FEBS Letters 581 (2007) 3749–3757

Minireview

ClpP: A distinctive family of cylindrical energy-dependent serine proteases

Angela Yeou Hsiung Yu, Walid A. Houry*

Department of Biochemistry, University of Toronto, Medical Sciences Building, 1 King's College Circle, Toronto, ON, Canada M5S 1A8

Received 20 March 2007; revised 16 April 2007; accepted 21 April 2007

Available online 8 May 2007

Edited by Robert Barouki

Abstract Processes maintaining protein homeostasis in the cell are governed by the activities of molecular chaperones that mainly assist in the folding of polypeptide chains and by a large class of proteases that regulate protein levels through degradation. ClpP proteases define a distinctive family of cylindrical, energy-dependent serine proteases that are highly conserved throughout bacteria and eukaryota. They typically interact with ATP-dependent AAA+ chaperones that bind and unfold target substrates and then translocate them into ClpP for degradation. Structural and functional studies have provided a detailed view of the mechanism of function of this class of proteases.
© 2007 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: ClpP; Cylindrical protease; Serine protease; ClpR; AAA+ chaperone

1. Overview

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

ClpP is a highly conserved serine protease present throughout bacteria and eukaryota; it seems to be absent in archaea [1], mollicutes [1], and some fungi (Table 1). ClpP was first identified in *Escherichia coli* [2,3], and more than 400 studies have been published on it since its discovery. The X-ray structure of ClpP has been solved from several different organisms (Figs. 1 and 2). All the structures show similar features: the protease is comprised of 14 subunits arranged into two heptameric rings forming a cylindrical-like structure which encloses a large chamber containing the protease active sites. Entrance into the chamber occurs through axial pores in the cylinder. ClpP typically forms complexes with AAA+ (ATPases associated with various cellular activities) chaperones that denature substrates and then translocate them through the axial pores into the proteolytic chamber of the protease for degradation. ClpP degrades proteins into peptides of about 7–8 residues [4], which are subsequently released from the chamber.

In this review, we will describe the distribution of Clp proteases in different kingdoms of life and then discuss in more detail ClpP structure and mechanism of function.

2. The distribution of ClpP proteases in different kingdoms of life

2.1. Bacterial ClpPs

ClpP is found in all bacteria sequenced to date except for *Mollicutes* [5]. Table 1 lists the number of ClpP copies present in selected organisms. The majority of bacteria contain only one copy of ClpP, although there are exceptions such as for actinobacteria, chlamydiae, cyanobacteria, and others (Table 1) which have multiple isoforms of the protease. In addition to ClpP, cyanobacteria contain a ClpP paralog that does not have all three residues of the Ser-His-Asp catalytic triad. This inactive version of ClpP is called ClpR and is mostly present in plants [6] in addition to cyanobacteria. For example, cyanobacteria *Synechococcus elongatus* and *Synechocystis PCC6803* contain three copies of ClpP and one copy of ClpR (Table 1 and Fig. 3) [7,8].

ClpP was first identified and mostly studied in *Escherichia coli*, which contains only one copy of the protease (Table 1). *E. coli* ClpP consists of 207 amino acids including an N-terminal prosequence which acts as a regulatory peptide and undergoes autocatalytic cleavage during folding to yield a mature ClpP of 193 residues [9]. In *E. coli*, ClpP forms complexes with AAA+ chaperones, ClpX and ClpA, which belong to the Clp/Hsp100 family. ClpA is an 83 kDa ATPase that contains two AAA+ domains, while ClpX contains only one AAA+ domain (46 kDa), which is homologous to the second AAA+ domain of ClpA. ClpA and ClpX are hexameric chaperones that can stack onto one or both ends of ClpP to form ClpAP or ClpXP holoenzyme complexes. Substrates are recognized, unfolded by ClpX or ClpA, and then threaded into the ClpP proteolytic chamber through the narrow axial pores for degradation (Fig. 4). Only the unfolding and threading by the chaperones require ATP binding and hydrolysis, while proteolysis by ClpP does not require nucleotide hydrolysis. The mechanism by which ClpX or ClpA binds to and unfolds substrates is poorly understood. Other bacteria contain ClpA paralogs, such as ClpC, ClpE, and ClpL [10].

The most general substrates for the *E. coli* ClpXP or ClpAP system are those with an SsrA C-terminal tag [11]. The SsrA tag consists of 11 hydrophobic residues that are added by a tmRNA, also called SsrA or 10Sa RNA, to the C-terminus

*Corresponding author. Fax: +1 416 978 8548.
E-mail address: walid.houry@utoronto.ca (W.A. Houry).

0014-5793/$32.00 © 2007 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2007.04.076

Table 1
The distribution of ClpP and ClpR across different organisms

Superkingdom	Phylum	Class	Species	ClpP	ClpR
Archaea			32 completely sequenced genomes		
Bacteria	Actinobacteria	Actinobacteria	*Mycobacterium tuberculosis H37Rv*	2	
	Bacteroidetes	Bacteroidetes	*Bacteroides fragilis NCTC 9343*	1	
	Chlorobi	Chlorobia	*Chlorobium chlorochromatii CaD3*	1	
	Chlamydiae	Chlamydiae	*Chlamydia muridarum Nigg*	2	
	Chlamydiae	Chlamydiae	*Chlamydophila pneumoniae AR39*	2	
	Cyanobacteria	–	*Prochlorococcus marinus str. MIT 9313*	3	1
	Cyanobacteria	–	*Synechococcus elongatus*	3	1
	Cyanobacteria	–	*Synechocystis PCC6803*	3	1
	Firmicutes	Clostridia	*Carboxydothermus hydrogenoformans Z-2901*	1	
	Firmicutes	Bacilli	*Bacillus subtilis str.168*	1	
	Firmicutes	Bacilli	*Streptococcus pneumoniae R6*	1	
	Firmicutes	Bacilli	*Lactobacillus acidophilus NCFM*	1	
	Firmicutes	Mollicutes	16 completely sequenced genomes		
	Spirochaetes	Spirochaetes	*Borrelia afzelii (strain PKo)*	2	
	Proteobacteria	Alphaproteobacteria	*Agrobacterium tumefaciens str. C58*	3	
	Proteobacteria	Alphaproteobacteria	*Anaplasma phagocytophilum HZ*	1	
	Proteobacteria	Alphaproteobacteria	*Rickettsia felis URRWXCal2*	1	
	Proteobacteria	Betaproteobacteria	*Bordetella pertussis Tohama I*	1	
	Proteobacteria	Betaproteobacteria	*Bordetella parapertussis 12822*	1	
	Proteobacteria	Betaproteobacteria	*Burkholderia cenocepacia AU 1054*	1	
	Proteobacteria	Betaproteobacteria	*Ralstonia eutropha JMP134*	1	
	Proteobacteria	Betaproteobacteria	*Chromobacterium violaceum ATCC 12472*	1	
	Proteobacteria	Betaproteobacteria	*Neisseria gonorrhoeae FA 1090*	1	
	Proteobacteria	Betaproteobacteria	*Neisseria meningitidis FAM18*	1	
	Proteobacteria	Deltaproteobacteria	*Desulfotalea psychrophila LSv54*	1	
	Proteobacteria	Deltaproteobacteria	*Geobacter metallireducens GS-15*	1	
	Proteobacteria	Epsilonproteobacteria	*Campylobacter jejuni RM1221*	1	
	Proteobacteria	Epsilonproteobacteria	*Helicobacter pylori J99*	1	
	Proteobacteria	Gammaproteobacteria	*Escherichia coli K12*	1	
	Proteobacteria	Gammaproteobacteria	*Salmonella typhimurium LT2*	1	
	Proteobacteria	Gammaproteobacteria	*Yersinia pestis Antiqua*	1	
	Proteobacteria	Gammaproteobacteria	*Haemophilus ducreyi 35000HP*	1	
	Proteobacteria	Gammaproteobacteria	*Pasteurella multocida subsp. Multocida str. Pm70*	1	
	Proteobacteria	Gammaproteobacteria	*Pseudomonas aeruginosa 2192*	2	
	Proteobacteria	Gammaproteobacteria	*Vibrio cholerae O1 biovar eltor str. N16961*	1	
	Proteobacteria	Gammaproteobacteria	*Xanthomonas axonopodis pv. citri str. 306*	1	
Eukaryota	Apicomplexa	Aconoidasida	*Plasmodium falciparum 3D7*	1	1
	Ascomycota	Saccharomycetes	*Saccharomyces cerevisiae*		
	Ascomycota	Schizosaccharomycetes	*Schizosaccharomyces pombe 972h-*		
	Ascomycota	Sordariomycetes	*Gibberella zeae PH-1 (incomplete sequence)*	1	
	Ascomycota	Sordariomycetes	*Neurospora crassa OR74A (incomplete sequence)*	1	
	Ascomycota	Eurotiomycetes	*Aspergillus fumigatus Af293*	1	
	Chlorophyta	Chlorophyceae	*Chlamydomonas reinhardtii*	4	5
	Streptophyta	–	*Arabidopsis thaliana*	6	4
	Nematoda	Chromadorea	*Caenorhabditis elegans*	1	
	Arthropoda	Insecta	*Drosophila melanogaster*	1	
	Chordata	Mammalia	*Mus musculus*	1	
	Chordata	Mammalia	*Rattus norvegicus*	1	
	Chordata	Mammalia	*Homo sapiens*	1	

E. coli ClpP, *Synechocystis PCC6803* ClpR, and *Arabidopsis* ClpR protein sequences were used to BLAST against all archaeal genomes up to date (NCBI March, 2007 version) and selected organisms from bacteria and eukaryota in NCBI. A cut-off of 40% identity with at least 150 amino acids aligned was typically used. The results for *Chlamydomonas reinhardtii* are based on published literature [29].

of nascent chains whose translation is stalled on the ribosome, targeting them for degradation [12–14].

Without ATPase components, only small peptides can enter the ClpP chamber [15,16]; hence, ClpP on its own cannot degrade folded proteins. Recently, however, a group at Bayer described the discovery of acyldepsipeptides (ADEPs) that, when added to ClpP, allow the protease to degrade folded native proteins in the absence of its cognate chaperones [17]. The drug was initially found to rescue mice challenged with lethal infections of *Enterococcus faecalis* and *Staphylococcus aureus*, and the target for the drug was then identified to be ClpP. ClpP is activated by ADEPs to degrade, or at least to cleave, folded proteins in the absence of the cognate Clp ATPases. The consequences of the action of the drug *in vivo*, is that the ADEPs cause unregulated proteolysis by ClpP, subsequently, triggering cell death in Gram-positive bacteria. Gram-negative bacteria, on the other hand, were found to be able to survive drug treatment because they have efficient efflux pumps to remove the drug from the cell, but deletion of those pumps or the use of permeabilizing agents made Gram-negative bacteria vulnerable to ADEP treatment as well [17]. The mechanism of how ADEPs activate ClpP in the absence

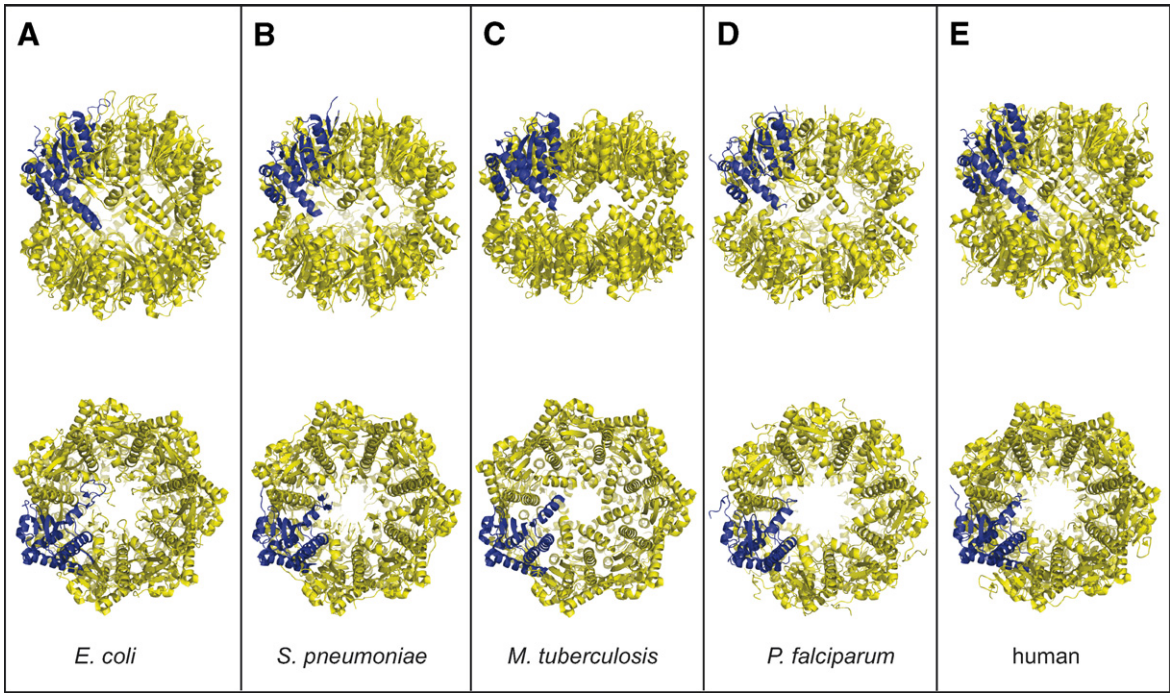

Fig. 1. The structures of tetradecameric ClpP. Shown are the side and top views of the X-ray structures of tetradecameric ClpPs from *E. coli* (1YG6), *S. pneumoniae* (1Y7O), *M. tuberculosis* (2CE3), *P. falciparum* (2F6I), and human (1TG6). The structures were drawn using PyMOL (http://pymol.sourceforge.net/).

of its cognate chaperones is currently unknown. The discovery of ADEPs may lead to development of a novel class of antimicrobial agents.

2.2. Human ClpP

In the human genome, ClpP is encoded on chromosome 19 [18]. Immunofluorescence studies illustrated that human ClpP is located in the mitochondrial matrix [19], although its role is still not known. The mature human ClpP has about 56 residues removed from its N-terminus [19,20], which include the mitochondrial targeting sequence and the prosequence. Mature human ClpP shares high sequence identity (56%) and similarity (71%) with *E. coli* ClpP (Fig. 3), however, human ClpP has an additional 28 residues present at its C-terminus. The function of the extended C-terminus is unknown although it seems to be a unique feature of mammalian ClpPs [21]. The only ATPase component identified so far for human ClpP is ClpX, which is encoded on chromosome 15 [18]. Human ClpX also contains an extended N-terminal sequence which targets ClpX to the mitochondria; the mature human ClpX shares 44% identity and 62% similarity with *E. coli* ClpX.

Unlike *E. coli* ClpP that exists solely as a double-ring tetradecamer, human ClpP exists as a single heptameric ring under physiological conditions [22]. In *E. coli*, the ClpP catalytic triads are compartmentalized inside two heptameric rings (Figs. 1 and 2), but in the human single-ring ClpP, they are exposed to the environment. Indeed, the active sites in human ClpP have been shown to be solvent accessible by trypsin digestion [23]. The accessible catalytic triads would be detrimental for the cell if they are in an active configuration. However, the heptameric human ClpP does not exhibit protease activity and has a very low peptidase activity compared to *E. coli* ClpP [23]. The lack of activity is likely due to an inactive orientation of the catalytic triads in human ClpP heptamer, which might result from the higher mobility in the handle region in the heptameric single ring assembly compared to the tetradecameric double-ring complex. The catalytic triad is located at the interface between the handle and head regions (Fig. 2A) and, hence, high mobility in the handle region is expected to disrupt the proper orientation of the catalytic triad.

The single-ring human ClpP forms a double ring upon binding to ClpX in the presence of ATP, which suggests that binding of the ATPase component might induce conformational changes in the ClpP heptameric rings leading to the formation of the double ring. The human ClpXP complex has protease activity as well as higher peptidase activity compared to heptameric human ClpP, implying that the catalytic residues are properly oriented in an active configuration to catalyze peptide bond hydrolysis [23]. Hence, in the human ClpP heptamer, the accessible catalytic sites are inactive to prevent uncontrolled proteolysis until the ClpP cylinder is properly formed and the active sites are sequestered from the environment upon binding of ClpX and the formation of ClpXP complexes. Such a paradigm of functional regulation has been observed for ClpPs in other organisms such as *Bacillus subtilis* [24].

Surprisingly, it was found that human ClpP can form a complex with *E. coli* ClpX *in vitro* [22]. The structure of the heterogeneous complex as visualized by electron microscopy was highly similar to that of the *E. coli* ClpXP complex. However, the inverse hetero-complex of human ClpX and *E. coli* ClpP could not be formed. Also, the *E. coli* ClpA ATPase has no affinity to human ClpP. It is known that human ClpXP does not recognize the same substrates as the *E. coli* ClpXP system. Interestingly, however, when human ClpP is associated with *E. coli* ClpX, the heterogeneous system can degrade *E. coli* ClpXP substrates. These data further confirm that substrate

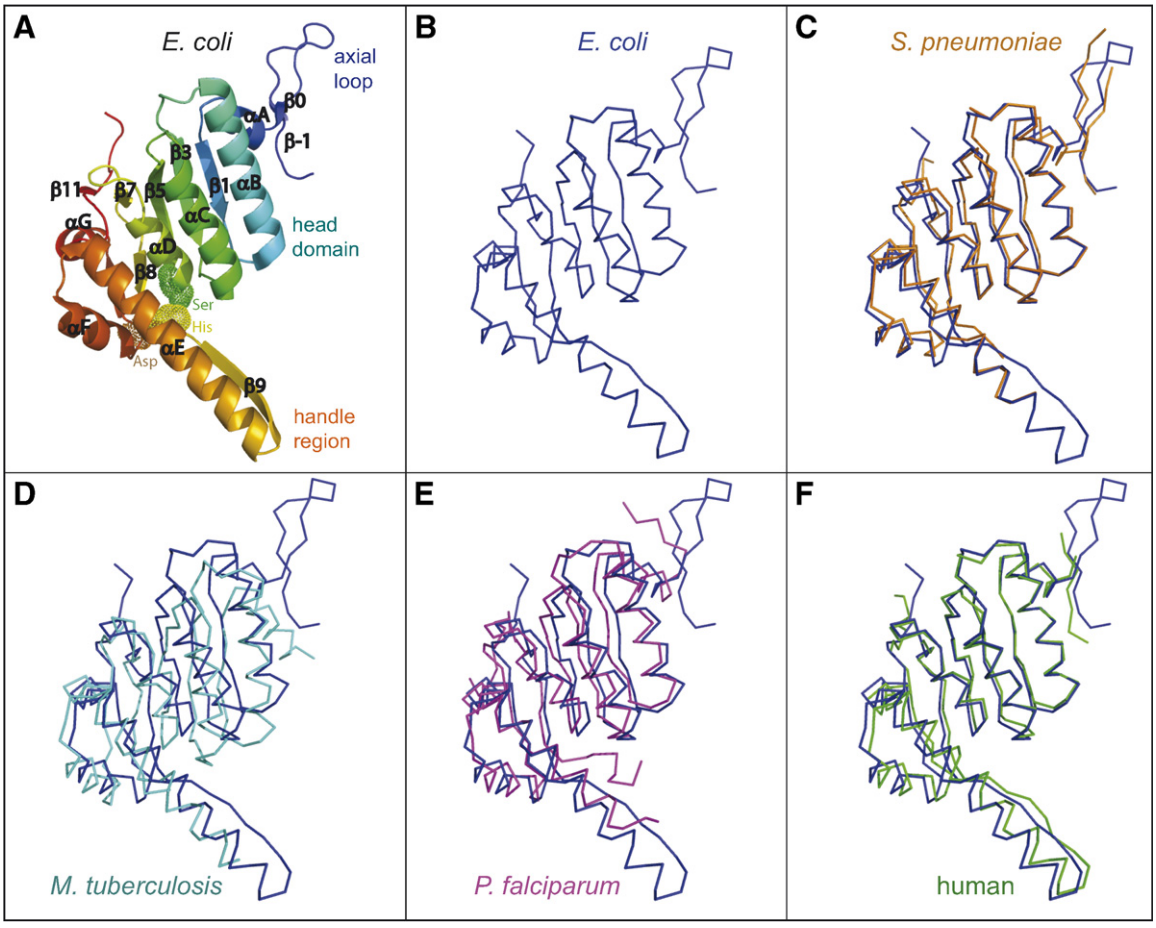

Fig. 2. The structures of the ClpP subunit. (A) Secondary structure elements of *E. coli* ClpP monomer are highlighted, the axial loops, head domain, and handle region are also indicated. Residues of the catalytic triad are shown as dotted spheres. (B)–(F) Overlay of the monomers from *S. pneumoniae*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

recognition is only dependent on the chaperone component of the system.

2.3. Plant ClpPs

The diversity and complexity of the Clp family is evident in higher plant organisms [25]. There are at least 10 ClpP-like proteins identified in the model plant *Arabidopsis thaliana* [26] (Table 1). There are 6 ClpP paralogs (ClpP1–6) and 4 ClpR paralogs (ClpR1–4). Phylogenetic studies indicate that some ClpR proteins (ClpR1, 3, 4) of *Arabidopsis thaliana* may have evolved from the cyanobacterial ClpR [6]. In addition, there are 10 Clp/Hsp100 AAA+ chaperones found in *A. thaliana* including 7 Class I (ClpB1–4, ClpC1–2, ClpD) and 3 Class II (ClpX1-3) which are associated with ClpPR complexes [25]. ClpT is an ortholog of the bacterial ClpS, which is a cofactor that binds to ClpA and affects substrate specificity in *E. coli* [27]. A new group of plant Clp chaperones with unknown function were discovered and also named as ClpS (ClpS1, ClpS2), but they have no similarity with bacterial ClpS and should not be confused with them. Although the function of plant ClpS1-2 is not clear, their sequences are homologous to the N-terminus of the chaperone ClpC but lack the AAA+ domains [6], which led to the suggestion that they may bind to the apical surface of ClpPR core complex, preventing association with the ATPases [25].

Most Clp ATPases and proteases in *Arabidopsis* are found in the chloroplast stroma including ClpB3, ClpC1–2, ClpD, ClpP1, CpP3–6, ClpR1–4, ClpS1–2, and ClpT. All of them are encoded in the nuclear genome, except for ClpP1 which is plastid-encoded [25,26]. The complexity of Clp proteins was also seen in other plants. For example, all Clp proteins except for ClpT were also identified in the plastids of non-green plants *Brassica rapa* roots and *Brassica oleracea* petals by mass spectrometry [6,25].

The ClpP oligomer in the chloroplast of *Arabidopsis* has been identified as composed of the subunits ClpP1, ClpP3–6, ClpR1–4 and ClpS1–2, forming a complex of 325–350 kDa [25]. Modeling studies indicate that ClpP and ClpR contribute to formation of the tetradecamer, whereas ClpS1–2 do not fit in the ring structure but rather seem to be associated with hydrophobic pockets on the axial sites on top of the ClpPR core complex [25]. In the mitochondria, the ClpP complex is a homotetradecamer of ClpP2 subunits [25,26]. From sequence alignments, ClpP2 is found to be the most closely related to human mitochondrial ClpP, and, hence, mitochondrial plant ClpP2 possibly forms a complex with plant ClpX which is also targeted to the plant mitochondria [28].

ClpS1-2 are only found in land plants but not in prokaryotes or green algae, which suggests that they probably have special functional roles in proteolysis in higher plants. Binding of

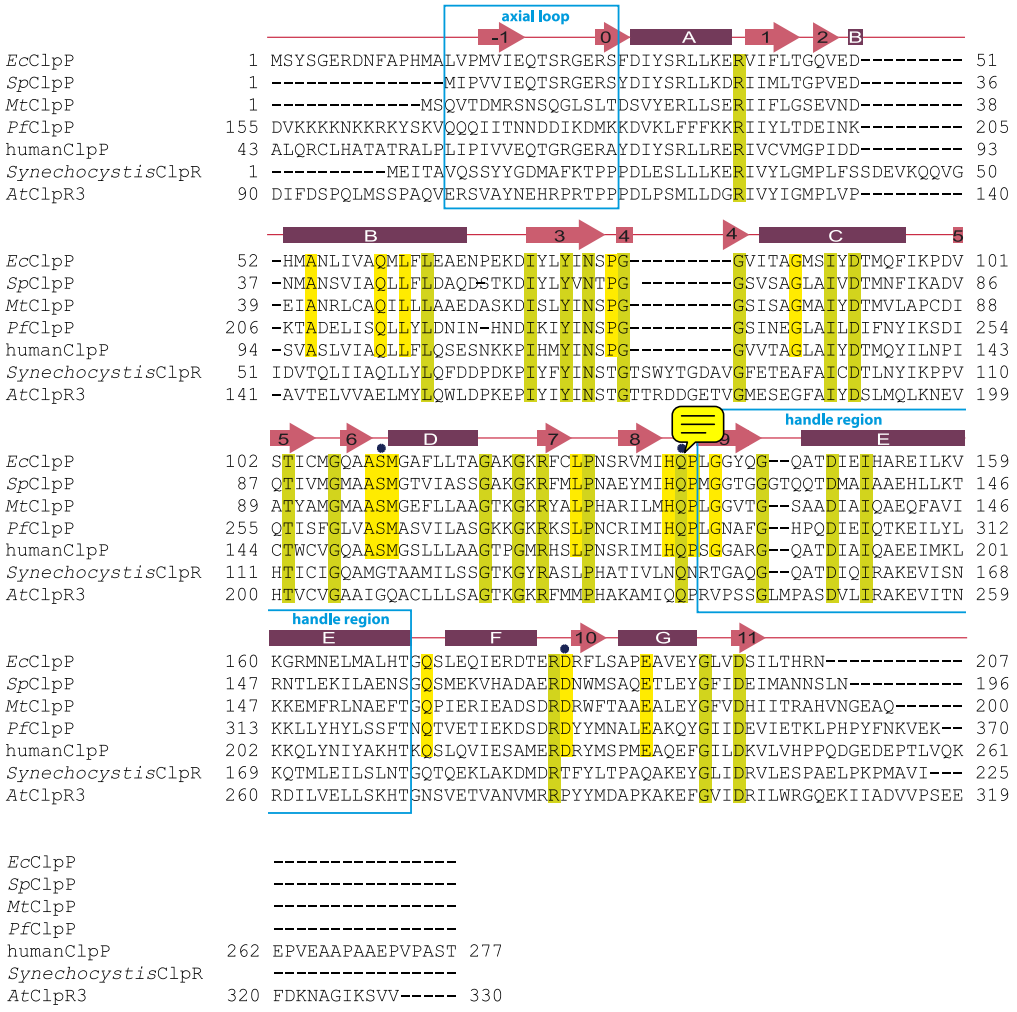

Fig. 3. Alignment of ClpP and ClpR sequences. Sequence alignments of ClpP from *E. coli*, *S. pneumoniae*, *M. tuberculosis*, *P. falciparum*, and human with ClpR from *Synechocystis PCC6803* and ClpR3 from *A. thaliana*. The alignment was generated using ClustalW and the Blosum62 matrix [49]. Identical residues in all seven sequences are highlighted in green, while identical residues in the five ClpP sequences are highlighted in yellow. The catalytic triad residues are indicated by blue dots. The boundaries for the axial loop and handle region and the numbering of the helices and strands refer to that of *E. coli* ClpP [34,37].

ClpS1–2 to the axial site of ClpPR core is likely to inhibit the binding of the cognate ATPase such as ClpC or ClpD. Therefore ClpS1–2 might regulate degradation by competitive binding with the ATPase chaperones [29]. The heterocomplex of ClpPR core is also found in the green alga *Chlamydomonas reinhardtii* stroma (also see Table 1), but ClpS1–2 do not exist in *C. reinhardtii* genome [25]. Instead, a 30-kDa insertion sequence (IS1) is found in ClpP1 of *C. reinhardtii*, which is proposed to have the same effect as ClpS1–2. The insertion domain located between helix 2 and strand 2 (Fig. 2) should protrude over the apical surface thus hindering ATPase binding [29].

The inactive ClpR is thought to have a regulatory role in proteolysis. ClpRs in *A. thaliana* have extended C-termini compared to *E. coli* ClpP (Fig. 3). The C-terminal extension in ClpR might fold on top of the proteolytic core and, hence, might control the interaction between the core and its chaperones. An insertion loop of 9–10 residues in ClpR1, ClpR3, and ClpR4 is found when aligned with *A. thaliana* ClpPs. According to homology modeling, the insertion loop is part of the substrate-binding cleft in the active site, thus this loop of the non-catalytic ClpR proteins might participate in presenting substrates to the catalytic triads of active ClpP neighbors [25].

The cellular functions of most ClpPs in plants are not yet known, but all ClpP paralogs are essential for chloroplast development. In *C. reinhardtii*, ClpP1 is responsible for degrading fully or incompletely assembled cytochrome b_6f and for high CO_2 tolerance [30]. ClpP1 is also vital for plastid development and plant viability in tobacco (*Nicotiana tabacum*) [31,32]. In *Arabidopsis*, ClpR2 is essential for Clp core complex assembly since reducing the levels of ClpR2 resulted in reduction of the ClpPRS protease complex levels, which led to a pale-green phenotype, delayed shoot development, reduced chloroplast size, decreased thylakoid accumulation, and increased plastoglobule levels [33].

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

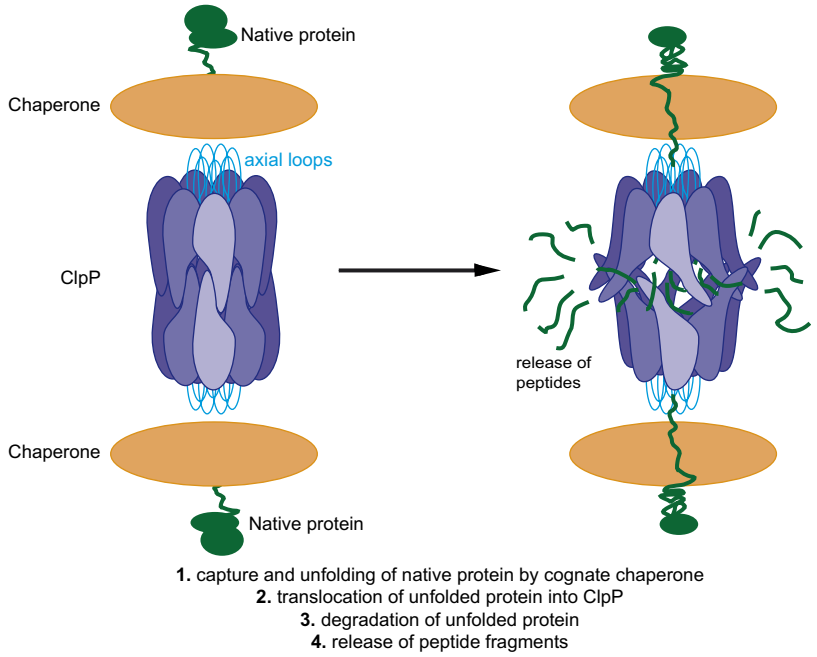

Fig. 4. Model of ClpP mechanism of function. Shown is a cartoon model of the ClpP tetradecamer with the axial loops indicated. For simplicity, the chaperones bound to ClpP are drawn as simple ellipses. In the proposed model, substrate proteins are unfolded and then translocated into the ClpP cylinder by the bound chaperone. Unfolded polypeptides enter the ClpP proteolytic chamber through the axial pores which are lined by the axial loops. The polypeptide chains are then degraded into small peptides. The peptides are proposed to be released through axial pores that are transiently formed as a result of the dynamics in the handle region of the ClpP subunits.

2.4. Yeast ClpPs

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

3. ClpP structure and function

3.1. ClpP X-ray structure

To date, ClpP structures have been solved from five different organisms (Fig. 1) including *E. coli* [34–36], *Streptococcus pneumoniae* [37], the malaria agent *Plasmodium falciparum* [38], *Mycobacterium tuberculosis* [39], and human [21]. The sequence alignment of these different ClpPs is given in Fig. 3. As expected, the overall ClpP structures from these organisms are very similar (Figs. 1 and 2).

The X-ray structure of ClpP from *E. coli* was the first to be solved (1TYF) [34]; it shows a cylindrical-shaped tetradecamer of about 300 kDa in molecular weight and 90 Å in both height and diameter [34]. The monomer (Fig. 2) is mainly composed of six repeats of α/β-fold (α*A*/β1/β2, α*B*/β3/β4, α*C*/β5/β6, α*D*/β7/β8, α*F*/β10, and α*G*/β11) with an additional protruding α/β unit (α*E*/β9). The subunits are held together mainly by hydrophobic interactions. Each monomer can be divided into a handle region (β strand 9 and E helix), which mediates ring-ring interaction, and a head domain comprised of residues 42–134 and 174–202 [34]. A heptameric ring is formed from the packing of head domains, exposing 7 handle regions, which intercalate with seven handle regions from the opposite heptameric ring, forming a tetradecamer with a spherical internal chamber of about 51 Å in diameter [40]. Although according to the X-ray structure, the handle region is the only area where two ClpP rings have contact, surprisingly, truncations in this region do not lead to the dissociation of the 2 rings, indicating a high degree of plasticity of this region [37]. Charge–charge interaction networks involving residues in the head domain seem to contribute to stabilizing the interactions between the two ClpP heptameric rings [37,41]. The two rings are found to dissociate under high sulfate concentration at low temperature [41].

A mutant ClpP(A153P) from *S. pneumoniae* was solved (Figs. 1and 2) and the major difference in the structure of this mutant from the first solved *E. coli* ClpP was in the N-terminal region. In *S. pneumoniae* ClpP(A153P), it was observed that the N-terminal residues Met16-Ser30 (using *E. coli* SwissProt numbering, Fig. 3) form an extended loop protruding from the body of the ClpP cylinder. These axial loops were not detected in the original *E. coli* ClpP structure due to the weak electron density in that region. However, recently solved X-ray structures of *E. coli* ClpP (e.g. 2FZS [36] and 1YG6 [35], shown in Fig. 1) and human ClpP (1TG6) [21] show the presence of these axial loops. These axial loops are an important feature of the ClpP structure and have been shown to mediate the interaction of ClpP with its cognate ATPases [37].

The mutation of Ala153 to Pro in the *S. pneumoniae* ClpP(A153P) structure causes disorder in the handle domain and renders the E helix shorter by two turns as compared to

the E helix in wildtype *E. coli* ClpP (Fig. 2) [37]. The fact that ClpP remains tetradecameric under these conditions is an explicit demonstration of the plasticity of the handle region. This region is also found to be disordered in the crystal structure of *P. falciparum* ClpP and *M. tuberculosis* ClpP1. Based on biochemical and NMR-based studies [37,42], we had proposed that this region of ClpP is highly dynamic and that the movement of the E helices results in the transient formation of equatorial side pores that allow for the exit of peptide fragments generated from the degradation of polypeptide chains inside the ClpP chamber (Fig. 4). This issue is further discussed below.

Very little is known about the *P. falciparum* ClpP. However, the sequence of this protease has about 150 extra amino acids at the N-terminus of the protein when compared to the sequence of the *E. coli* ClpP. It is predicted that this extra N-terminal sequence targets the nuclear encoded protease to the apicoplast organelle present in plasmodium, although this has not been experimentally demonstrated. The apicoplast is an organelle that is homologous to chloroplast of plants and is found in apicomplexan parasites such as *P. falciparum* [43]. The apicoplast is an ancient feature of this group of organisms and is thought to have been acquired by the process of endosymbiosis. It contains proteins that are nuclear encoded and then transported to this organelle and also proteins that are encoded and expressed by its own genome and expression machinery.

The structure of *M. tuberculosis* ClpP1 has two unique features [39]. First, there is an extended αA helix at the N-terminus (Fig. 2) which renders the axial pores smaller compared to those in other ClpPs. Second, the monomers are tilted inwards compared to other ClpPs. The rotated orientation of the monomers results in equatorial side pores, which were suggested to transiently form in ClpP [37,42]. This rotated orientation of the monomers may assist in substrate entry through the axial pores, or product release from the equatorial pores [39]. Like human ClpP, *M. tuberculosis* ClpP1 is also found to form heptamers under normal conditions [39], although in the crystal it formed a tetradecamer. It should be noted, however, that the structure of *M. tuberculosis* ClpP1 might not be the predominant physiological assembly since there are two ClpPs in *M. tuberculosis* (ClpP1 and ClpP2, Table 1) which seem to be on the same operon and might form a hetero-oligomer.

The unique feature in the structure of human ClpP is the presence of an extended C-terminus with 28 additional residues located on the periphery of the heptamer, forming a flexible loop which extends out of the surface of the oligomer. The loop is unstructured and is not observed in the solved X-ray structure of human ClpP (Figs. 1 and 2). This C-terminal extension is found to affect the assembly of human ClpP heptamer since deletion of this C-terminus resulted in structurally unstable ClpP [21]. However, the deletion also resulted in an increased affinity of human ClpP to human ClpX, suggesting that the C-terminus might hinder the interaction between ClpX and ClpP [21].

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

3.2. ClpP mechanism of function

ClpP is a classical serine protease whereby each subunit in the ClpP homotetradecamer has an active site consisting of the three canonical residues: Ser, His, and Asp (Figs. 2A and 3). Hence, the ClpP homotetradecamer contains 14 active sites within its proteolytic chamber. Fig. 2A shows the location of the catalytic triad at the junction between the head and handle domains in each ClpP monomer. For the case of the hetero ClpPR complexes in plants and cyanobacteria, the number of active sites would be reduced. It has been shown that mutations in the handle region are often found to decrease ClpP activity due to the close proximity of the active site to the handle region, resulting in the disruption of the catalytic triad alignment [42]. Systematic studies on ClpP substrate cleaving specificity have not yet been published, however, according to one report [36], *E. coli* ClpP was found to preferentially cleave after charged and branched-chain amino acids. There are indications that the catalytic residues undergo a rearrangement to an active configuration upon substrate binding [36].

The only access to the catalytic chamber is through the narrow axial pores which allow entry of small peptides of about 30 residues in length [15]. The N-terminal axial loops contribute to the axial pores. The axial loops can be divided into two parts [35–37]. There are 7–8 hydrophobic residues, termed the axial pore lining, that line the axial pores followed by about 9 residues, the axial protrusion, that protrude from the apical surface of the ClpP cylinder (Fig. 2). These axial loops are required for the interaction of ClpP with its cognate ATPases [35,37]. Mutants with the first seven N-terminal residues deleted from mature ClpP were shown to be unable to degrade folded protein substrates in the presence of the ATPases, but exhibited higher peptidase activity, suggesting that the truncations did not inactivate the protease but had disrupted its interaction with the chaperones [37]. The increased peptidase activity is probably due to the enlargement of the axial pores resulting from the deletion of the N-terminal residues.

On the apical surface of ClpP, about 54 Å away from the axial pores, there are seven grooves of about 10 Å in diameter. These grooves are mainly composed of conserved hydrophobic residues and are thought to provide binding pockets for specific loop regions present in the structure of the chaperones. These loop regions are highly conserved in AAA+ chaperones that bind to ClpP [44] such as the 'IGF' loop in *E. coli* ClpX or 'IGL' loop in *E. coli* ClpA. Hence, the chaperone–protease interaction is mainly mediated by loop–groove interactions rather by interactions between large surfaces. This mode of interaction might circumvent the symmetry mismatch that exists between ClpP, which has a sevenfold symmetry, and its interacting chaperones, which typically have sixfold symmetry. It is not known how many loop–groove contacts are maintained between ClpX hexamer and ClpP oligomer in an active ClpXP complex. The symmetry mismatch between chaperone rings and ClpP rings have been proposed to be an evolutionary result to allow the rings to reciprocate or rotate about each other, enhancing the rate of translocation of unfolded substrates [45], although no experimental evidence for such rotation is currently available. Alternatively, the loop–groove interaction could provide the chaperone–protease complex with more structural flexibility that might be necessary for the complex to achieve substrate protein unfolding, translocation, protein degradation, and peptide product release.

ClpP processively degrades substrates and generates peptides of average length of about 7–8 residues [4,15]. The mechanisms by which unfolded polypeptides enter ClpP and how the degraded products are released from the protease are still poorly understood. Crystallographic data [34,36] indicate that the substrate protein is probably captured close to the active site by a hydrogen bonding network involving residues in β4 and β9 and that the interaction is further stabilized by hydrophobic interactions with the wall of the ClpP chamber. It is proposed that the substrate will be horizontally positioned in the proteolytic chamber near the equator of the ClpP cylinder and oriented in a clockwise N to C direction when viewed from the axial channel through which the substrate entered [36]. This would be in agreement with a C to N translocation of substrates into the ClpP proteolytic chamber.

Different hypotheses exist about the mechanism of product release. One model proposes that the generated peptides exit the ClpP proteolytic chamber by passive diffusion through the same axial pores used for substrate entry [34,46,47]. However, this implies that the bound chaperones have to dissociate from the ClpP cylinder to allow for peptide exit. Such a mechanism might be inefficient and would probably render the degradation of polypeptide chains nonprocessive. This mechanism has been suggested for the proteasome in which the dissociation of the ATPase units has been reported to coincide with product release [48]. The second model suggests that peptides are released through side pores that are transiently formed at the interface between the two heptameric ClpP rings. It has been established that the handle region exhibits high plasticity [37], and NMR studies further confirmed that the E helices exchange between at least two structurally distinct conformations, suggesting the existence of transiently formed dynamic side pores which could act as exit pores for degraded products [42]. Biochemical data also confirmed this model for peptide exit. An alanine residue at position 153 on the E helix of *E. coli* ClpP was mutated to cysteine, A153C, allowing the formation of a disulfide bond between E helices of two subunits from the two opposite heptameric ClpP rings. The crosslinking of the two rings inhibited movements of the handle regions. A small dipeptide was able to diffuse into the chamber of disulfide-linked inactive ClpP but was trapped inside the chamber under oxidizing condition. The trapped substrate was shown to be released after the system was placed in reducing conditions. The results strongly suggested that the equatorial regions of the ClpP barrel provide the exit sites of degraded products [42]. The observation that *M. tuberculosis* ClpP1 and human ClpP exist mainly as single heptameric rings [23,39] and that human ClpP forms the double ring structure only upon binding the chaperone ClpX [23], seems to further support the peptide exit model through transiently formed equatorial side pores.

4. Concluding remarks

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

Acknowledgements: The authors thank Guillaume Thibault and Dr. Anna Gribun for careful reading of the manuscript. We also thank Jennifer K.W. Huen for help with Fig. 4. W.A.H. is Canadian Institutes of Health Research New Investigator. This work is supported by a grant from the Canadian Institutes of Health Research.

References

[1] Wong, P. and Houry, W.A. (2004) Chaperone networks in bacteria: analysis of protein homeostasis in minimal cells. J. Struct. Biol. 146, 79–89.
[2] Katayama-Fujimura, Y., Gottesman, S. and Maurizi, M.R. (1987) A multiple-component, ATP-dependent protease from Escherichia coli. J. Biol. Chem. 262, 4477–4485.
[3] Katayama, Y., Gottesman, S., Pumphrey, J., Rudikoff, S., Clark, W.P. and Maurizi, M.R. (1988) The two-component, ATP-dependent Clp protease of *Escherichia coli*. Purification, cloning, and mutational analysis of the ATP-binding component. J. Biol. Chem. 263, 15226–15236.
[4] Choi, K.H. and Licht, S. (2005) Control of peptide product sizes by the energy-dependent protease ClpAP. Biochemistry 44, 13921–13931.
[5] Wojtyra, U.A., Thibault, G., Tuite, A. and 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, 48981–48990.
[6] Peltier, J.B., Ytterberg, J., Liberles, D.A., Roepstorff, P. and van Wijk, K.J. (2001) Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of Arabidopsis thaliana. J. Biol. Chem. 276, 16318–16327.
[7] Schelin, J., Lindmark, F. and Clarke, A.K. (2002) The clpP multigene family for the ATP-dependent Clp protease in the cyanobacterium *Synechococcus*. Microbiology 148, 2255–2265.
[8] Stanne, T.M., Pojidaeva, E., Andersson, F.I. and Clarke, A.K. (2007) Distinctive types of ATP-dependent Clp proteases in cyanobacteria. J. Biol. Chem. 282, 14394–14402.
[9] Maurizi, M.R., Clark, W.P., Kim, S.H. and Gottesman, S. (1990) Clp P represents a unique family of serine proteases. J. Biol. Chem. 265, 12546–12552.
[10] Butler, S.M., Festa, R.A., Pearce, M.J. and Darwin, K.H. (2006) Self-compartmentalized bacterial proteases and pathogenesis. Mol. Microbiol. 60, 553–562.
[11] Keiler, K.C., Waller, P.R. and Sauer, R.T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science 271, 990–993.
[12] Gottesman, S., Roche, E., Zhou, Y. and 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, 1338–1347.
[13] Weber-Ban, E.U., Reid, B.G., Miranker, A.D. and Horwich, A.L. (1999) Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. Nature 401, 90–93.
[14] Gillet, R. and Felden, B. (2001) Emerging views on tmRNA-mediated protein tagging and ribosome rescue. Mol. Microbiol. 42, 879–885.
[15] Thompson, M.W., Singh, S.K. and Maurizi, M.R. (1994) Processive degradation of proteins by the ATP-dependent Clp protease from Escherichia coli. Requirement for the multiple

array of active sites in ClpP but not ATP hydrolysis. J. Biol. Chem. 269, 18209–18215.
[16] Gottesman, S., Maurizi, M.R. and Wickner, S. (1997) Regulatory subunits of energy-dependent proteases. Cell 91, 435–438.
[17] Brotz-Oesterhelt, H. et al. (2005) Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. Nat. Med. 11, 1082–1087.
[18] Corydon, T.J. et al. (2000) Human and mouse mitochondrial orthologs of bacterial ClpX. Mamm. Genome 11, 899–905.
[19] de Sagarra, M.R., Mayo, I., Marco, S., Rodriguez-Vilarino, S., Oliva, J., Carrascosa, J.L. and Casta n, J.G. (1999) Mitochondrial localization and oligomeric structure of HClpP, the human homologue of *E. coli* ClpP. J. Mol. Biol. 292, 819–825.
[20] von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Domain structure of mitochondrial and chloroplast targeting peptides. Eur. J. Biochem. 180, 535–545.
[21] Kang, S.G., Maurizi, M.R., Thompson, M., Mueser, T. and Ahvazi, B. (2004) Crystallography and mutagenesis point to an essential role for the N-terminus of human mitochondrial ClpP. J. Struct. Biol. 148, 338–352.
[22] Kang, S.G., Ortega, J., Singh, S.K., Wang, N., Huang, N.N., Steven, A.C. and Maurizi, M.R. (2002) Functional proteolytic complexes of the human mitochondrial ATP-dependent protease, hClpXP. J. Biol. Chem. 277, 21095–21102.
[23] Kang, S.G., Dimitrova, M.N., Ortega, J., Ginsburg, A. and Maurizi, M.R. (2005) Human mitochondrial ClpP is a stable heptamer that assembles into a tetradecamer in the presence of ClpX. J. Biol. Chem. 280, 35424–35432.
[24] Kirstein, J., Schlothauer, T., Dougan, D.A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B. and Turgay, K. (2006) Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. EMBO J. 25, 1481–1491.
[25] Peltier, J.B. et al. (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, 4768–4781.
[26] Sjogren, L.L., Stanne, T.M., Zheng, B., Sutinen, S. and Clarke, A.K. (2006) Structural and functional insights into the chloroplast ATP-dependent Clp protease in Arabidopsis. Plant Cell 18, 2635–2649.
[27] Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D.A. and Bukau, B. (2006) ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. Nature 439, 753–756.
[28] Halperin, T., Zheng, B., Itzhaki, H., Clarke, A.K. and Adam, Z. (2001) Plant mitochondria contain proteolytic and regulatory subunits of the ATP-dependent Clp protease. Plant Mol. Biol. 45, 461–468.
[29] Majeran, W., Friso, G., van Wijk, K.J. and Vallon, O. (2005) The chloroplast ClpP complex in *Chlamydomonas reinhardtii* contains an unusual high molecular mass subunit with a large apical domain. FEBS J. 272, 5558–5571.
[30] Majeran, W., Wollman, F.A. and Vallon, O. (2000) Evidence for a role of ClpP in the degradation of the chloroplast cytochrome *b*(6)f complex. Plant Cell 12, 137–150.
[31] Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T. and Hashimoto, T. (2001) The chloroplast clpP gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in tobacco. Plant Cell Physiol. 42, 264–273.
[32] Kuroda, H. and Maliga, P. (2002) Overexpression of the clpP 5′-untranslated region in a chimeric context causes a mutant phenotype, suggesting competition for a clpP-specific RNA maturation factor in tobacco chloroplasts. Plant Physiol. 129, 1600–1606.
[33] Rudella, A., Friso, G., Alonso, J.M., Ecker, J.R. and van Wijk, K.J. (2006) Downregulation of ClpR2 leads to reduced accumulation of the ClpPRS protease complex and defects in chloroplast biogenesis in *Arabidopsis*. Plant Cell 18, 1704–1721.
[34] Wang, J., Hartling, J.A. and Flanagan, J.M. (1997) The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. Cell 91, 447–456.
[35] Bewley, M.C., Graziano, V., Griffin, K. and 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.
[36] Szyk, A. and Maurizi, M.R. (2006) Crystal structure at 1.9A of *E. coli* ClpP with a peptide covalently bound at the active site. J. Struct. Biol. 156, 165–174.
[37] Gribun, A., Kimber, M.S., Ching, R., Sprangers, R., Fiebig, K.M. and 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, 16185–16196.
[38] Vedadi, M. et al. (2007) Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms. Mol. Biochem. Parasitol. 151, 100–110.
[39] Ingvarsson, H., Mate, M.J., Hogbom, M., Portnoi, D., Benaroudj, N., Alzari, P.M., Ortiz-Lombardia, M. and 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, 249–259.
[40] Wang, J., Hartling, J.A. and Flanagan, J.M. (1998) Crystal structure determination of *Escherichia coli* ClpP starting from an EM-derived mask. J. Struct. Biol. 124, 151–163.
[41] Maurizi, M.R., Singh, S.K., Thompson, M.W., Kessel, M. and Ginsburg, A. (1998) Molecular properties of ClpAP protease of *Escherichia coli*: ATP-dependent association of ClpA and clpP. Biochemistry 37, 7778–7786.
[42] Sprangers, R., Gribun, A., Hwang, P.M., Houry, W.A. and Kay, L.E. (2005) Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. Proc. Natl. Acad. Sci. USA 102, 16678–16683.
[43] Waller, R.F. and McFadden, G.I. (2005) The apicoplast: a review of the derived plastid of apicomplexan parasites. Curr. Issues Mol. Biol. 7, 57–79.
[44] Kim, D.Y. and Kim, K.K. (2003) Crystal structure of ClpX molecular chaperone from Helicobacter pylori. J. Biol. Chem. 278, 50664–50670.
[45] Beuron, F., Maurizi, M.R., Belnap, D.M., Kocsis, E., Booy, F.P., Kessel, M. and Steven, A.C. (1998) At sixes and sevens: characterization of the symmetry mismatch of the ClpAP chaperone-assisted protease. J. Struct. Biol. 123, 248–259.
[46] Thompson, M.W. and Maurizi, M.R. (1994) Activity and specificity of *Escherichia coli* ClpAP protease in cleaving model peptide substrates. J. Biol. Chem. 269, 18201–18208.
[47] Kim, Y.I., Burton, R.E., Burton, B.M., Sauer, R.T. and Baker, T.A. (2000) Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. Mol. Cell 5, 639–648.
[48] Babbitt, S.E. et al. (2005) ATP hydrolysis-dependent disassembly of the 26S proteasome is part of the catalytic cycle. Cell 121, 553–565.
[49] Henikoff, S. and Henikoff, J.G. (1993) Performance evaluation of amino acid substitution matrices. Proteins 17, 49–61.