rate that was relatively slow to that performed by the *E. coli* ClpAP protease. Interestingly, no degradation was observed if ClpC was replaced with *E. coli* ClpA, suggesting that the ClpP3/R core does not associate to chaperone counterpart in *E. coli* (**Paper I**).

3.2. Inactivation of ClpP3 and reactivation of ClpR

Sequence alignments of ClpR with ClpP3 and *E. coli* ClpP revealed the apparent absence of the three active site amino acids in the ClpR subunit that would constitute the catalytic triad of Ser-type proteases (Clarke et al. 1998, Fig. 7). To test this proposal, a version of the ClpP3/R core was over-expressed and purified in which the active site Ser residue in ClpP3 was changed to Ala, thereby inactivating all catalytic activity in this subunit. This version of the core complex, called SynClpP3 S101A, not only formed the 270 kDa oligomer that matched the size of the wild type ClpP3/R tetradecamer, but it also showed no disturbance to the normal ClpC association. When used in proteolytic assays, however, the ClpP3S101A/R core was unable to degrade α-casein (**Paper I**), confirming that ClpR did not contribute to the degradative activity of ClpP3/R.

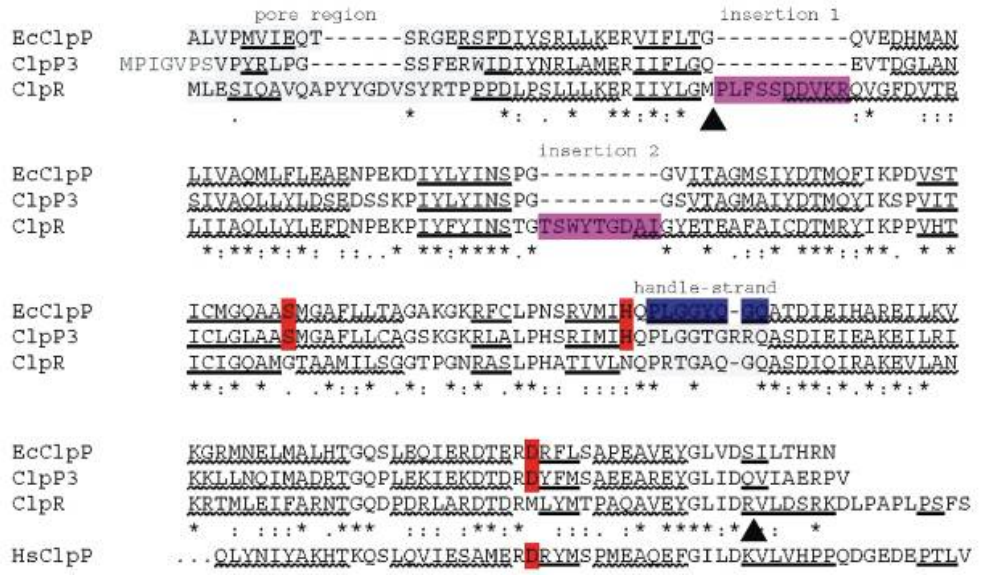


Figure 7. Alignment of *Synechococcus* ClpR and ClpP3 with *E. coli* ClpP. The proteolytic active sites are marked in red and the extensions in ClpR marked in magenta (from Andersson et al. 2009).

To investigate if the inclusion of the inactive ClpR subunit limited the overall proteolytic activity of the ClpP3/R core, attempts were made to restore the proteolytic activity of ClpR. If the slower activity of the ClpP3/R core compared to EcClpP is due to fewer active sites, then in theory a reactivated ClpR should enhance proteolytic activity considerably. The first changes made to ClpR were to restore the three active site amino acids along with the removal of the two extension regions (Fig. 7), but these modifications failed to increase the proteolytic activity of the ClpP3/R complex. We next made more extensive modifications to ClpR by replacing the sequence from Met38 to Arg212 with the corresponding region from ClpP3. This chimeric version of ClpR was then co-expressed with either the wild type or inactivated ClpP3 in *E. coli*, with the proteolytic activity of the different core complexes tested against α-casein. For the core containing the inactive ClpP3 subunit with the modified ClpR, its proteolytic activity was similar to that of the wild type ClpP3/R complex, demonstrating that the more extensive changes to ClpR had indeed restored catalytic activity. However, the proteolytic activity of the core containing active ClpP3 with the modified ClpR was also similar to that of the wild type complex despite all subunits now being catalytically active. This suggested at the time that the lack of activity in wild type ClpR was not rate-limiting for the overall activity of the ClpP3/R core (**Paper I)**. Later, however, we

observed by native-PAGE that the core containing the reactivated ClpR subunit formed fewer stable tetradecamers than the wild type (Tryggvesson unpublished), raising the possibility that reactivating the ClpR subunit could indeed increase the proteolytic activity of the ClpP3/R core but that this potential gain would be compromised by increased instability of the oligomer.

3.3. Modeling of ClpR and ClpP3/R

Our group in collaboration with several structural chemistry laboratories has made numerous attempts to crystallize the *Synechococcus* ClpP3/R core but failed to produce crystals of sufficient quality and size to obtain reliable x-ray diffraction data. As a consequence, we have modelled the ClpR protein and the ClpP3/R complex using structures available for ClpP proteins from different bacteria such as *E. coli* and *Streptococcus* (Wang et al. 1997, Gribun et al. 2005). The resulting models highlighted several distinct features in the ClpP3/R compared to the EcClpP complex (**Paper I**). The second of the two internal extensions in ClpR, which is the one conserved for all ClpR orthologs extends over the region corresponding to the substrate specificity pocket in EcClpP. This extension on ClpR reaches further into the pore chamber than the corresponding region in the EcClpP structure, which would almost certainly disrupt the substrate interacting pocket containing the catalytic triad. This could explain the gradual loss of the active sites in ClpR since some cyanobacteria still retain part of the catalytic triad in their ClpR sequence.

Another interesting feature revealed from the modeling studies is the structure at the pore entrance of ClpP3/R (**Paper I**). The calculated model suggests that there are flexible parts in the N-terminus of both subunits that extend further out from the lining of the pore entrance than in EcClpP. These parts are probably important for core binding and interaction with ClpC and are missing from the EcClpP model, which could explain at least in part why EcClpP is unable to associate to ClpC. Examining the central pore channel reveals more amino acids with hydrophobic properties present in the ClpP3/R complex than in EcClpP. The effect of this is that the pore channel has a very narrow diameter compared to EcClpP. The model suggests that the entrance pore might for all practical purposes be closed when the core complex is not bound to ClpC. This more closed pore might also be the reason to the lack of peptidase activity displayed by ClpP3/R, which would infer that the rate limiting factor for proteolysis is the unfolding activity of ClpC.

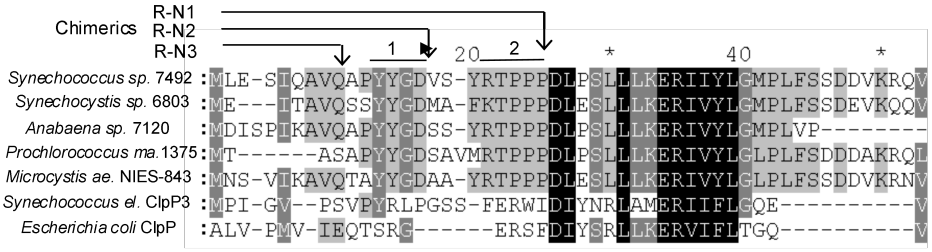
In Paper I, we characterized the structure and function of the essential ClpP3/R core in vitro. The recombinant ClpP3/R complex is proteolytically active in association with its chaperone partner ClpC but it lacks peptidase activity against a range of synthetic peptides. The ClpR protein was shown to be proteolytically inactive consist with its

apparent lack of a Ser-type catalytic triad. The ClpP3/R core consists of two identical heptameric rings, each with three ClpP3 and four ClpR subunits arranged in an alternating configuration; this stoichiometry of active to inactive subunits mirrors that within the β-subunits of the eukaryotic 26S proteasome.

3.4. Analysis of the N-terminal region of ClpP3 and ClpR

The next step in the characterization of ClpP3/R and the potential function of ClpR was to investigate the role of the N-terminal region of both types of subunits given that the modelling suggested that these regions might play an important role in the interaction with ClpC. Although some sequence similarity exists between the N-terminal of EcClpP and ClpP3, none occurs between EcClpP and ClpR. Since the ClpP3/R model (**Paper I**) showed the N-terminus of ClpR protruding further out from the pore than that of EcClpP, the lack of sequence conservation between these two proteins might not be too surprising.

A



B

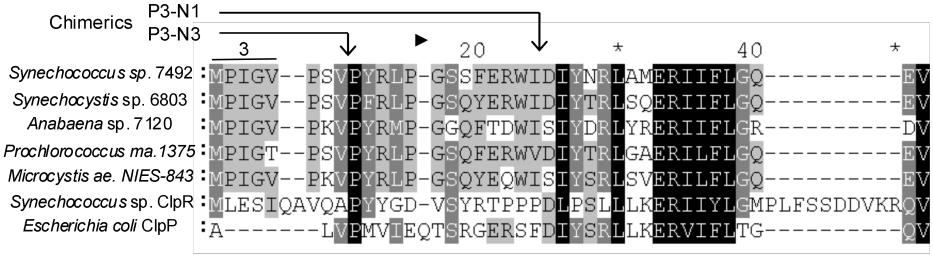


Figure 8. N-terminal chimerics of ClpR and ClpP3. A. Sequence alignments of selected cyanobacterial ClpR proteins in conjunction with *Synechococcus* ClpP3 and EcClpP. Arrows illustrate the N-terminal regions in ClpR that were substituted with the corresponding region in ClpP3, with the conserved Tyr- (1) and Pro-motifs (2) in ClpR underlined. **B.** Sequence alignments of selected cyanobacterial ClpP3 proteins along with *Synechococcus* ClpR and EcClpP. Arrows illustrate the N-terminal regions in ClpP3 that were substituted with the corresponding region in ClpR, with the conserved MPIG-motif (3) in ClpP3 underlined.

Two conserved regions in the N-terminus of ClpR were identified as being potentially important for the interaction between ClpC and ClpP3/R. (**Paper II**). The first region termed the **Tyr motif (**YYGD) is situated 12-15 amino acids from the N-terminus of ClpR (Fig. 8A, region 1). The second motif, named the **Pro motif** (RTPPP) is located 19-23 amino acids from the start of the N-terminus (Fig. 8A, region 2). Three modified versions of ClpR were prepared in which different lengths of the N-terminus were replaced with the corresponding sequence from ClpP3 (Fig. 8A). These three ClpR variants were all co-expressed with wild type ClpP3 in *E. coli* and purified. Degradation assays revealed that the core containing ClpR N-1 was proteolytically inactive while the core containing R-N2 displayed rates similar to those of the wild type ClpP3/R core. Interestingly, the core complex containing the ClpR-N3 chimeric degraded α-casein almost twice as fast as the wild type core (**Paper II**). This enhanced degradation activity could have been due to either the size of the entrance pore being larger in the chimeric core complex, thus allowing easier access to the catalytic sites, or to a stronger association to ClpC enabling more efficient translocation of the unfolded substrate. Tests of the interaction to ClpC by measuring the degree that each chimeric core stimulates the ATPases activity of ClpC revealed that the core containing R-N1 had lost the capacity to bind to ClpC. The R-N2 core complex stimulated ATP hydrolysis by ClpC only slightly less than the wild type core, suggesting the loss of the Tyr motif has little effect on ClpC association. The R-N3 core, however, stimulated ClpC ATPase activity almost twice that of the wild type ClpP3/R (**Paper II**). It should be noted that all the three ClpR chimeric core complexes were unable to function as peptidases against the synthetic peptides used earlier for the wild type ClpP3/R.

Removal of the Pro motif in ClpR was also shown to affect the assembly of the ClpP3/R core. The R-N1 core formed two types of oligomers with one larger than the wild type complex. This suggests that the Pro motif plays an important role both in the association to ClpC as well as in core assembly (**Paper II**). In comparison, the role of the Tyr motif in ClpR is less clear. The R-N2 core could bind to ClpC and showed no obvious change in proteolytic activity. However, the faster proteolysis by the R-N3 core suggests that the addition of the MPIG motif from ClpP3 to all ClpR subunits might compensate for a reduced ClpC affinity caused by the loss of the Tyr motif. Indeed, the enhanced proteolytic rate of the R-N3 core indicates that the N-terminal region of ClpP3 plays a critical role in the functioning of the ClpP3/R proteolytic core.

Since the changes to the N-terminus of ClpR had such dramatic effects on ClpP3/R we decided to also modify similar regions of the N-terminus of ClpP3. Sequence comparison of cyanobacterial ClpP3 orthologs revealed a highly conserved motif in the first six amino acids (MPIGVP) that was different and more variable in the ClpR orthologs (Fig. 8B). As a consequence, two chimeric versions of ClpP3 were prepared in which different lengths of the N-terminus were substituted for the corresponding

regions in ClpR and then over-expressed with the wild type ClpR (Fig. 8B). Assays revealed that both modified cores lacked proteolytic activity and had a lower binding affinity to ClpC (**Paper II)**, indicating that the MPIG motif in the N-terminus of ClpP3 is crucial for association to ClpC and potentially also for substrate translocation through the pore. Overall, both ClpR and ClpP3 possess conserved N-terminal motifs that are essential for ClpC binding and substrate degradation.

All five chimeric core complexes along with wild type ClpP3/R were separated by native-PAGE to analyze their oligomeric stability. All but two of the chimerics only formed a stable oligomer consistent with the ClpP3/R tetradecamer. The ClpR N-1 core formed two complexes, one matching the size of wild type ClpP3/R and another ca. 90 kDa larger that would be consistent with an octodecamer. In contrast, the P3-N1 core was unstable, with ca. half of the protein appearing as an indistinct smear. It is known that the N-terminal region of EcClpP is important for the interaction with the chaperone partner. Removal of the first 22 residues from EcClpP prevents binding to ClpA or ClpX (Gribun et al. 2005, Jennings et al. 2008). In EcClpP, Arg12 and Ser23 were identified as being necessary for heptameric ring formation (Wang et al. 1997). Interestingly, all cyanobacterial ClpP3 orthologs have an Arg in a matching position (Arg11 in ClpP3) but not the Ser, whereas ClpR sequences possess the Ser (Ser27 in ClpR) but not the Arg. Given that hydrogen bonding between these residues is crucial for core assembly, then the loss of the Ser27 residue in the R-N1 chimeric would explain its structural instability. Furthermore, the removal of Arg11 in the ClpP3 subunits of the P3-N1 chimeric would also explain its instability (**Paper II**).

3.5. Core specificity factor in ClpC

Since ClpP3/R did not associate to EcClpA, and EcClpP did not bind to ClpC (**Paper I**), we next searched for conserved domains in ClpC that could confer its core specificity. Given that the C-terminal P-loop is present in all Hsp100 proteins that interact with a ClpP proteolytic core, additional factors must be involving in conferring the specificity of ClpC for ClpP3/R. To search for such potential motifs, sequence alignments of the C-terminal region of various ClpC orthologs from both cyanobacteria and vascular plants were prepared. These alignments revealed an eight amino acid domain (YNRIRSLV) just downstream of the P-loop that was conserved in all the ClpC orthologs but absent from ClpA from various bacteria. To examine the possible importance of this domain, a recombinant version of ClpC was prepared in which the identified domain was replaced with the corresponding shorter sequence in EcClpA. This modified form of ClpC, termed ClpCA, was over-expressed in *E. coli* and purified. Proteolytic assays showed no observable degradation of α-casein when ClpCA was combined with ClpP3/R whereas there was degradation when EcClpP was added. The degradation rate of ClpCA with

EcClpP was faster than that of the ClpCP3/R protease but slower than that of EcClpAP (**Paper II**). ATPase assays confirmed that the core specificity of ClpCA had changed from that of wild type ClpC to EcClpA. These results confirmed that the region identified from the alignment, now referred to as the **R-motif**, is essential for association between ClpC and ClpP3/R. Apart from ClpC orthologs, the R-motif is also present in the ClpD protein from vascular plants, suggesting that this Hsp100 protein closely related to ClpC is also likely to function with the chloroplast ClpP/R core.

Another factor that could influence the specificity between ClpC and ClpP3/R is the amino acid composition of the hydrophobic cleft on the surface of the proteolytic core. This hydrophobic cleft in EcClpP is formed by Tyr60, Tyr62, Phe82, Ile90, Phe112 and Leu189 (Bewley et al. 2006). All but one of these residues are present in cyanobacterial ClpP3 and ClpR orthologs, with Phe112 changed to a Leu or Val in ClpP3 or to an Ala in ClpR. Interestingly, changing Phe112 to an Ala in EcClpP destabilizes its association to ClpA, suggesting that this single amino acid variation in ClpR could contribute to the specificity between ClpC and ClpP3/R.

In Paper II, we have shown that conserved regions in the N-terminus of both ClpR and ClpP3 are important for the association to ClpC and for the stable oligomerization of the ClpP3/R tetradecamer. We have also identified a domain in the C-terminus of ClpC that confers its specific interaction with the ClpP3/R proteolytic core.

3.6. The ClpS adaptor proteins

As previously mentioned, the ClpAP protease in *E. coli* degrades N-end rule substrates when associated to the small adaptor protein ClpS, which binds to destabilizing residues at the N-terminus on substrate proteins. The amino acids known to function as destabilizing signals in *E. coli* are Leu, Phe, Tyr and Trp. An earlier study by our group demonstrated that *Synechococcus* ClpS1 binds to ClpC *in vitro* (Andersson et al. 2006), and as such we were interested in further characterizing the function of this protein and elucidate if it targeted N-end rule substrates in *Synechococcus.* A recombinant version of ClpS1 was prepared with an N-terminal His_6-tag for purification. We found that if ClpS1 was added to ClpC and ClpP3/R the degradation of α-casein was blocked (**Paper I**). To investigate is ClpS1 also promotes degradation of N-end rule substrates we used a GFP-protein modified with an added N-end rule degradation tag (FR-GFP). The FR-GFP substrate was stable in the presence of ClpC and ClpP3/R but was completely degraded after the addition of ClpS1 (**Paper I**), indicating that ClpS1 associates to ClpC and alters its specificity to N-end rule protein substrates.

3.6.1. Identification of ClpS2

Again as mentioned earlier, it has been known for some time that cyanobacteria possess a second ClpS paralog that is so far unique to cyanobacteria (Lupas and Koretke 2003). A phylogenetic analysis suggests that the split between ClpS1 and ClpS2 occurred well before the endosymbiotic event. Interestingly it appears that only one copy (ClpS1) was transferred to plants *via* the endosymbiotic event, and that this chloroplast-localized ClpS in *Arabidopsis* interacts with ClpC1 and the chloroplast-localized Clp proteolytic core (Nishimura et al. 2013). Examination of the ORF for *Synechococcus* ClpS2 showed that it encoded a protein with an unusually long N-terminal extension that was not present in other ClpS2 orthologs. Attempts to inactivate the *clpS2* gene all failed to produce viable transformants, suggesting that the function of ClpS2 is essential for phototrophic growth in *Synechococcus*. This apparent importance of ClpS2 is in contrast to ClpS1 (and other bacterial ClpS orthologs), which can be deleted without causing any visible phenotypic changes (Ståhlberg et al. in preparation). A polyclonal antibody was prepared against a synthetic peptide of 15 amino acids from the C-terminus of ClpS2. This antibody detected a 17 kDa protein that corresponded to the predicted polypeptide from the full-length ORF. This 17 kDa protein was equally distributed between the soluble and membrane fractions in *Synechococcus*, which suggested it associated with ClpC which was also found in both subcellular locations (Stanne et al. 2007). As a consequence, ClpS2 was assumed to be a 17 kDa polypeptide in *Synechococcus* and that the unusually long N-terminal extension was important for its membrane interaction and critical function.

In order to study the functional characteristics of ClpS2 more closely, we prepared a recombinant ClpS2 protein carrying an N-terminal His_6-tag, a strategy that had been successful in purifying ClpS1 (Andersson et al. 2006). This recombinant ClpS2 protein, however, proved to be highly insoluble during storage and had to be prepared fresh for each set of experiments. As expected for a ClpS adaptor, the addition of ClpS2 blocked the degradation of α-casein by the ClpCP3/R protease *in vitro*. However, ClpS2 did not stimulate degradation of N-end rule substrates despite several different types of N-end rule substrates being tested.

At that time, new data raised questions over the true identity of the 17 kDa protein previously thought to be ClpS2. In another project, a modified version of the full-length *clpS2* gene coding for a protein with a C-terminal His_6-tag was transformed into wild type *Synechococcus* to replace the native *clpS2* gene. Correct insertion of the construct and its complete segregation was confirmed by PCR, but when immunoblots were performed with the ClpS2 antibody the same 17 kDa protein was identified in the transformants without the change in size from the wild type protein expected from the

His_6-tag. Attempts were then made to map the 5'-end of the *clpS2* mRNA in wild type *Synechococcus* using 5' RACE but these failed to produce sufficient cDNA product for sequencing. A new antibody was therefore prepared but this time to the full-length ClpS2 recombinant protein that included the N-terminal extension. Immunoblots with this new ClpS2 antibody no longer detected the 17 kDa protein but instead recognized a smaller, less abundant protein of 12.5 kDa; a size that is more typical of ClpS orthologs in other organisms. The new antibody also detected a slightly larger protein in the ClpS2 transformants as would be expected if the 12.5 kDa contained the C-terminal His_6-tag (**Paper III**). Re-examination of the *clpS2* gene revealed in position 64 a Val_{GTG} codon that is known to often function as an alternative translational start codon for many bacterial genes, and as we had previously shown for the *Synechococcus clpB1* gene (Eriksson and Clarke 1996). Translation from this Val codon would produce a 12.5 kDa protein, and as such we prepared a new recombinant N-terminal His-tagged version of ClpS2 in which the Val64 was changed to Met. We also repeated the localization studies with the new antibody and revealed that all ClpS2 was present in the soluble fraction along with ClpS1 (**Paper III**).

3.6.2. Quantification of ClpS proteins *in vivo*

Immunoblotting with specific antibodies was used to determine the amount of ClpS1, ClpS2 and ClpC in wild type *Synechococcus* relative to the cellular chlorophyll (Chl) content. A standard curve was made using known amounts of recombinant ClpS1, ClpS2 and ClpC in order to calculate the amount of each Clp protein in cell extracts. The calculated amount of ClpS1 was several times higher (0.093 pmol monomer/μg Chl) than that of ClpS2 (0.0012 pmol monomer/μg Chl), while the level of ClpC hexamer was 0.13 pmol/μg Chl. The relatively small amount of ClpS2 would explain why the 5'-RACE analysis failed given that the amount of *clpS2* transcript is almost certainly low. It remains unclear how much of the ClpC hexamers in *Synechococcus* are likely to have either ClpS1 or ClpS2 attached since there are disagreements over how many ClpS monomers are needed to bind to EcClpA to alter its substrate specificity (Hou et al. 2008, De Donatis et al. 2010).

3.6.3. Substrate specificity of ClpS1 and ClpS2

To investigate any possible functional differences between ClpS1 and ClpS2, the recombinant versions of both proteins were used in a series of proteolytic assays with ClpCP3/R. A range of protein substrates were tested containing N-terminal sequences that function as either destabilizing (Phe, Leu and Tyr) or stabilizing residues (Met or Ala) in *E. coli*. ClpS2 was first shown to block the degradation of α-casein by ClpCP3/R as did ClpS1 (**Paper III**). We then tested the substrate FM-L-GFP that was known to be

targeted by both ClpS1 and EcClpS, but no degradation was observed when ClpS2 was added, indicating that the two ClpS adaptors in *Synechococcus* have different substrates specificities *in vitro* (**Paper III**). Additional substrates were then tested to examine this potential difference in specificity between ClpS1 and ClpS2. Two types of proteins were used, the green fluorescent protein (GFP) with different sequences added to the N-terminus or a version of the Dps protein from *E. coli* (DNA binding proteins from starved cells) that is an *in vivo* substrate for the ClpAPS protease and carries an N-degron at the N-terminus (Schmidt et al. 2009). Many of the substrates tested were not recognized by either ClpS1 or ClpS2. The substrate YLFL-GFP, however, which contains a primary destabilizing amino acid (Tyr) at the N-terminus and has been shown to be a *in vitro* substrate for EcClpAPS (Wang et al. 2008) was recognized by ClpS1 but not ClpS2. In contrast, the substrates LVK-GFP and LVK-Dps that both carry a degradation tag from *E. coli* Dps were not targeted by ClpS1 for degradation but were by ClpS2 (**Paper III**). Not only did this demonstrate that ClpS1 and ClpS2 recognize different destabilizing amino acids at the N-terminus, but that ClpS1 does not recognize all the types of N-end rule substrates recognized by EcClpS.

The level of conservation between ClpS1 and ClpS2 is high in regions essential for binding to the chaperone but not for residues important for substrate interaction (**Paper III**) (Schueneman et al. 2009, Nishimura et al. 2013). The residues Met40 and Met62, which are crucial for substrate interaction by EcClpS are also present in ClpS1 but in ClpS2 are changed to Phe and Thr, respectively (Nishimura et al. 2013); the difference in these two amino acid positions could explain why ClpS1 and ClpS2 have different substrate specificity. Although ClpS1 contains the two Met residues, however, it did not target the LVK-GFP/Dps substrates for degradation, suggesting ClpS1 is not the direct functional equivalent of EcClpS.

3.6.4. Conserved amino acid regions in ClpS2

The proteolytic assays demonstrated that ClpS1 and ClpS2 recognized different destabilizing residues at the N-terminus of various N-end rule substrates. Since the function of ClpS2 also appears to be essential for cell viability, we searched for conserved motifs within each type of adaptor that could underlie their different roles. A MEME analysis (Bailey et al. 1994) was performed on ClpS1 and ClpS2 orthologs from 26 cyanobacterial species to identify conserved regions in the two ClpS groups. The most striking find was that all 26 sequences of ClpS2 used in the analysis contained the motif MAPLE close to the C-terminal end, whereas the corresponding region in ClpS1 orthologs showed no obvious sequence conservation (**Paper III**). The exact role of this conserved region in ClpS2 remains unclear. Also conserved at the very C-terminus of many ClpS2 proteins (12 of 26) are two Ala residues. It is known that such twin Ala

residues at the C-terminus can function as a degron for SsrA-mediated degradation in bacteria (Levchenko et al., 1997, Flynn et al. 2001). This suggests that ClpS2 in certain cyanobacteria might be regulated by SsrA-mediated degradation, making them potential targets for the ClpXP protease. This possibility is supported by the fact that ClpS2 levels are relatively high in the *Synechococcus ΔclpP1* mutant that lacks a functional ClpXP1/P2 protease (Ståhlberg et al. in preparation). Despite this, we did not observe any degradation *in vitro* of recombinant ClpS2 when added to ClpP1/P2 and EcClpX.

3.6.5. Potential substrates for ClpS2

Until recently, the only substrates for EcClpS that had been identified were PATase and Dps (Ninnis 2009, Schmidt 2009). Of these, Dps was originally identified as a substrate for the ClpXP protease, but was later demonstrated to be modified *in vivo* to become a substrate for ClpAPS by removal of the first five N-terminal amino acids and thereby creating an N-degron recognition signal (Flynn et al. 2003, Schmidt et al. 2009). A *Synechococcus* ortholog to this protein, DpsA (Pena and Bullerjahn 1995) has a potential N-degron site at the same location as Dps in *E. coli*. Since ClpS2 (but not ClpS1) degrades LVK-tagged proteins *in vitro*, it seems likely that *Synechococcus* DpsA is an *in vivo* substrate for ClpS2. This proposal is strengthened by the observation that DpsA does not accumulate in the *Synechococcus ΔclpS1* line, which is inconsistent with ClpS1 targeting DpsA for degradation *in vivo* (Ståhlberg et al. in preparation).

Since ClpS2 recognized LVK-tagged proteins *in vitro*, we searched the *Synechococcus* proteome for proteins containing the sequence LVK at or close to the N-terminus to hopefully identify potential substrates. Ten candidates were eventually found, several of which were hypothetical proteins of no known function. Included in the candidates with known function was an arabinose efflux permease that functions in other bacteria as a transmembrane transporter protein and has been linked to drug resistance, the repressor of the heat shock response HrcA (Zuber et al., 1994, Roberts et al., 1996), anthranilate phosphoribosyl transferase involved in aromatic amino acid synthesis (Mayans et al. 2002), and the M-subunit of the proton-translocating NADH-quinone oxidoreductase NDH-1 (Battchikova et al. 2011).

In Paper III, we demonstrate that the ClpS paralog so far unique to cyanobacteria (ClpS2) functions as an adaptor protein for the essential ClpCP3/R protease in vitro. We also reveal that ClpS1 and ClpS2 have different substrate specificity in vitro. We also show that ClpS1, although considered as the ortholog of EcClpS, does not recognize Dps, a known substrate for EcClpS. Dps is instead targeted for degradation by ClpS2 in vitro, making this protein a likely in vivo substrate for ClpS2.

4. Future prospects and remaining questions.

The studies presented in this thesis have shown that ClpR is important for the interaction with ClpC (**Paper II**) and the assembly of the ClpP3/R proteolytic core (**Paper I**). Despite this, the underlying reason for the evolution of ClpR and what selective advantage it confers to the Clp protease in photosynthetic organisms remains a mystery. The architectural similarities between the eukaryotic 26S proteasome and the Clp proteases in photosynthetic organisms is also intriguing, particularly the inclusion of identical amounts of inactive subunits within the proteolytic core. Does the incorporation of inactive subunits somehow confer tighter regulation of proteolysis, or do they enable a more diverse range of substrates to be recognized? Also of interest is the appearance of multiple ClpR paralogs in photosynthetic eukaryotes. As previously mentioned, *Arabidopsis* has four ClpR paralogs that assemble into the one Clp proteolytic core in the chloroplast, suggesting that the role of ClpR has been reinforced during evolution. The fact that the ClpR-containing Clp proteases are essential for cell viability in cyanobacteria and photosynthetic eukaryotes further emphasizes the importance of determining the exact function of ClpR.

4.1. Structures of ClpP3/R and the ClpS adaptor proteins

A critical step in defining the exact role of ClpR will be the crystallization and structural resolution of a ClpR-containing proteolytic core such as ClpP3/R from *Synechococcus*. Many attempts have been made by our group to produce crystals of ClpP3/R of sufficient size and quality for x-ray diffraction studies but all have so far failed, although efforts are ongoing. Not until such atomic level structure is known will the exact interaction between the ClpR and ClpP3 subunits be resolved, and what affect the ClpR subunit has on the substrate-binding groove of the proteolytic core. It will also be interesting to compare the structures of ClpS1 and ClpS2 with those of EcClpS and other defined ClpS structures, and if the difference in substrate specificity observed *in vitro* could be explained by structural variations. Modeling can provide some clues to the functional difference between ClpS1 and ClpS2 but high resolution structures will probably be needed to support these proposals. Although we envisage few problems in crystallizing these small, soluble proteins, an alternative for analyzing these structures if difficulties arise could be Nuclear Magnetic Resonance.

4.2. *In vivo* substrates for the ClpCP3/R protease

Another important feature that requires continued effort is the identification of native protein substrates for the ClpCP3/R protease. Several studies have identified *in vivo* substrates for Clp proteases from different bacteria or plants using various trap techniques, pull-down assays or degradation assays (Flynn et al. 2003, Sjögren et al. 2006, Stanne et al. 2009, Nishimura et al. 2013, Humbard et al. 2013). We have attempted the trap approach by transforming into wild type *Synechococcus* a construct expressing the inactive ClpP3/R proteolytic core used in Paper I under an inducible promoter. However, this construct either produced too little of the core complex for purification from the transformed cells or proved lethal due to likely interference with the native ClpP3/R complex. An alternative approach would be to simply add the recombinant inactive ClpP3/R core with or without recombinant ClpC to large-scale cell protein extracts from wild type *Synechococcus*. The ClpP3/R trap would then be purified on a Ni^{2+}-affinity column and washed to remove nonspecific proteins. The ClpP3/R complex would then be eluted and disrupted with urea to release the protein substrates trapped inside. The sample would then once more be run on the Ni^{2+}-affinity column to bind the His-tagged Clp proteins, leaving the previously trapped substrates to flow through. The potential protein substrates could then be identified by mass spectrometry.

4.3. *In vivo* substrates for ClpS1 and ClpS2

Another remaining question is the identity of the *in vivo* substrates for ClpS1 and ClpS2. One *in vitro* approach to answer this question could be the use of large synthetic peptide libraries to screen for binding of either recombinant ClpS1 or ClpS2 using specific antibodies. This method has been successfully used on a smaller scale to investigate which N-terminal amino acids are recognized by EcClpS (Erbse et al. 2006) and we are in the process of collaborating with this same group to test the binding specificity of ClpS1 and ClpS2 using the same peptide library. A more comprehensive screen, however, would require larger peptide libraries, either commercially available or requiring custom synthesis, but these are likely to be prohibitively expensive.

To obtain information about the *in vivo* substrates of ClpS1 and ClpS2, an alternative approach will be needed. A recent study has identified almost 100 putative substrates for EcClpS using the recombinant protein immobilized on a column and applying cell extracts from *E. coli* (Humbard et al. 2013). Since large amounts of recombinant ClpS1 and ClpS2 with N-terminal His-tags can be readily purified, a similar approach to identifying *in vivo* substrates for the *Synechococcus* adaptor proteins could be used. If it proves necessary to use another affinity tag to immobilize the ClpS

proteins to a column, the ClpS1 and ClpS2 proteins could be expressed as fusion proteins with either glutathione-S transferase (GST) or the maltose-binding protein (MBP). The added advantage of using such fusion partners is that they can be used separately as a negative control for nonspecific binding to the column.

4.4. Is ClpS2 degraded by the ClpXP1/P2 protease *in vivo*?

The possibility that ClpS2 is a substrate for the ClpXP1/P2 protease is intriguing. Although no degradation of recombinant ClpS2 was observed when using ClpP1/P2 with EcClpX, this does not exclude the possibility that such degradation occurs *in vivo* with *Synechococcus* ClpX. Although recombinant ClpP1/P2 can be readily purified by over-expression, all attempts to purify *Synechococcus* ClpX in an active state have so far failed mainly due to its extreme insolubility (Ståhlberg et al. in preparation). As an alternative, a chimeric form of the more soluble EcClpX could be prepared in which the N-terminal domain responsible for substrate recognition and binding could be replaced with the corresponding region from *Synechococcus* ClpX. If this chimeric ClpX can be purified and remain active, it could then be used to test if ClpS2 is indeed a substrate of the ClpXP1/P2 protease *in vitro*.

4.5. Which regions in ClpS1 and ClpS2 are important for substrate recognition?

Site-directed mutagenesis could be used to investigate which residues in ClpS1 and ClpS2 are important for substrate recognition. An obvious target would be the Phe40 and Thr62 residues in ClpS2 and testing if changing them to Met would change the substrate specificity to that of ClpS1 or EcClpS. Another conserved motif of interest is the C-terminal MAPLE of ClpS2, and whether mutating this sequence would affect the substrate specificity of ClpS2 or its association to ClpC.

5. References

Adam Z., Adamska I., Nakabayashi K., Ostersetzer O., Haussuhl K., Manuell A., Zheng B., Vallon O., Rodermel S.R., Shinozaki K., Clarke AK. (2001) Chloroplast and mitochondrial proteases in Arabidopsis. A proposed nomenclature. *Plant Physiol*. **125**: 1912-8.

Akiyama Y., Ogura T., Ito K. (1994). Involvement of FtsH in protein assembly into and through the membrane. I. Mutations that reduce retention efficiency of a cytoplasmic reporter. *J. Biol. Chem.* **269**: 5218-24

Andersson F.I., Blakytny R., Kirstein J., Turgay K., Bukau B., Mogk A., Clarke A.K.. (2006). Cyanobacterial ClpC/HSP100 Protein Displays Intrinsic Chaperone Activity." The *J.Biol.Chem.* 281: 5468–75.

Apel W., Schulze W.X., Bock R. (2010). Identification of protein stability determinants in chloroplasts. *Plant. J.* **63**: 636–50.

Bachmair A., Becker F., Schell J. (1993) Use of a reporter transgene to generate arabidopsis mutants in ubiquitin-dependent protein degradation. *Proc Natl Acad Sci U.S. A.* **90**:418-21.

Bailey TL; Elkan C 1994 "Fitting a mixture model by expectation of maximalization to discover motifs in polymers".

Baker, T.A, and Sauer, R.T. (2006). ATP-Dependent Proteases of Bacteria: Recognition Logic and Operating Principles. *Trends. Biochem. Sci* **31**: 647–53.

Banecki B., Wawrzynow A., Puzewicz J., Georgopoulos C., Zylicz M. (2001). Structure-function analysis of the zinc-binding region of the Clpx molecular chaperone.*J. Biol. Chem*. **276** :18843-8.

Battchikova, N., Eisenhut, M., Aro, E. (2011). Cyanobacterial NDH-1 complexes : Novel insights and remaining puzzles . *Biochim. Biophys. Acta*. **1807:** 935–944.

Baugh J.M., Viktorova E.G., Pilipenko E.V (2009). Proteasomes can degrade a significant proportion of cellular proteins independent of ubiquitination. *J Mol Biol.* **386**: 814-27.

Bachmair A., Finley D., Varshavsky A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science.* **234**: 179-86.

Baumeister W., Walz J., Zühl F., Seemüller E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell.* **92**: 367-80.

Bochtler M., Hartmann C., Song H. K., Bourenkov G. P., Bartunik H. D., Huber,R.(2000).The structures of HslU and the ATP- dependent protease HslU – HslV, *Nature.* **403**: 13091–13096.

Bolon D.N, Wah D.A, Hersch G.L, Baker T.A, Sauer R.T. (2004). Bivalent tethering of SspB to ClpXP is required for efficient substrate delivery: a protein-design study. *Mol. Cell*. **13**: 443-9.

Botos I., Melnikov E.E., Cherry S., Khalatova A.G., Rasulova F.S., Tropea J.E., Maurizi M.R., Rotanova T.V., Gustchina A., Wlodawer A. (2004). Crystal structure of the AAA+ alpha domain of E. coli **Lon** protease at 1.9A resolution. *J. Struct. Biol.* **146**:113-22.

Bewley M. C., Graziano V., Griffin K., Flanagan J. M. (2006). The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. *J Struct. Biol*. **153**: 113–128.

Burns KE, Darwin KH (2010). Pupylation versus ubiquitylation: tagging for proteasome-dependent degradation. *Cell Microbiol*. **12**: 424-31.

Burton B.M., Baker T.A. (2005). Remodeling protein complexes: insights from the AAA+ unfoldase ClpX and Mu transposase. *Protein Sci*. **14** :1945-54.

Cha S.S., An Y.J., Lee C.R., Lee H.S., Kim Y.G., Kim S.J., Kwon K.K., De Donatis G.M., Lee J.H., Maurizi M.R., Kang S.G. (2010) . Crystal structure of Lon protease: molecular architecture of gated entry to a sequestered degradation chamber. *EMBO J*. **29**: 3520-30.

Chaney, M., Grande, R., Wigneshweraraj, S.R., Cannon, W., Casaz, P, Gallegos, M.T.,Schumacher, J., Jones, S., Elderkin, S., Dago, A.E., Morett, E., Buck, M **(**2001). Binding of Transcriptional Activators to Sigma 54 in the Presence of the Transition State Analog ADP-Aluminum Fluoride: Insights into Activator Mechanochemical Action. *Gene. Dev.* **15** 2282–94.

Charette M.F, Henderson G.W, Markovitz A. (1981) ATP hydrolysis-dependent protease activity of the lon (capR) protein of Escherichia coli K-12. *Proc. Natl. Acad. Sci. U. S .A*. **78**: :4728-32.

Chuang S.E., Burland V., Plunkett G 3rd., Daniels D.L., Blattner F.R. Sequence analysis of four new heat-shock genes constituting the hslTS/ibpAB and hslVU operons in *Escherichia coli*. *Gene*. **134**:1-6.

Clarke A.K**.**, Schelin J., Porankiewicz J. (1998). Inactivation of the clpP1 gene for the proteolytic subunit of the ATP-dependent Clp protease in the cyanobacterium Synechococcus limits growth and light acclimation. *Plant. Mol. Biol*. **37**: 791-801.

Constan D., Froehlich J.E., Rangarajan S., Keegstra K. (2004) A stromal Hsp100 protein is required for normal chloroplast development and function in Arabidopsis. *Plant Physiol*. **136**: 3605-15.

Damerau K., St John A.C. (1993). Role of Clp protease subunits in degradation of carbon starvation proteins in *Escherichia coli*. *J.Bacteriol*. **175**:53-63.

Darwin K.H (2009). Prokaryotic ubiquitin-like protein (Pup), proteasomes and pathogenesis. *Nat Rev Microbiol*. **7**: 485-91..

De Donatis G.M, Singh S.K., Viswanathan S., Maurizi M.R. (2010). A single ClpS monomer is sufficient to direct the activity of the ClpA hexamer. *J. Biol. Chem*. **285**:8771-81.

Derré I., Rapoport G., Devine K., Rose M., Msadek T. (1999) ClpE, a novel type of HSP100 ATPase, is part of the CtsR heat shock regulon of Bacillus subtilis. *Mol Microbiol*. **32**: 581-93.

Dines M., Sendersky E., Schwarz R., Adir N. (2006). Crystallization of sparingly soluble stress-related proteins from cyanobacteria by controlled urea solublization. *J. Struct.Biol*. **158**:116-21.

Dougan, D. a, Mogk, A., Zeth, K., Turgay, K., & Bukau, B. (2002a). AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Letters*, **529**, 6–10.

Dougan, D.A., Reid, B.G., Horwich, A.L., Bukau, B. (2002b) ClpS, a Substrate Modulator of the ClpAP Machine. *Mol. Cell* **9**: 673–683

Dougan D.A., Truscott K.N., Zeth K. (2010). The bacterial N-end rule pathway: expect the unexpected *Mol. Microbiol.* **76**: 545-58.

Dougan D.A., Micevski D., Truscott K.N. (2012). N-end rule pathway: from recognition by N-recognins, to destruction by AAA+proteases. *Biochim. Biophys. Acta*. **1823**: 83-91.

Ellis, R.J. 2006. "Molecular Chaperones: Assisting Assembly in Addition to Folding. *Trends in Biochemical Sciences* **31**: 395–401.

Erbse A., Schmidt R., Bornemann T., Schneider-Mergener J., Mogk A., Zahn R., Dougan D.A., Bukau B. (2006) ClpS Is an Essential Component of the N-End Rule Pathway in *Escherichia Coli*. *Nature.* **439**: 753–6.

Eriksson, M.J, Clarke, A.K (1996). The Heat Shock Protein ClpB Mediates the Development of Thermotolerance in the Cyanobacterium Synechococcus Sp. Strain PCC 7942. *J. Bacteriol.* **178**: 4839–46.

Eriksson, M.J., Schelin, J., Miskiewicz, E.W.A., Clarke, A.K (2001). Novel Form of ClpB / HSP100 Protein in the Cyanobacterium Synechococcus *J Bacteriol* **183**: 7392–7396.

Farrell C.M., Grossman A.D., Sauer R.T. (2005). Cytoplasmic degradation of ssrA-tagged proteins. *Mol Microbiol.* **57**: 1750-61.

Farrell C.M., Baker T.A., Sauer R.T. (2007). Altered specificity of a AAA+ protease.. *Mol Cell*. **25**: 161-6.

Flanagan J.M., Wall J.S., Capel M.S., Schneider D.K., Shanklin J. (1995) Scanning transmission electron microscopy and small-angle scattering provide evidence that native *Escherichia coli* ClpP is a tetradecamer with an axial pore. *Biochemistry*. **34:** 10910-7.

Finlay D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem*.**78**: 477-513.

Flores-Pérez Ú., Jarvis P (2013). Molecular chaperone involvement in chloroplast protein import. *Biochim Biophys Acta*. **1833**: 332-40.

Flynn J.M., Levchenko I., Seidel M., Wickner S.H, Sauer R.T., Baker T.A. (2001). Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc. Natl. Acad. Sci. U. S. A* . **98**:10584-9.

Flynn J.M., Neher S.B., Kim Y.I., Sauer R.T., Baker T.A. (2003). Proteomic Discovery of Cellular Substrates of the ClpXP Protease Reveals Five Classes of ClpX-Recognition Signals. *Mol. Cell* **11**: 671–83.

Flynn J.M., Levchenko I., Sauer R.T., Baker T.A. (2004). Modulating substrate choice: the SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation. *Genes. Dev*. **18**: 2292-301.

Forouzan D., Ammelburg M., Hobel C.F., Ströh L.J., Sessler N., Martin J., Lupas A.N. (2012). The archaeal proteasome is regulated by a network of AAA ATPases. *J.Biol.Chem.* **287**:39254-62.

Führer F., Langklotz S., Narberhaus F. (2006) The C-terminal end of LpxC is required for degradation by the FtsH protease. *Mol. Microbiol*. **59**:1025-36.

Führer F., Müller A., Baumann H., Langklotz S., Kutscher B., Narberhaus F. (2007). Sequence and length recognition of the C-terminal turnover element of LpxC, a soluble substrate of the membrane-bound FtsH protease.*J. Mol. Biol*.**372**::485-96.

García-Lorenzo M., Sjödin A., Jansson S., Funk C. (2006). Protease gene families in Populus and Arabidopsis. *BMC. Plant. Biol*. **6**: 30.

Glickman M.H.,Rubin D.M., Fried V.A., Finley D. (1998) The regulatory particle of the Saccharomyces cerevisiae proteasome. *Mol Cell Biol.* **18**:3149-62.

Glynn S.E., Martin A., Nager A.R., Baker T.A., Sauer R.T. (2009). Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell.* **139**: 744-56.

Goldberg A.L., Moerschell RP., Chung C.H., Maurizi M.R. (1994) ATP-dependent protease La (lon) from Escherichia coli. *Methods Enzymol.* **244**: 350-75

Gonda D.K., Bachmair A., Wünning I., Tobias J.W., Lane W.S., Varshavsky A. (1989) Universality and structure of the N-end rule. *J.Biol.Chem.* **264**: 16700-12.

Gonzalez M., Frank E.G., Levine A.S., Woodgate R. (1998). Lon-mediated proteolysis of the Escherichia coli UmuD mutagenesis protein: in vitro degradation and identification of residues required for proteolysis. *Genes. Dev.* **12**: 3889-99.

Gottesman, S., Clark, W.P., Maurizi M.R. (1990). The ATP-dependent Clp Protease of Escherichia coli. Sequence of ClpA and identification of a Clp-specific substrate. *J.Bio.Chem*. **265**: 7886-93.

Gottesman S, Maurizi MR.(1992). Regulation by proteolysis: energy-dependent proteases and their targets. *Microbiol Rev*. **56** 592-621.

Gottesman S., Clark W.P., Crecy-lagard V De., Maurizi R. (1993) ClpX , an Alternative Subunit for the ATP-dependent Clp Protease of Escherichia coli. *J. Biol. Chem*. **268**: 22618–26.

Gottesman, S. (1996). Proteases and Their Targets in Escherichia Coli. *Annu. Rev. Gen.* **30**: 465–506.

Gottesman S., Roche E., Zhou Y., Sauer R.T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes. Dev*.**12** 1338-47.

Gribun A., Kimber M. S., Ching R., Sprangers R., Fiebig K. M., Houry, W.A. (2005). The ClpP double ring tetradecameric protease exhibits plastic ring-ring interactions, and the N termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation. *J.Biol.Chem*, **280**: 16185–96.

Groll M., Ditzel L., Löwe J., Stock D., Bochtler M., Bartunik H.D., Huber R . (1997). Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature*. **386**: 463-71.

Gur E., Sauer R.T. (2008). Recognition of misfolded proteins by Lon, a AAA(+) protease. *Genes. Dev*. **22**: 2267-77.

Gur E., Ottofueling R., Dougan D.A. (2013). *Regulated Proteolysis in Microorganisms* Dougan D.A, editor. Dordrecht: Springer Netherlands.

Hanson, P.I, and Whiteheart, S.W. (2005) AAA+ Proteins: Have Engine, Will Work. *Nature Reviews. Molecular Cell Biology* **6** 519–29.

Hengge-Aronis R. (2000). A role for the sigma S subunit of RNA polymerase in the regulation of bacterial virulence. *Adv. Exp. Med. Biol*. **485**: 85-93.

Hepowit N.L., Uthandi S., Miranda H.V., Toniutti M., Prunetti L., Olivarez O., De Vera I.M., Fanucci G.E., Chen S., Maupin-Furlow J.A. (2012). Archaeal JAB1/MPN/MOV34 metalloenzyme (HvJAMM1) cleaves ubiquitin-like small archaeal modifier proteins (SAMPs) from protein-conjugates. *Mol. Microbiol*. **86**: 971-87.

Herman C., Ogura T., Tomoyasu T., Hiraga S., Akiyama Y., Ito K., Thomas R., D'Ari R., Bouloc P. (1993). Cell growth and lambda phage development controlled by the same essential *Escherichia coli* gene, ftsH/hflB. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 10861-5.

Herman C., Thévenet D., Bouloc P., Walker G.C., D'Ari R.(1998).Degradation of carboxy-terminal-tagged cytoplasmic proteins by the Escherichia coli protease HflB (FtsH). *Genes. Dev*. **12**: 1348-5

Hou J.Y., Sauer R.T., Baker T.A. (2008). Distinct Structural Elements of the Adaptor ClpS Are Required for Regulating Degradation by ClpAP. *Nat. Struct. Mol. Biol.* **15**: 288–294.

Humbard M.A., Miranda H.V., Lim J.M, Krause D.J., Pritz J.R., Zhou G., Chen S., Wells L., Maupin-Furlow J.A. (2010). Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *Nature*. ***463***: 54-60.

Humbard M.A., Surkov S., De Donatis G.M., Jenkins L.M., Maurizi M.R. (2013). The N-degradome of Escherichia coli: limited proteolysis in vivo generates a large pool of proteins bearing N-degrons. *J. Biol. Chem.* **288** :28913-24.

Hwang, C.S., Shemorry A,. Varshavsky, A. (2010). N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science.* 327: 973–7.

.Ingvarsson H., Maté M.J., Högbom M., Portnoï D., Benaroudj N., Alzari P.M., Ortiz-Lombardía M., Unge T. (2007). Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of Mycobacterium tuberculosis ClpP1. *Acta. Crystallogr. D. Biol. Crystallogr*. **63** 249-59.

Jennings L.D., Bohon J., Chance M.R., Licht S., (2008a). The ClpP N-terminus coordinates substrate access with protease active site reactivity. *Biochemistry* **47**, 11031-40

Jennings L.D., Lun D.S., Médard M., Licht S. (2008b). ClpP hydrolyzes a protein substrate processively in the absence of the ClpA ATPase: mechanistic studies of ATP-independent proteolysis. *Biochemistry*. **47**:11536-46.

Joshi S.A., Hersch G.L., Baker T.A., Sauer R.T. (2004). Communication between ClpX and ClpP during substrate processing and degradation. *Nat. Struct. Mol. Biol*. **11**: 404-11.

Katayama, Y., Gottesmans, S., Pumphreyll, J., Rudikoffl, S., Clark, W.P., and Maurizi, M.R (1988) The Two-Component , ATP-Dependent Clp Protease of Escherichia Coli **263:**15226–15236.

Kang S.G, Maurizi MR, Thompson M., Mueser T., Ahvazi B..(2004). Crystallography and mutagenesis point to an essential role for the N-terminus of human mitochondrial ClpP. *J Struct Biol*. **148**: 338-52.

Kato Y., Miura E., Ido K., Ifuku K., Sakamoto W. (2009).The variegated mutants lacking chloroplastic FtsHs are defective in D1 degradation and accumulate reactive oxygen species. *Plant. Physiol.* **151**: 1790-801.

Kato., Y., Sun, X., Zhang., L., Sakamoto., W. (2012). Cooperative D1 degradation in the photosystem II repair mediated by chloroplastic proteases in Arabidopsis. *Plant. Physiol*, **159**:1428–39.

Katz C., Ron E.Z (2008) Dual role of FtsH in regulating lipopolysaccharide biosynthesis in *Escherichia coli. J. Bacteriol.*;**190**::7117-22.

Keiler K.C., Waller P.R., Sauer R.T (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science.* **271**: 990-3.

Kessel M., Maurizi M.R., Kim B., Kocsis E., Trus B.L., Singh SK., Steven A.C. (1995) Homology in structural organization between E. coli ClpAP protease and the eukaryotic 26 S proteasome.*J Mol Biol* **250**:587-94

Kessel M., Wu W., Gottesman S., Kocsis E., Steven A.C., Maurizi M.R. (1996). Six-fold rotational symmetry of ClpQ, the E. coli homolog of the 20S proteasome, and its ATP-dependent activator, ClpY. *FEBS Lett.* **398**: 274-8.

Kim DY, Kim KK.(2008). The structural basis for the activation and peptide recognition of bacterial ClpP. *J. Mol. Biol*.**379**:760-71.

Kim Y.I., Levchenko I., Fraczkowska K., Woodruff R.V., Sauer R.T., Baker T. A (2001) Letters Molecular Determinants of Complex Formation between Clp / Hsp100 ATPases and the ClpP Peptidase. *Nat. Struct. Biol*. **8**: 268–271

Kihara A., Akiyama Y., Ito K. (1995). FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proc. Natl. Acad. Sci U. S. A.* **92**: 4532-6.

Kirstein J., Schlothauer T., Dougan D.A., Lilie H., Tischendorf G., Mogk A., Bukau B., Turgay K. (2006) Adaptor protein controlled oligomerization activates the AAA+ protein ClpC*. EMBO J* **25**:1481-91

Koodathingal P., Jaffe N.E., Kraut D.A., Prakash S., Fishbain S., Herman C., Matouschek A..(2009). ATP-dependent proteases differ substantially in their ability to unfold globular proteins. *J. Biol. Chem*. ;**284**: 18674-84.

Kress, Wolfgang, Zeljka Maglica, and Eilika Weber-Ban. 2009. "Clp Chaperone-Proteases: Structure and Function." *Research in Microbiology* 160 (9) (November): 618–28..

Krüger E., Völker U., Hecker M (1994) Stress induction of clpC in Bacillus subtilis and its involvement in stress tolerance. . *J Bacteriol* **76**:3360-7.

Kuo M.S., Chen K.P., Wu W.F. (2004). Regulation of RcsA by the ClpYQ (HslUV) protease in *Escherichia coli*. *Microbiology.* **150**: 437-46.

Kwon Y.T, Reiss Y., Fried V.A, Hershko A, Yoon J.K, Gonda D.K, Sangan P, Copeland N.G, Jenkins N.A., Varshavsky A. (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl. Acad. Sci. U. S. A.* **95** 7898-903.

Lee B.G., Kim M.K., Song H.K. (2011). Structural insights into the conformational diversity of ClpP from *Bacillus subtilis*. *Mol. Cells*. **32**: 589-95.

Levchenko I, Yamauchi M, Baker TA. (1997) . ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. *Genes. Dev*. **11**:1561-72.

Lies M., Maurizi M.R. (2008). Turnover of endogenous SsrA-tagged proteins mediated by ATP-dependent proteases in *Escherichia coli*. *J. Biol. Chem.* **283:** 22918-29.

Loewen P.C., Hengge-Aronis R. (1994) The role of the sigma factor sigma S (KatF) in bacterial global regulation. *Annu. Rev. Microbiol*. **48:** 53-80.

López-Otín, C., & Bond, J. (2008). Proteases: multifunctional enzymes in life and disease. *J.Biol.Chem* **283** 30433–3043

Lupas A., Zühl F., Tamura T., Wolf S., Nagy I., De Mot R., Baumeister W. (1997).Eubacterial proteasomes. *Mol Biol Rep.* **24:** 125-31.

Lupas A.N., Koretke K.K (2003). Bioinformatic Analysis of ClpS, a Protein Module Involved in Prokaryotic and Eukaryotic Protein Degradation. *J. Struct. Biol.* **141**: 77–83.

Löwe J., Stock D., Jap B., Zwickl P., Baumeister W., Huber R. (1995). Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. *Science.* **268**:533-9.

Martin W., Stoebe B., Goremykin V., Hapsmann S., Hasegawa M., Kowallik K.V (1998).Gene transfer to the nucleus and the evolution of chloroplasts. *Nature.* **393**: 162-5.

Martin A., Baker T.A., Sauer R.T. (2005). Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature.* **437**: 1115-20.

Martin A., Baker T.A., Sauer R.T (2007) Distinct Static and dynamic interactions control ATPAse-peptidase communication in a AAA+ protease *Mol. Cell.* **27**: 41-52

Martin A., Baker T.A., Sauer R.T. (2008a). Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol. Cell*. **29**: 441-50.

Martin A., Baker T.A, Sauer R.T. (2008b) Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat. Struct. Mol. Biol.* **15**: 1147-51

Maurizi M. R., Clark W. P., Katayama Y., Rudikoff S., Pumphrey J., Bowers B., Gottesman, S. (1990). Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of Escherichia coli. *J.Biol.Chem*, ***265***(21), 12536–12545.

Maurizi M.R. (1991) ATP-promoted interaction between ClpA and ClpP in activation of Clp protease from *Escherichia coli*. *Biochem. Soc. Trans*.**19**:719-23.

Mayans O., Ivens A., Nissen L. J., Kirschner K., Wilmanns., M. (2002). Structural analysis of two enzymes catalysing reverse metabolic reactions implies common ancestry. *EMBO J*, **21**. 3245–54.

Miethke M., Hecker M., Gerth U. (2006). Involvement of Bacillus subtilis ClpE in CtsR degradation and protein quality control. *J Bacteriol*. **188**: 4610-9.

Miranda H V., Nembhard N., Su D., Hepowit N., Krause D.J, Pritz J.R., Phillips C., Söll D., Maupin-Furlow J.A. (2011) E1- and ubiquitin-like proteins provide a direct link between protein conjugation and sulfur transfer in archaea. *Proc Natl Acad Sci U.S. A.* **108** :4417-22.

Mogk A., Schmidt R., Bukau B., (2007). The N-End Rule Pathway for Regulated Proteolysis: Prokaryotic and Eukaryotic Strategies. *Trend. Cel.l Biol.* **17**: 165–172.

Nagpal J., Tan J.L., Truscott K.N., Heras B, Dougan D.A. (2013). Control of protein function through regulated protein degradation: biotechnological and biomedical applications. *J.Mol. Microbiol. Biotechnol*. **23** :335-44

Neher S.B., Sauer R.T., Baker T.A. (2003). Distinct peptide signals in the UmuD and UmuD' subunits of UmuD/D' mediate tethering and substrate processing by the ClpXP protease. *Proc Natl Acad Sci U.S. A.* **100**: 13219-24.

Neher S.B., Villén J., Oakes E.C., Bakalarski C.E., Sauer R.T., Gygi S.P., Baker T.A. (2006). Proteomic profiling of ClpXP substrates after DNA damage reveals extensive instability within SOS regulon. *Mol. Cell*. **22**:193–204.

Neuwald, A. F., Aravind, L., Spouge, J. L., & Koonin, E. V. (1999). Assembly , Operation , and Disassembly of Protein Complexes AAA + A Class of Chaperone-Like ATPases Associated with the Assembly , Operation , and Disassembly of Protein Complexes, *Genome Res* **9** 27–43.

Ninnis R.L., Spall S.K., Talbo G.H., Truscott K.N., Dougan D.A (2009) Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in Escherichia coli. *EMBO J.* **28**: 1732-44.

Nishimura K., Asakura Y., Friso G., Kim J., Oh S.H., Rutschow H., Ponnala L., van Wijk K.J. (2013). ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in Arabidopsis. *Plant Cell.* **25**: 2276-301.

Ogura T., Inoue K., Tatsuta T., Suzaki T., Karata K., Young K., Su L.H, Fierke C.A, Jackman J.E, Raetz C.R, Coleman J, Tomoyasu T, Matsuzawa H. (1999). Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in *Escherichia coli. Mol. Microbiol*. **31**:833-44.

Ortega J., Singh S.K, Ishikawa T., Maurizi M.R, Steven A.C (2000). Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Mol Cell*. **6** :1515-21.

Ortega J., Lee H.S, Maurizi M.R, Steven A.C.(2002). Alternating translocation of protein substrates from both ends of ClpXP protease. *EMBO J*. **21**: 4938-49.

Ostersetzer O., Kato Y., Adam Z., Sakamoto W. (2007). Multiple intracellular locations of Lon protease in Arabidopsis: evidence for the localization of AtLon4 to chloroplasts. *Plant. Cell. Physiol.* 2007 **48**: 881-5.

Park S., Rodermel S.R 2004 Mutations in ClpC2/Hsp100 suppress the requirement for FtsH in thylakoid membrane biogenesis. *Proc Natl Acad Sci U.S. A.* **101**:12765-70

Pak M., Wickner S. (1997) Mechanism of protein remodeling by ClpA chaperone *Proc Natl Acad Sci U S A* **94**: 4901-6

Pena M. M. O., Bullerjahn, G. S. (1995). The DpsA Protein of Synechococcus sp. Strain PCC7942 Is a DNA-binding Hemoprotein: LINKAGE OF THE Dps AND BACTERIOFERRITIN PROTEIN FAMILIES. *J. Biol. Chem.* **270**. 22478–22482.

Peltier J-B., Ripoll D.R., Friso G., Rudella A., Cai Y., Ytterberg J., Giacomelli L., Pillardy J., van Wijk K.J. (2004). Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. *J Biol Chem*.**279**: 4768–81.

Pietrosiuk A., Lenherr E.D., Falk S., Bönemann G., Kopp J., Zentgraf H., Sinning I., Mogk A. (2011). Molecular basis for the unique role of the AAA+ chaperone ClpV in type VI protein secretion. *J.Biol.Chem.* **286**: 30010-21.

Plas J.Van Der., Hegeman H., Vrieze G De., Tuyl M., Borrias M., Weisbeek P. (1990). Genomic integration system based on pBR322 sequences for the eyanobacterium Syneckococcus sp. PCC7942: transfer o f genes encoding plastoeyanin and ferredoxin (Recombinant DNA; homologous recombination; integration platform; plasmid rescue; Anabaena) *Gene*. **95**: 39–48.

Porankiewicz J., Wang J., Clarke A.K. (1999). New insights into the ATP-dependent Clp protease: Escherichia coli and beyond. *Mol Microbiol.* **32**: 449-58.

Ravid T., Hochstrasser M. (2008) Diversity of degradation signals in the ubiquitin-proteasome system. *Nat Rev Mol Cell Biol*. **9:** 679-90.

Roberts R.C., Toochinda C., Avedissian M,. Baldini R.L,. Gomes S.L,. Shapiro L (1996). Identification of a Caulobacter crescentus operon encoding hrcA, involved in negatively regulating heat-inducible transcription, and the chaperone gene grpE. *J. Bacteriol*. **178**: 1829-41.

Rohrwild M., Coux O., Huang H.C., Moerschell R.P., Yoo S.J., Seol J.H., Chung C.H., Goldberg A.L. (1996). HslV-HslU: A novel ATP-dependent protease complex in Escherichia coli related to the eukaryotic proteasome. *Proc Natl Acad Sci U.S. A.* **93**: 5808-13.

Román-Hernández G, Hou J.Y, Grant R.A., Sauer R.T., Baker T.A.(2011) The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. *Mol. Cell* **43:** 217–28.

Roos D.S., Crawford M.J, Donald R.G., Fraunholz M., Harb O.S., He C.Y., Kissinger J.C, Shaw M.K., Striepen B (2002) Mining the Plasmodium genome database to define organellar function: what does the apicoplast do? Philos. *Trans. R. Soc. Lond. B. Biol. Sci*. **357**: 35-46.

Rosano G.L., Bruch E.M., Ceccarelli E.A. (2011)Insights into the Clp/HSP100 chaperone system from chloroplasts of Arabidopsis thaliana. *J.Biol.Chem.* **286**: 29671-80.

Rotanova T.V., Melnikov E.E., Khalatova A.G., Makhovskaya O.V., Botos I., Wlodawer A., Gustchina A. (2004). Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains. *Eur. J. Biochem*. **271**: 4865-71.

Rugarli E.I., Langer T. (2006). Translating m-AAA protease function in mitochondria to hereditary spastic paraplegia. *Trends. Mol. Med.* **12**: 262-9.

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

Sauer, R. T., Bolon, D.N., Burton, B.M., Burton, R.E., Flynn, J.M., Grant, R.A, Hersch, G.L., Joshi, S.A., Kenniston, J.A., Levchenko, I., Neher, S.B., Oakes, E. S.C., Siddiqui, S.M., Wah, D.A., Baker, T.A. (2004). Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell*, **119,** 9–18.

Sauer, R. T., & Baker, T. a. (2011). AAA+ proteases: ATP-fueled machines of protein destruction. *Annual Review of Biochemistry*, *80*, 587–612. doi:10.1146/annurev-biochem-060408-172623

Schelin J., Lindmark F., Clarke A.K. (2002). The clpP multigene family for the ATP-dependent Clp protease in the cyanobacterium Synechococcus. *Microbiology.* **148**: 2255-65.

Schiener J., Witt S., Hayer-Hartl M., Guckenberger R. (2005). How to orient the functional GroEL-SR1 mutant for atomic force microscopy investigation. *Biochem. Biophys. Res. Commun*. **328**: 477-83.

Schirmer, E C, J R Glover, M a Singer, and S Lindquist, HSP100/Clp Proteins: A Common Mechanism Explains Diverse Functions. (1996) *Trends Biochem Sci.*, **21** 289–96

Schmidt R., Zahn R., Bukau B., Mogk. A. (2009) ClpS Is the Recognition Component for Escherichia Coli Substrates of the N-End Rule Degradation Pathway. *Mol. Microbiol.* **72**: 506–517.

Schwenkert S., Soll J., Bölter B. (2011). Protein import into chloroplasts--how chaperones feature into the game. *Biochim Biophys Acta*. **1808**: 901-11.

Seong I.S., Oh J.Y., Lee J.W., Tanaka K., Chung C.H. (2000). The HslU ATPase acts as a molecular chaperone in prevention of aggregation of SulA, an inhibitor of cell division in Escherichia coli. *FEBS Lett.* **477** 224-9.

Shanklin J., DeWitt N.D., Flanagan J.M. (1995). The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of *Escherichia coli* ClpP and ClpA: an archetypal two-component ATP-dependent protease. *Plant Cell*.**7**: 1713-22.

Shimohata N., Chiba S., Saikawa N., Ito K., Akiyama Y. (2002). The Cpx stress response system of Escherichia coli senses plasma membrane proteins and controls HtpX, a membrane protease with a cytosolic active site. *Genes. Cells.* **7**: 653-62.

Shin D.H., Lee C.S., Chung C.H., Suh S.W. (1996). Molecular symmetry of the ClpP component of the ATP-dependent **Clp** protease, an *Escherichia coli* homolog of 20 S proteasome. *J. Mol. Biol*. **262:** :71-6.

Shinagawa H., Iwasaki H., Kato T., Nakata A. (1988). RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. *Proc Natl Acad Sci U. S. A.* **85**: 1806-10.

Siddiqui S.M., Sauer R.T., Baker T.A. (2004). Role of the Processing Pore of the ClpX AAA+ ATPase in the Recognition and Engagement of Specific Protein Substrates. *Genes. Dev.* **18**: 369–74.

Singh S.K., Maurizi M.R. (1994) Mutational analysis demonstrates different functional roles for the two ATP-binding sites in ClpAP protease from *Escherichia coli. J Biol Chem* **269**:29537-45.

Singh S.K., Rozycki J., Ortega J., Ishikawa T., Lo J., Steven A.C, Maurizi M.R. (2001). Functional domains of the ClpA and ClpX molecular chaperones identified by limited proteolysis and deletion analysis. *J.Biol. Chem*. **276**: 29420-9.

Sjögren, L. L. E , Macdonald, T. M., Sutinen, S. and Clarke, A. K. (2004). Inactivation of the clpC1 Gene Encoding a Chloroplast Hsp100 Molecular Chaperone Causes Growth Retardation , Leaf Chlorosis , Lower Photosynthetic Activity , and a Specific Reduction in Photosystem Content , *Plant Physiol*.**136**:4114-26

Sjögren, L. L. E., Stanne, T. M., Zheng, B., Sutinen, S., & Clarke, A. K. (2006). Structural and functional insights into the chloroplast ATP-dependent Clp protease in Arabidopsis. *Plant Cell*. **18**: 2635–49

Sjögren, L. L. E., Clarke, A. K. (2011). Assembly of the chloroplast ATP-dependent Clp protease in Arabidopsis is regulated by the ClpT accessory proteins. *Plant Cell*, ***23***: 322–32.

Sjögren LL., Tanabe N., Lymperopoulos P., Khan N.Z., Rodermel S.R., Aronsson H, Clarke A.K. (2014) Quantitative analysis of the chloroplast molecular chaperone ClpC/Hsp93 in Arabidopsis reveals new insights into its localization, interaction with the Clp proteolytic core, and functional importance. *J.Biol.Chem.* **289**: 11318-30.

Smith D.M., Chang S.C., Park S., Finley D., Cheng Y., Goldberg A.L. (2007).

Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. *Mol Cell.* **27**: 731-44.

Sokolenko A., Pojidaeva E., Zinchenko V., Panichkin V., Glaser V.M., Herrmann R.G., Shestakov S.V. (2002). The gene complement for proteolysis in the cyanobacterium Synechocystis sp. PCC 6803 and Arabidopsis thaliana chloroplasts. *Curr. Genet*. **41**: 291-310.

Sonezaki S., Ishii Y., Okita K., Sugino T., Kondo A., Kato Y. (1995). Overproduction and purification of SulA fusion protein in *Escherichia coli* and its degradation by Lon protease in vitro. *Appl. Microbiol. Biotechnol*. **43**: 304-9.

Stanne, Tara M, Elena Pojidaeva, Fredrik I Andersson, and Adrian K Clarke. 2007. "Distinctive Types of ATP-Dependent Clp Proteases in Cyanobacteria." *J.Biol.Chem* **282**: 14394–402..

Stanne TM, Sjögren LL, Koussevitzky S, Clarke AK.(2009). Identification of new protein substrates for the chloroplast ATP-dependent Clp protease supports its constitutive role in Arabidopsis. *Biochem J*. **417**: 257-68.

Tao L., Biswas I. (2013) ClpL is required for folding of CtsR in Streptococcus mutans. *J Bacteriol.* **195**: 576-84..

Tasaki T., Mulder L.C., Iwamatsu A., Lee M.J., Davydov I.V., Varshavsky A., Muesing M., Kwon Y.T. (2005). A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol. Cell. Biol*. **25**: 7120-36.

Thompson M.W., Maurizi M.R. (1994) Activity and specificity of Escherichia coli ClpAP protease in cleaving model peptide substrates. *J. Biol. Chem*.**269**: 18201-8.

Tobias J.W, Shrader T.E., Rocap G., Varshavsky A. (1991) The N-end rule in bacteria. *Science.* **254**: 1374-7.

Tomoyasu T. , Yamanaka K., Murata K., Suzaki T., Bouloc P., Kato A., Niki H., Hiraga S., Ogura T. (1993) Topology and subcellular localization of FtsH protein in Escherichia coli. *J Bacteriol.***175**: 1352-7.

Tomoyasu, T., Gamer, J., Bukau, B., Kanemori, M., Mori, H., Rutman, A.J., Oppenheim, A.B., Yura, T., Yamanaka, K., Niki, H. "Escherichia Coli FtsH Is a Membrane-Bound, ATP-Dependent Protease Which Degrades the Heat-Shock Transcription Factor Sigma 32. (1995) *EMBO J .1* 4, 2551–60

Tran T.D., Kwon H.Y., Kim E.H., Kim K.W., Briles D.E., Pyo S., Rhee D.K. (2011) Decrease in penicillin susceptibility due to heat shock protein ClpL in Streptococcus pneumoniae. *Antimicrob Agents Chemother*. **55**: 2714-28

Tsilibaris V., Maenhaut-Michel G., Van Melderen L. (2006). Biological roles of the Lon ATP-dependent protease. *Res. Microbiol*. **157**: 701-13.

Turgay K., Hamoen LW., Venema G., Dubnau D (1997) Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of Bacillus subtilis. *Genes Dev*.**11** :119-28.

Tu G.F., Reid G.E, Zhang J.G, Moritz R.L, Simpson R.J. (1995). C-terminal extension of truncated recombinant proteins in Escherichia coli with a 10Sa RNA decapeptide. *J Biol Chem*. **270**: 9322-6.

Wang J., Hartling J. A., Flanagan J. M. (1997). Resolution The Structure of ClpP at 2 . 3 A Suggests a Model for ATP-Dependent Proteolysis *Cell* **91,** 447–456.

Wang KH, Oakes ES, Sauer RT, Baker TA. (2008) Tuning the strength of a bacterial N-end rule degradation signal *J. Biol. Chem*. **283**: 24600-7.

Wawrzynow A., Wojtkowiak D., Marszalek J., Banecki B., Jonsen M., Graves B., Georgopoulos C., Zylicz M. (1995) The ClpX heat-shock protein of *Escherichia coli,* the ATP-dependent substrate specificity component of the ClpP-ClpX protease, is a novel molecular chaperone. *EMBO J*.**14**:1867-77.

Weaver L.M., Froehlich J.E., Amasino R.M. (1999). Chloroplast-targeted ERD1 protein declines but its mRNA increases during senescence in Arabidopsis. *Plant Phys.* Apr;**119:** 1209–16.

Weibezahn J, Tessarz P, Schlieker C., Zahn R., Maglica Z, Lee S, Zentgraf H, Weber-Ban E.U., Dougan D.A, Tsai F.T, Mogk A, Bukau B. (2004) Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell.*;**119** : 653-65.

Weichart D., Querfurth N., Dreger M., Hengge-Aronis R. (2003). Global role for ClpP-containing proteases in stationary-phase adaptation of *Escherichia coli. J Bacteriol*. **185**: 115-25.

Wickner S., Gottesman S., Skowyra D., Hoskins J., McKenney K., Maurizi M.R (1994). A molecular chaperone, ClpA, functions like DnaK and DnaJ.. *Proc Natl Acad Sci U.S.A.* **91*:*** 12218-22.

Wickner, S. (1999). Posttranslational Quality Control: Folding, Refolding, and Degrading Proteins. *Science*, *286* 1888–1893.

Wojtyra U.A., Thibault G., Tuite A., Houry W.A. (2003) The N-terminal zinc binding domain of ClpX is a dimerization domain that modulates the chaperone function. *J.Biol.Chem.* **278**: 48981-90.

Varshavsky, A. (2003). The N-End Rule and Regulation of Apoptosis. Nat.Cell.Biol.**5**: 373–6

Varshavsky A. (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci*.

Zhang, D., Kato Y., Zhang L., Fujimoto M., Tsutsumi N., Sodmergen,. Sakamoto W. (2010). The FtsH protease heterocomplex in Arabidopsis: dispensability of type-B protease activity for proper chloroplast development. *The Plant Cell*. ***22***: 3710–25.

Zheng B., Halperin T., Hruskova-Heidingsfeldova O., Adam Z., Clarke A.K. (2002) Characterization of Chloroplast Clp proteins in Arabidopsis: Localization, tissue specificity and stress responses. *Physiol. Plant*. **114** :92-101.

Zimmerman S.B., Minton A.P. (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct*. **22**: 27-65.

Zuber, U., Schumann, W. (1994). CIRCE, a novel heat shock element involved in regulation of heat shock operon dnaK of Bacillus subtilis. *J.Bacteriol*. **176**: 1359–63.

Zwickl P., Ng D., Woo K.M, Klenk H.P., Goldberg A.L. (1999) An archaebacterial ATPase, homologous to ATPases in the eukaryotic 26 S proteasome, activates protein breakdown by 20 S proteasomes. *J.Biol.Chem.* **274**: 26008-14.

Populärvetenskaplig sammanfattning.

Under åren som har gått så har rätt många frågat mig vad det egentligen är som jag har jobbat med under min tid som doktorand. Mina försök att förklara detta för släkt och vänner har inte varit speciellt framgångsrika. Men jag hoppas kunna göra ett bättre försök nu.

Cellen är den mest grundläggande beståndsdelen i levande organismer. En cell består av många olika delar och ämnen som har olika funktion. En mycket viktig beståndsdel i cellen är proteiner. Proteiner består av långa kedjor av aminosyror som närmast skulle kunna liknas vid byggklossar. Dessa klossar sätts samman i olika långa kedjor, och ofta så måste sedan dessa kedjor veckas och vikas för att proteinet ska kunna fungera ordentligt.

Proteiner är förstås mycket viktiga för att organismer ska kunna fungera. Bland annat så utgör de huvuddelen av de enzymer som är med och reglerar cellens funktioner och även i själva funktionern. Enzymer är till exempel med och kopierar DNA vid celldelning, och tillverkar även det RNA som behövs för att tillverka nya proteiner. Men precis som alla verktyg så slits proteinerna när de används. När proteiner blir utslitna eller skadade eller inte längre behövs i cellen så måste de på något sätt tas bort. Utslitna proteiner kan annars klumpa ihop sig med andra proteiner, och detta kan i värsta fall leda till att cellens funktioner störs så mycket att den dör. För att skydda cellerna mot denna typ av skador så finns det **chaperoner** och **proteaser**. Chaperoner är en typ av proteiner vars uppgift är att se till att andra proteiner har rätt struktur eller form (ofta så är chaoperoner med när proteiner sätts samman, och hjälper dem att få rätt struktur vid vikningen). De fungerar som en slags kontrollmekanism och känner av ifall proteiner är felveckade och kan utgöra en fara för cellen genom att klumpa ihop sig och bilda aggregat. Chaperonerna kan då göra två saker, antingen kan de försöka ”vika tillbaka” de felveckade proteinerna så att de får rätt struktur och kan fungera igen. Om detta inte är möjligt så kan de kalla dit ett proteas. Detta är en grupp proteiner som bryter ned andra proteiner och gör det möjligt att återvinna aminosyrorna till att tillverka nya protein.

En grupp av protener som innehåller både proteaser och chaperoner är **Clp-proteinerna**. Dessa finns i de flesta organismer, från bakterier till djur och växter. Men det som utmärker Clp-proteinerna är att de verkar vara speciellt viktiga för organismer som kan fotosyntetiserna, det finns många fler Clp-proteriner i växter än i djur, t ex. Dock så har man mest studerat Clp-proteaser i bakterien *E.coli* tidigare, beroende på att Clp-proteinera här är relativt få och enkla så att det har varit (relativt) lätt att arbeta med dem.

Cyanobakterien *Synechococcus elongatus* (*Synechoccous*) som vi har arbetat med med har 11 st Clp-proteiner och minst två olika typer av Clp-komplex som fungerar som

proteaser. Ett av dessa är essentiellt, det vill säga det är nödvändigt för att dessa bakterier ska kunna överleva. Detta komplex har döpts till ClpP3/R. Det som gör detta komplex intressant att undersöka är att det består av två olika typer av protein, ClpP3 och ClpR. Dessa bildar ett komplex som sedan i sin tur binder till chaperonen ClpC för att kunna bryta ned proteiner. Mitt arbete har mest fokuserat på studier av ClpP3/R komplexet genom olika typer av biokemiska analyser. Vi har inte studerat funktionen av proteinet i cellerna utan har använt metoder för att massproducera ClpP3/R i en annan typ av bakterier för att sedan kunna analysera detta protein med s.k *in vitro* metoder. Vi har även använt samma typ av teknik för att framställa proteiner som är med och reglerar funktionen av ClpP3/R, s.k adaptorproteiner.

En av de mest spännande sakerna som vi har upptäckt gäller hur ClpP3/R-komplexet är uppbyggt. Vi kunde bestämma att ClpP3/R består av tre stycken ClpP3 och fyra stycken ClpR i en ring. Och att två av dessa ringar sedan bildar själva Clp-proteaset som bryter ned andra proteiner. Detta var första gången som ett Clp-proteas bestående av två olika Clp-proteiner kunde karakteriseras biokemiskt. Vi har också hittat ett antal områden (ska motiv) i ClpP3 och ClpR som är viktiga för att det ska kunna binda till varandra och bilda ett fungerande proteas, och för att detta proteas ska kunna binda till ClpC. Främst så hart vi gjort detta genom att modifiera ClpP3 och ClpR på olika sätt. Vi har nämligen bytt delar av ClpR mot delar av ClpP3 och vice versa, och sedan undersökt hur dessa förändringar har påverkat ClpP3/R. Vi upptäckte också att trots att ClpR är inaktivt på egen hand (det saknar de delar som behövs för att kunna bryta ned proteiner) så blir inte komplexet mer aktivt om man återaktiverar ClpR. Det betyder att ClpR troligen har andra funktioner som är viktiga för cellen, men vi vet fortfarande inte exakt vilka dessa är.

Vi har också kunnat identifiera det motiv i ClpC som är nödvändig för att ClpC ska kunna binda till ClpP3/R och bryta ned proteiner. Detta gjorde vi genom att igen byta delar av ClpC mot delar i ett annat protein från *E.coli*, ClpA. ClpA binder till ClpP-proteaset från *E.coli*, och fungerar bara med detta proteas, inte ClpP3/R. Genom att identifiera ett motiv i ClpC och placera detta på ClpA så kunde vi få detta protein (kallat ClpCA) att binda till ClpP3/R och fungera ihop med det.

Vi har också undersökt hur adaptorproteinerna ClpS1 och ClpS2 påverkar Clp-proteaset. ClpS finns i de flesta organismer som har Clp-proteas, men cyanobakterier verkar vara lite speciella på så sätt att de har två stycken ClpS-proteiner. ClpS1, som är det som är mest likt ClpS i andra organismer, och ClpS2, som verkar vara helt unikt just för cyanobakterier. ClpS har också studerats en hel del i *E.coli*, där man har upptäckt att ClpS känner igen en viss sorts ”etiketter” på proteiner som ska brytas ned, och för dessa proteiner till ClpAP-proteaset (*E.colis* närmaste funktionella motsvarigheten till ClpCP3/R) samtidigt som ClpS blockerar ClpAP från att bryta ned andra proteiner. Det

intressanta här är att ClpS är icke-essentiellt i alla organismer där det hittills har undersökts, vilket alltså betyder att det inte måste finnas där för att organismerna ska kunna överleva. Man kan ta bort det från cellen utan att det verkar påverka dem. Detta verkar dock inte gälla för ClpS2. När vi försökte ta bort ClpS2 från cyanobakterierna som vi studerar så överlevde de inte detta. Däremot så gick det att ta bort ClpS1, som då verkar var mer likt det ”vanliga” ClpS. Men samtidigt så pekar de resultat vi har fått på att ClpS1 ändå inte fungerar helt likadant som ”vanligt” ClpS i andra organismer. Vi har gjort studier där vi jämförde ClpS1 och ClpS2 och upptäckte att de verkar känna igen olika ”etiketter” på proteiner som ska brytas ned. Det betyder att ClpS1 inte fungerar som alla andra ClpS, utan att dessa proteiner verkar ha blivit specialicerade på att känna igen olika etiketter. Vi har ännu inte kunnat ta reda på exakt vilka olika proteiner som ClpS1 och ClpS2 känner igen i cyanobakterier, men resultaten visar i alla fall att ClpS-proteinerna i cyanobakterier verkar fungera på ett speciellt sätt jämfört med andra organismer. Sammantaget så har resultaten från studierna i den här avhandlingen lett till ny kunskap om hur det essentiella Clp-proteaset i fotosyntetiserande organismer fungerar, och också hur adaptorproteinerna som interagerar med dem fungerar.

Acknowledgements

Well, this has been a very long journey indeed. And there are may people that have helped me along the way.

First I want to thank my supervisor **Adrian Clarke.** Thanks for taking me on as a PhD-student. Whenever there has been a problem with my projects you have managed to help me find a solution. Which often ment purifying a new version of a protein... :-) Thanks for good discussions, and of course thank you for your patience during these years. And I definitively owe you a box of beer now after these last stressful months! Thanks!

Tanks to my co-supervisor **Henrik Aronsson** for support and encouragement.

Frida Ståhlberg: First: sorry for having forced you to live through the chaos of my part of our office for the last years. And thanks for good collaborations, a **lot** of help and many interesting and funny discussions over the years.

Nori Tabakani: thanks for all the help with cloning problems. You really have the magic touch when it comes to cloning.!

The "old Clp-gang".

Fredrik Andersson: For good collaboration, and for introducing me to the protein purification field. Which of course means long, long days in the cold room. But also nice results when you finally have the proteins. And of course it was worse in your days when we didn't have computers for the Äktas... :-)

Lars Sjögren: for a lot of lab help with Western Blots and many, many fun and nice discussions on both scientific and non-scientific subjects. The music played in the lab now can not meet up to the standard I got used when you took care of that.

Tara Stanne: For good collaborations and for being a nice and friendly person in general. :-)

Elena Pojidaeca: for good collaborations and also for introducing me to the art of running native gels.

Other people at Botan:

Mats Töpel for help with phylogenetic analysis and other nice suggestions for data searches. This was very helpful and gave new perspectives to the questions asked.

Mats X Andersson: thanks for all the help over the years, both with theoretical and practical questions. And of course also for being a very good friend in general. I'm sure going to miss our lunches! Or at least the interesting. funny and quite often bizarre discussions.

Sazzad Karim: thanks for all the encouragement and pep talk over the years. And of course also for scientific help and advice.

And of course a **big** thank you to all nice people that I've met at Botan over the years. **Tobias P** for introducing me to "Curb Your Enthusiasm" and other TV-series. And for generally having the same type of bizarre humour and taste for horror moves as me. The whole "fika" gang that has varied a lot over the years: **Johanna, Jenny. Mikael, Henrik J, Per F, Oskkar** as the "present group". In the past **Marianne, Josefin, Maria, Sara, Patrik B, Ceclia O. And many more.**

Anders N, thanks for the template for the thesis, that really helped a lot and saved a lot of time for me. Thanks for good discussions, both scientific and non-scientific.

Niklas for help with all computer problems.

Thanks **Monica** and **Vivian** for technical help.

Tack till alla hjälpsamma personer på Bioenv som håller igång verksamheten och ställer upp med hjälp om det behövs.

Sven Toresson: tack för all möjlig hjälp genom åren. Huset skulle förmodligen rasa ihop rätt snabbt om inte du fanns på plats (eller I alla fall våra frysar och lab). Tack speciellt för hjälp I samband med tryckningen av avhandlingen.

Tack till alla trevliga människor som jag har haft att göra med på Botan genom åren. Jag kommer verkligen att sakna er alla jättemycket, det har varit trevligt och kul att jobba här. Och jag har säkert glömt att nämna många som borde omnämnas. Men ni får tackas genom denna kollektiva hälsning.

Världen utanför: (det finns ju faktiskt en sådan också)

Tack till tisdags/torsdags/måndagsgruppen: **Patrik, Olof, Jonas, Rikard, Alexander, Johan, Jonatan** och **Kristoffer**. Det har varit oerhört bra att ha haft en möjlighet att (bokstavligen) fly verkligheten i alla fall en gång i veckan. Jag har dessutom lärt mig att uppskatta te mycket mer än vad jag annars skulle ha gjort, måste jag säga. Ni är också alldeles fantastiska vänner även utanför rollspelens och spelens värld!

Tack också till alla promenadkompisar genom åren som har sett till att jag också har fått lite motion då och då och inte fastat på lab... Speciellt **Jonas** och **Jonatan** ska tackas för detta. Jag hoppas ha lite mer tid att bli social nu och kunna promenera lite mer. Och eftersom Jonatan nu även ska tackas för att ha introducerat mig i ölbryggningens spännande värld så lär det nog behövas mer motion i framtiden också.

Tack till alla andra andra vänner. Tack till **Madde** för att du släpande med mig på quiz första gången när det behövdes utfyllnad av laget, det ledde ju till att jag fick många nya trevliga vänner och bekanta där. Tack Peter S, Cecilia, Anna, Anki, Liselotte och många, många fler för att ni är så trevliga och bra vänner!

Tack också till moster **Karin**, du trodde hela tiden på mig. Det hade varit roligt om du hade fått se mig bli klar, men så blev det ju tyvärr inte.

Sist, men egentligen främst, ett stort tack till mina föräldrar **Sten och Eva.** Ni har hela tiden trott på mig och stöttat och uppmuntrat mig. Tack!!!