

Comparative genomics and functional roles of the ATP-dependent proteases Lon and Clp during cytosolic protein degradation

Dilip Chandu, Dipankar Nandi *

Abstract

The general pathway involving adenosine triphosphate (ATP)-dependent proteases and ATP-independent peptidases during cytosolic protein degradation is conserved, with differences in the enzymes utilized, in organisms from different kingdoms. Lon and caseinolytic protease (Clp) are key enzymes responsible for the ATP-dependent degradation of cytosolic proteins in *Escherichia coli*. Orthologs of *E. coli* Lon and Clp were searched for, followed by multiple sequence alignment of active site residues, in genomes from seventeen organisms, including representatives from eubacteria, archaea, and eukaryotes. Lon orthologs, unlike ClpP and ClpQ, are present in most organisms studied. The roles of these proteases as essential enzymes and in the virulence of some organisms are discussed.

Keywords: ATP-dependent proteases; Clp; Comparative and functional genomics; Cytosolic protein degradation; Lon

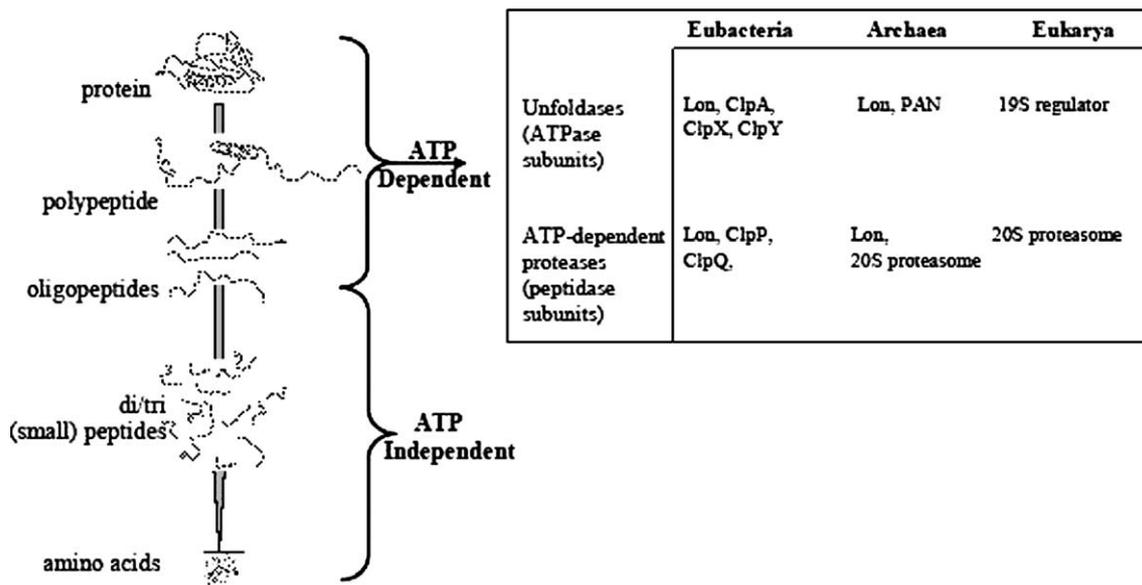
1. Introduction

Cytosolic protein degradation is an active area of investigation and its importance in homeostasis and regulation of important cellular pathways is increasingly appreciated. In general, proteins with well folded structures are stable. Under certain conditions, e.g., unstable proteins, or as proteins age or become damaged, hydrophobic residues that are normally buried are exposed on the surface. Chaperones bind to these hydrophobic residues and attempt to refold these proteins. If refolding is not possible, these misfolded proteins may bind to regulatory components of proteases and undergo degradation. Consequently, cells lacking chaperones or proteases accumulate large amounts of aggregated proteins, demonstrating a close relationship between protein folding and degradation (reviewed in [7,17]).

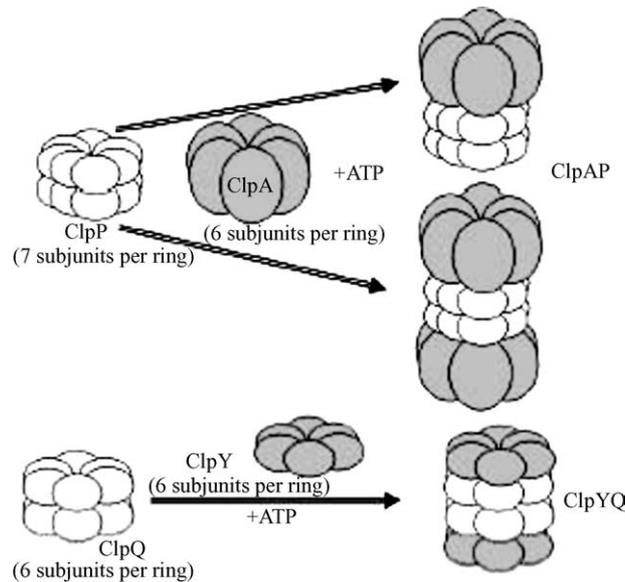
The cytosolic pathway leading to release of amino acids from an intact protein can be divided into two parts based on the utilization of metabolic energy. The proximal ATP-dependent steps are followed by ATP-independent events (Fig. 1a). This basic model of protein unfolding followed by degradation is conserved in eubacteria, archaea and eu-

karyotes and involves several steps. First, ATP-dependent enzymes are responsible for the recognition, unfolding and cleavage of substrate proteins into large peptides (Fig. 1a). These enzymes harbor a domain conferring ATPase activity, belonging to the ATPase family associated with various cellular activities (AAA) in the same polypeptide (e.g., Lon) or associated with a different polypeptide-containing ATPase activity (e.g., ClpA, ClpX, ClpY, 19S regulator) (reviewed in [7,17]). Interestingly, the unavailability of tRNAs during polypeptide biosynthesis results in the addition of an eleven-amino-acid peptide tag (AANDENYALAA) encoded by transfer and messenger RNA (tmRNA) to the C-terminus of the truncated polypeptide, and proteins containing tmRNA encoded peptide tags are degraded rapidly (reviewed in [12]). Substrate proteins are unfolded by regulators via conformational changes driven by ATP hydrolysis. Next, unfolded substrate proteins are actively translocated into the proteolytic chambers present in the proteases, which make the initial cuts to generate large peptides. These are cleaved into smaller peptides and, finally, into amino acids, which are recycled into the cellular pool. The ATP-independent steps of the cytosolic protein degradation pathway are often referred to as “downstream processing of peptides” [5].

26S proteasomes (19S regulator + 20S proteasomes) are responsible for most cytosolic protein degradation in eukary-



(a)



(b)

Fig. 1. General scheme and enzymes involved in cytosolic protein degradation. (a) The intracellular protein degradation pathway leading to the release of free amino acids from intact proteins is represented in this schematic diagram. The initial steps are dependent on ATP-dependent proteases whereas the latter steps are performed by ATP-independent peptidases. The regulatory and protease subunits of ATP-dependent enzymes from different kingdoms are tabulated in the inset. (b) The protease and regulatory components of ClpAP (top) and ClpYQ (bottom) associate and are stable in the presence of ATP to form active complexes that unfold and degrade proteins. Lon harbors both the peptidase and regulatory domains in the same polypeptide and is not shown.

otes. However, Lon and ClpP are the key enzymes responsible for degradation of cytosolic proteins in *Escherichia coli*. To detect orthologs of these proteases, genome analysis of completed sequences of representative organisms (TBLASTN searches) using full length *E. coli* protease sequences was performed (Table 1). The results of this analysis are represented in Table 1. Furthermore, amino acid sequences important for catalysis in Lon and Clp proteases are discussed. We utilized two criteria to select a particular gene as an ortholog: (i) the E-value depicting the chance factor

to be as low as possible (Table 1, parentheses); and (ii) selected orthologs possessing conserved active site residues. Orthologs with a significantly low E-value but lacking active site residues were observed for several enzymes but were not considered for multiple sequence alignment performed using ClustalW. It is possible that some proteins with a high E-value may be important; however, this would require detailed modeling and biochemical studies which do not fall into the scope of this study. In addition, we have reviewed the literature on Lon, Clp proteases and their orthologs with

Table 1

List of *E. coli* peptidases selected for whole genome analysis of representative organisms from various branches of the evolutionary tree

	Ec	Hs	Mm	At	Ce	Dm	Sc	Ta	Ss	Bs	Ll	Mt	Ml	Mge	Mpe	Mpn	Mpu
Lon	sms (8e-04)	pLon (e-139)	Lon (e-140)	nLonm (e-144)	mLon (e-136)	LD30525 (e-148)	PIM1 (e-133)	Lon (6e-20)	LonA (0.0)					Lon (e-157)	Lon (e-169)	Lon (e-161)	Lon (e-147)
	b0955 (0.005)	Lonlike (e-138)	mLon (e-137)	Lon1 (e-137)	Lon (e-115)	CD8798 (e-127)											
		mLon (e-138)	mLon1 (2e-95)	mLon (e-136)													
		mLon1 (e-136)		Lon (e-120)													
		Lon- plantlike (e-134)		Lonlike (6e-89)													
ClpP		ClpP (1e-49)	ClpP (4e-66)	ClpP1 (9e-61)	ClpP (3e-56)	CG5045 (4e-61)				ClpP (5e-74)	ClpP (2e-60)	ClpP (1e-49)	ClpP (3e-49)				
				nClpP1 (4e-58)								ClpP2 (3e-49)	ClpP2 (4e-49)				
				nClpP3 (4e-50)													
				nClpP4 (3c-58)													
ClpQ										HsIV (1e-40)							

Numbers in parentheses represent the E-value, reflecting the significance of the similarity between the homologue and the respective *E. coli* peptidase.

The abbreviations used are as follows: Ec, *Escherichia coli*; Hs, *Homo sapiens*; Mm, *Mus musculus*; At, *Arabidopsis thaliana*; Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Sc, *Saccharomyces cerevisiae*; Ta, *Thermoplasma acidophilum*; Ss, *Sulfolobus solfataricus*; Bs, *Bacillus subtilis*; Ll, *Lactococcus lactis*; Mt, *Mycobacterium tuberculosis*; Ml, *Mycobacterium leprae*; Mge, *Mycoplasma genitalium*; Mpe, *Mycoplasma penetrans*; Mpn, *Mycoplasma pneumoniae*; Mpu, *Mycoplasma pulmonis*.

respect to their structures and roles of these proteases in growth and virulence. Proteases involved in cytosolic protein degradation in bacteria recognize different substrates with overlapping specificities [48]. In fact, *E. coli* lacking three proteases, *lon*, *clpP* and *clpQ*, but not a single protease, is required to prevent the degradation of homoserine *trans*-succinylase, the first enzyme involved in biosynthesis of methionine [1]. A catalogue of some key in vitro and in vivo substrates hydrolyzed by Lon, ClpP and ClpQ is presented in Table 2.

2. Lon

Lon protease is a multimer with each subunit (87 kDa) having three distinct domains: an N-terminal domain with an unassigned function, a central domain with ATPase activity and the C-terminal proteolytic domain harboring the active site serine (Ser-679). In eukaryotes, Lon paralogs are present in the mitochondrial matrix, peroxisomes or plastids. Cryo-electron microscopy has demonstrated that yeast mitochondrial Lon exists as a ring-shaped heptamer [38]. Lon and its orthologs display a conserved ATP binding motif "GPPGVGKT" in the central domain and ATP binding presumably causes a conformational change. In fact, in the absence of ATP, yeast mitochondrial Lon consists of asymmetric rings whereas ATP binding results in formation of symmetric rings [38].

Comparison of Lon amino acid sequences from different sources revealed the presence of a Ser–Lys catalytic dyad. Furthermore, site-specific mutagenesis of the Ser–Lys catalytic dyad residues demonstrated that these were essential for catalysis in VP4, a Lon ortholog encoded by an infectious bursal disease virus strain P2 [2]. Recently, the crystal structure of the proteolytic domain of *E. coli* Lon revealed the presence of a unique fold and modeling of the active site demonstrated the close association of Lys-722 with Ser-679 [4]. Interestingly, the Lon Ser-679 mutant, but not an ATPase domain mutant, is able to complement UV sensitivity and capsule overproduction in a strain lacking Lon activity, probably due to substrate sequestration [44]. This apparent complementation suggests that substrate binding followed by unfolding and proteolysis are coupled events.

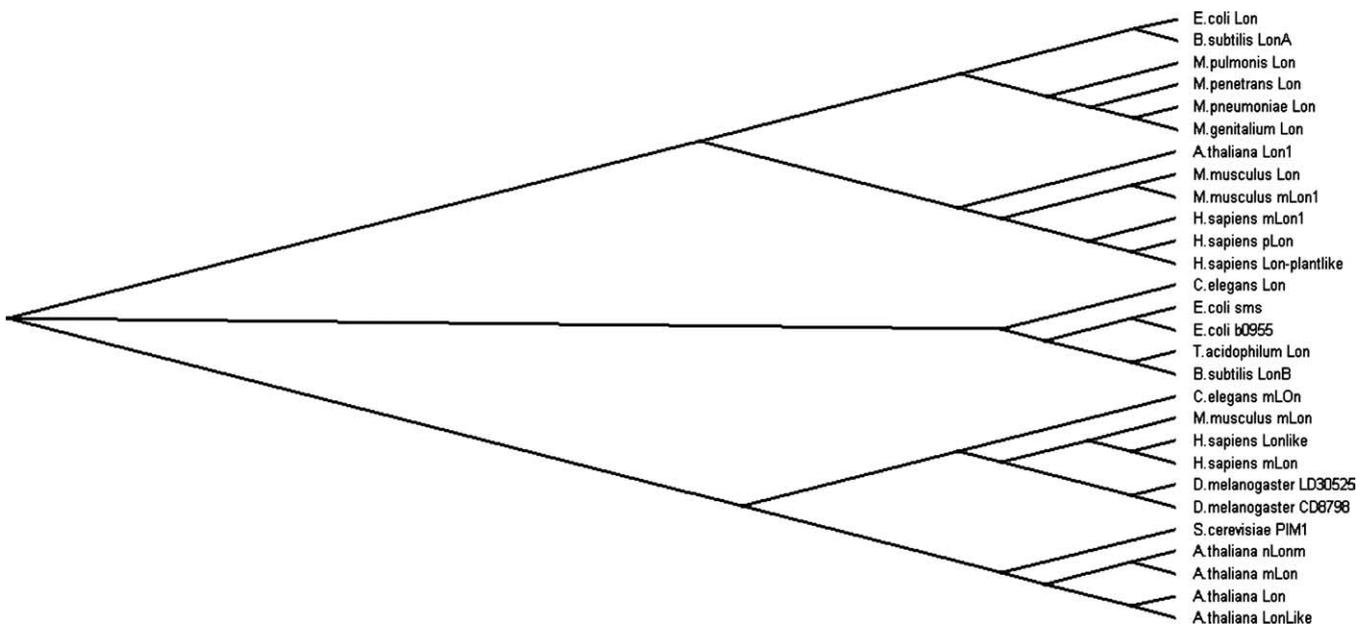
Lon orthologs are widely conserved in different kingdoms (Table 1). Residues corresponding to the catalytic dyad, Ser-679 and Lys-722, in *E. coli* Lon, are present in most orthologs (Fig. 2a). Archaeobacterial representatives showed mixed distribution. *Sulfolobus*, a crenarchaeote, does not harbor a significant Lon ortholog, unlike the euryarchaeote member *Thermoplasma* (Table 1, Fig. 2). As shown in Fig. 2b, the *T. acidophilum* Lon is not closely related to *E. coli* Lon and further biochemical analysis is required. Lon orthologs are present in *Methanococcus jannaschii*, *Pyrococcus furiosus*, *Methanosarcina mazei* and *Halobacterium* sp. NRC-1 (data not shown), suggesting that Lon is fairly well distributed in archaea. In fact, Lon from the

```

E.coli_Lon      PDFYEKRDIHVHVEGATPKDGPSAGIAMCTALVSLTGNPVRADVAMTGEITLRGQVLPVGGIKELL-----AAHRGGI 731
B.subtilis_LonA PDFHEKYDIHIHVEGAVPKDGPSAGITMATALVSLTGRAVSRVEMTGEITLRGRVLPVGGIKEKAL-----GAHRAGL 729
M.pneumoniae_Lon PNLFKKVDINIHVPGGGIPKDGSAGAAALVTAIISSLTGKKVDPKIAMTGEITLRGKVMITIGGVKEKTI-----SAYRGGV 754
M.genitalium_Lon PSLFKKIDINIHVPGGGIPKDGSAGAAALVTAIISSLTGKKVDPVAMTGEITLRGKVLVIGGVKEKTI-----SAYRGGV 754
M.penetrans_Lon PKIFGEIDLHVHVPSSGIPKDGSAGIALTTAIISSLTGKKVDPKIAMTGEITLRGRVLIIGGVKEKTI-----SAYRGGV 740
M.pulmonis_Lon  DFDWENNSIHIHVPDGPATPKDGPSAGVTFTTAIISSLTGKKVSPVGMTEITLRGKVLVIGGVKEKSI-----AANKFGI 799
M.musculus_Lon  FDLLDNTDIHLHFPAGAVTKDGPSAGVTIVTCLASLFSGRLVRSADVAMTGEITLRGLVLPVGGIKDKVL-----AAHRAGL 795
M.musculus_mLon1 FDLLDNTDIHLHFPAGAVTKDGPSAGVTIVTCLASLFSGRLVRSADVAMTGEITLRGLVLPVGGIKDKVL-----AAHRAGL 375
H.sapiens_pLon  FDLLDNTDIHLHFPAGAVTKDGPSAGVTIVTCLASLFSGRLVRSADVAMTGEITLRGLVLPVGGIKDKVL-----AAHRAGL 795
H.sapiens_Lon-plant1 FDLLDNTDIHLHFPAGAVTKDGPSAGVTIVTCLASLFSGRLVRSADVAMTGEITLRGLVLPVGGIKDKVL-----AAHRAGL 658
H.sapiens_mLon1  FDLLDNTDIHLHFPAGAVTKDGPSAGVTIVTCLASLFSGRLVRSADVAMTGEITLRGLVLPVGGIKDKVL-----AAHRAGL 524
A.thaliana_Lon1  MNVLDGRDIHIHFPAGAVPKDGPSAGVTLVLTALVSLFSQKRVRADTAMTGEITLRGLVLPVGGIKDKIL-----AAHRYGI 504
H.sapiens_Lonlike NDYLVTSIHLHVPAGATPKDGPSAGCTIVTALLSLAMGRVPRQNLAMTGEVSLTGKILPVGGIKEKTI-----AAKRAGV 793
H.sapiens_mLon  NDYLVTSIHLHVPAGATPKDGPSAGCAIVTALLSLAMGRVPRQNLAMTGEVSLTGKILPVGGIKEKTI-----AAKRAGV 910
M.musculus_mLon  NDYLVTSIHLHVPAGATPKDGPSAGCTIVTALLSLALGQVPLQNLAMTGEVSLTGKILPVGGIKEKTI-----AAKRAGV 896
D.melanogaster_LD30525 NLFLEQEHILHVPAGATPKDGPSAGITIIITLVSLATGKQVPRQDIAMTGEVSLGKQVLPVGGIKEKTI-----AARRSGV 758
D.melanogaster_CD8798 NLFLEQEHILHVPAGATPKDGPSAGITIIITLVSLATGKQVPRQDIAMTGEVSLGKQVLPVGGIKEKTI-----AARRSGV 932
C.elegans_mLon   NKFFDKAHIHVPAGATPKDGPSAGVTLVSSLLSLALDRPVVQDMAMTGEISLTGKVLVIGGVKIREKVI-----AARRVGA 930
A.thaliana_nLonm NQFFANSKLHLHVPAGATPKDGPSAGCTMITSLLSLATKKPVRKDLAMTGEVTLTGRILPVGGIKEKTI-----AARRSQI 894
A.thaliana_mLon  NQFFANSKLHLHVPAGATPKDGPSAGCTMITSLLSLATKKPVRKDLAMTGEVTLTGRILPVGGIKEKTI-----AARRSQI 384
A.thaliana_Lon  NLFANSKLHLHVPAGATPKDGPSAGCTMITSFSLSLAMKLVKRDAMTGEVTLTGRILPVGGIKEKTI-----AARRSQI 880
A.thaliana_LonLike NKLANSKLHLHVPAGATPKDGPSAGCTMITSLLSLALKKPVRKDLAMTGEVTLTGRILAIIGGVKEKTI-----AARRSQV 898
S.cerevisiae_PIM1 NRFFEKASIHLCPEGATPKDGPSAGVTMATSFLSLALNKSIDPTVAMTGEITLTGKVLRIIGGLREKAV-----AAKRSGA 1067
C.elegans_Lon   --TLEDQDIHVHLPAGAVNKDGPSAGTGLACALVSLATNIPLRSDAAVTGEISLTGHVLPVGGIKEKVL-----AAQREGL 724
B.subtilis_LonB GMKPSDYDIHINFPGG-IPIDGPSAGIAMAGIFSAIHKIPIDNTVAMTGEISLNLGLVVKPIGGVIPKIK-----AAKQSGA 497
T.acidophilum_Lon -----MDIHIQFVGTVEGVEGDSASVS IATAVISAIENIPVDQSVAMTGSLSVRGDVLPVGGVTAKVE-----AAIEAGL 577
E.coli_sms      GLQMDQDVFNVVGKVTET-SADLALLAMVSSLRDRLPQDLVVFGEVGLAGEIRVPSGQERIS-----EAAKHGF 424
E.coli_b0955    -----FSASLTFEQSYSEVDGDSASMAELCALISALADVPNQSIATIGSVDPQFGRAQPVGGLNEKIEGFFAICQQRELTGK 497

```

(a)



(b)

Fig. 2. Comparison of *E. coli* Lon orthologs from representative organisms. (a) Multiple sequence alignment of conserved residues surrounding the active site residues of Lon and its orthologs was performed using ClustalW. Serine and lysine residues that form the catalytic dyad are shown in bold. (b) The dendrogram depicts the relatedness between different Lon orthologs. The data from ClustalW was utilized to create the tree using the program TreeView 1.6.6. (<http://taxonomy.zoology.gla.ac.uk/rod/rod.html>). Accession numbers of orthologs are as follows: *A. thaliana*_Lon, NP_566258; *A. thaliana*_Lon1, AAO00734; *A. thaliana*_Lonlike, NP_566259; *A. thaliana*_mLon, AAL09722 (m); *A. thaliana*_nLonm, AAB48000 (m); *B. subtilis*_LonA, NP_390698; *B. subtilis*_LonB, NP_390699; *C. elegans*_Lon, NP_499577 (n); *C. elegans*_mLon, NP_492796 (m); *D. melanogaster*_LD30525, AAK93211; *D. melanogaster*_CG8798, AAF49134; *E. coli*_Lon, AAC73542; *E. coli*_sms, NP_418806; *E. coli*_b0955, NP_415475; *H. sapiens*_mLon, AAA61616 (m); *H. sapiens*_mLon1, BAB55278; *H. sapiens*_Lon-plantlike, CAD38889; *H. sapiens*_Lonlike, CAA52291 (m); *H. sapiens*_pLon, CAD68987 (p); *M. musculus*_Lon, BAC34137 (m); *M. musculus*_mLon, AAN85210 (m); *M. musculus*_mLon1, BAC35908; *M. genitalium*_Lon, NP_072905; *M. penetrans*_Lon, NP_758079; *M. pneumoniae*_Lon, NP_110020; *M. pulmonis*_Lon, NP_326350; *S. cerevisiae*_PIM1, CAA52634; *T. acidophilum*_Lon, NP_394540. Alphabets in parenthesis after the accession numbers of certain eukaryotic orthologs represent the respective annotated localizations: m, mitochondrial; n, nuclear and p, peroxisomal.

Table 2

Catalogue of known substrates hydrolyzed by Lon, ClpP and ClpQ

Enzyme	EC number	In vitro substrates	Cellular substrates
Lon	3.4.21.53	Bacteriophage λ protein N Casein CNBr fragments of BSA Denatured IgG Glucagon Oxidized insulin B chain Fluorogenic peptides, e.g., Suc-AAF-AMC	SulA, RcsA Bacteriophage λ protein N Abnormal/misfolded proteins Aggregated proteins, aged proteins Oxidized mitochondrial aconitase Mutant alkaline phosphatase PhoA61 UmuC and UmuD, subunits of error prone DNA polymerase CcrM, an essential DNA methyl transferase in <i>C. crescentus</i>
ClpP	3.4.21.92	Bacteriophage λ protein O β -Galactosidase fusion proteins Casein BSA Glucagon SsrA-tagged proteins Oxidized insulin B chain Fluorogenic peptides, e.g., Suc-LY-AMC	Proteins induced on C-starvation Abnormal proteins, canavanylated proteins Heat shock damaged-proteins σ^S , SsrA-tagged proteins, λ protein O ClpA, RepA, Phage P1 PhD N-end rule substrates UmuD/UmuD', dimer of error prone DNA polymerase MazE, endogenous regulator of cell death in <i>E. coli</i> Dps, DNA protein induced during starvation DksA, DnaK suppressor protein <i>C. crescentus</i> CtrA, a key transcriptional regulator <i>B. subtilis</i> ComK, positive regulator of competence, CtsR, heat shock regulon repressor, and MurAA, involved in synthesis of peptidoglycan <i>Salmonella</i> FlhD and FlhC, regulators of flagellum biosynthesis Phage Mu transposase and mutated vir repressor
ClpQ	3.4.25.-	Casein Insulin B chain Fluorogenic peptides, e.g., Cbz-GGL- β NA	SulA RcsA σ^{32} Abnormal/misfolded proteins Heat shock damaged-proteins

archaeobacterium *Thermococcus kodakaraensis* has been characterized: it utilizes ATP to hydrolyze folded proteins and is membrane-bound [9]. *Lactococcus* and the eubacterial pathogenic representatives *Mycobacterium tuberculosis* and *M. leprae* lack *lon* orthologs in their genomes; however, *M. smegmatis*, a non-pathogenic relative, harbors Lon [23]. All four *Mycoplasma* species analyzed harbor Lon orthologs and further biochemical characterization will be required to address whether these Lon orthologs possess proteolytic activity. Interestingly, two Lon paralogs, sms (b4389) and b0955, were detected in the *E. coli* genome (Fig. 2a) by BLASTP analysis and both possess the conserved catalytic serine of Lon (Fig. 2a); b0955 lines up well with the catalytic lysine whereas sms contains arginine instead of lysine. Both paralogs lack the ATP binding motif (data not shown) and appear to be more similar to Lon from *T. acidophilum* compared to *E. coli* (Fig. 2b). Further studies are required to investigate whether they bind to ATPase regulators and function as proteases.

E. coli cells lacking Lon are viable but display well characterized defects, for example, sensitivity to ultraviolet light, methylmethane sulfonate and overproduction of capsular polysaccharide. The absence of Lon leads to the stabilization of SulA and RcsA, which are responsible for the sensitivity to UV and capsular polysaccharide overproduction, respectively. Lon is the most efficient protease in degrading mis-

folded proteins in cells lacking DnaK, the heat shock protein (Hsp)-70 ortholog in *E. coli* [43]. In agreement with these findings, *E. coli* lacking Lon contains three times more aggregated proteins compared to the wild type [35]. However, overexpression of Lon in *E. coli* leads to reduced growth due to activation of the YoeB toxin which induces cleavage of translated mRNAs leading to cell death [6].

Lon orthologs appear to have acquired specialized functions in different organisms. A dramatic phenotype is observed in *Myxococcus xanthus*, a Gram-negative soil bacterium that forms fruiting bodies on nutrient starvation and possesses two Lon paralogs. Lon V is essential for vegetative growth [41] whereas Lon D is required for fruiting body formation [42]. In the *Caulobacter lon* mutant, cell division does not go to completion and cell cycle control of initiation of replication is disturbed due to the stabilization of a Lon substrate, an adenine DNA methyltransferase, CcrM [47]. In eukaryotes, Lon plays a major role in mitochondrial function. An increase in expression of Lon leads to a concomitant increase in mitochondrial biogenesis in hepatoma cells [28]. Together, it appears that Lon is an important protease, conserved in different kingdoms. Further elucidation of the crystal structure of full length Lon will be useful in understanding the mechanisms utilized by a protease that harbors both the regulatory and catalytic sites within a single polypeptide.

3. Clp

Clps are barrel-shaped ATP-dependent proteases composed of two components: one catalytic and the other regulatory (Fig. 1b). Two distinct Clp proteolytic subunits of ~23 kDa are known: ClpP and ClpQ (HslV). ClpP is a serine protease, whereas ClpQ belongs to the N-terminal hydrolase family of enzymes with Thr-1 as the active site residue, similar to the β subunit of 20S proteasomes. Interestingly, catalytically inactive ClpP paralogs, known as ClpR, are found associated with active Clp complexes in cyanobacteria and plants [33,34]. Clp regulators belong to the Clp/Hsp100 family of ATPases, ClpA, B, X and Y (HslU), and are structurally and functionally distinct. ClpA and ClpB possess two nucleotide binding domains (NBD) whereas ClpX contains a single NBD. ClpY contains an extra domain (I domain) which is inserted into the NBD domain (reviewed in [7]). ClpB exclusively acts as a chaperone whereas ClpA and ClpX function as chaperones and also bind to ClpP to form the ATP-dependent proteases ClpAP or ClpXP. ClpY, however, specifically interacts with ClpQ to form the ClpYQ protease. The presence of a tripeptide motif matching (LIVMP)-G-(FL), identified by sequence alignments, mutation analysis and biochemical studies, is an indicator of Clp ATPase binding to ClpP [22]. This motif is present in ClpA and ClpX but is absent in ClpB and ClpY (data not shown), which explains the lack of binding of these regulators to ClpP. The presence of this motif is useful for predicting ClpP binding subunits in other organisms [33]. Three Clp regulators are known in *B. subtilis*: the two NBD containing regulators, ClpC and ClpE, and the single NBD containing regulator, ClpX. Four distinct ClpP binding regulators are encoded in the *Arabidopsis thaliana* genome: ClpC, ClpD, ClpS and ClpX [33].

The structure and biochemistry of Clp proteases has been the focus of several studies. ClpP comprises two heptameric rings consisting of an internal central chamber ~50 Å in diameter with 14 serine active sites (Ser-111), one from each subunit, and contains narrow axial pores of ~10 Å diameter for substrate entry [45]. ClpP by itself is able to hydrolyze small peptides but not large peptides or proteins, as access to the proteolytic chamber requires passage through small axial pores. ATP and its non-hydrolyzable forms promote Clp oligomerization (regulator plus protease complexes); however only ATP can support protein unfolding and degradation [36]. ClpA or ClpX forms a single hexameric ring that associates at one or either end of ClpP (Fig. 1b). In fact, ClpP has the potential to form mixed complexes with ClpA and ClpX binding to either end to form ClpAPX complexes, which have the potential to bind to a greater range of cellular substrates [14]. ClpA and ClpX bind, unfold and translocate green fluorescent protein-tmRNA encoded peptide tag to ClpP [36]. In fact, unfolding of substrate protein by ClpXP is rate-limiting [21] and requires a lot of energy [20]. ATP-dependent enzymes

involved in cytosolic degradation unfold structural elements next to the degradation signal; thus the local structure next to the degradation signal greatly influences the amount of ATP utilized during unfolding of the substrate [20,27]. Cryoelectron microscopy and ClpP mutants were used to demonstrate that substrate proteins bound to distal surfaces followed by accumulation at the inner sides of ATPases and translocation into the catalytic chamber of ClpP [15].

ClpP orthologs are present in some organisms selected for analysis (Table 1). The catalytic triad in *E. coli* ClpP consists of Ser-111, His-136 and Asp-185 and these residues are conserved in ClpP orthologs (Fig. 3). Most archaeal genomes appear to lack ClpP subunits, including *T. acidophilum* and *S. solfataricus*. In addition, *M. jannaschii*, *P. furiosus*, *Halobacterium* sp. NRC-1 and *M. mazei* lack ClpP orthologs (data not shown). None of the genomes of *Mycoplasma* species analyzed harbor ClpP orthologs; these are also absent in *S. cerevisiae* and *S. pombe* (data not shown). *L. lactis*, *M. tuberculosis* and *M. leprae*, which lack *lon*, harbor *clpP* orthologs (Table 1; Fig. 3). Interestingly, the genome of *A. thaliana* encodes six catalytically active paralogs of ClpP [33], some of which are included in our analysis (Fig. 3). In *A. thaliana*, plastid Clp containing complexes are composed of several subunits (ClpP1,3-6) whereas the mitochondrial Clp complex is composed of only the ClpP2 subunit [33]. The differential subunit composition of ClpP complexes in plastids and mitochondria in *A. thaliana* may modulate proteolysis in different organelles.

ClpAP and ClpXP proteases are responsible for several cellular activities and display distinct substrate specificities (reviewed in [7,17]). In general, ClpAP is able to trap different substrate proteins whereas ClpXP is better at trapping substrate proteins possessing a targeting motif. Using a catalytically inactive ClpP mutant followed by identification of the trapped substrates by mass spectrometry, several ClpP substrates, including DnaK suppressor protein (DksA) and the DNA protective protein induced during starvation (Dps), have been identified. Interestingly, RpoS is the most abundant ClpXP trapped substrate. This strategy led to the identification of at least five distinct signals, present at either the N-terminus or the C-terminus of ClpXP substrates [8]. ClpP complexes are also responsible for degradation of proteins targeted by tmRNA. Functional insights into the tmRNA tagging system have revealed that ClpA and ClpX bind to different stretches of the tmRNA encoded peptide tag. Most likely, ClpXP is the major protease involved in degradation of tmRNA encoded peptide tagged proteins in vivo [13].

ClpP is involved in survival of *E. coli* during the stationary phase and levels of several proteins (e.g., Dps, 30S ribosomal protein S6 RS-6B etc.) are modulated in *E. coli* lacking ClpP in the stationary phase [46]. In different organisms ClpP proteins are known to participate in various physiological processes; for example, it is involved in cell cycle progression from G1 to S in *Caulobacter* [16]. *B. subtilis* lacking ClpP display multiple defects, including the inability to grow at high temperatures, competence development,

<i>E. coli</i> _ClpP	VSTICMGQAASMGAFLLTAGAKGKRFCLPNSRVMIHQ	138
<i>B. subtilis</i> _ClpP	VSTICIGMAASMGAFLLAAGEKGRYALPNSEVMIHQ	125
<i>H. sapiens</i> _ClpP	ICTWCVGQAASMGSLLLAAGTPGMRHSLPNSRIMIHQ	180
<i>M. musculus</i> _ClpP	ICTWCVGQAASMGSLLLAAGSPGMRHSLPNSRIMIHQ	176
<i>C. elegans</i> _ClpP	VSTWVIGQAASMGSLLLCAGEKGRSALPNSRIMVHQ	132
<i>A. thaliana</i> _ClpP1	ISTICLGQAASMASLLLAAGAKGQRRSLPNATVMIHQ	154
<i>D. melanogaster</i> _CG5045	IATWCVGQAASMGSLLLAAGAPGMRYSLPNARIMIHQ	151
<i>A. thaliana</i> _nClpP1	VSTVCGLAASMGAFLLSAGTKGKRYSLPNSRIMIHQ	220
<i>L. lactis</i> _ClpP	VQTIVMGMAASMGTI IASSGKGRFMLPNAEYLHQ	126
<i>M. tuberculosis</i> _ClpP	IATYAMGMAASMGFLLAAGTKGKRYALPHARILMHQ	125
<i>M. leprae</i> _ClpP	IATYAMGMAASMGFLLAAGTRGKRYALPHARILMHQ	149
<i>A. thaliana</i> _nClpP3	VSTVCLGLAASMGAFLLASGSKGKRYCMPNSKVMHQ	191
<i>A. thaliana</i> _nClpP4	VSTIALGIAASTAS IILGAGTKGKRFAMPNTRIMIHQ	185
<i>M. tuberculosis</i> _ClpP2	IQTVCLGQAASAAAVLLAAGTPGKRMALPNARVLHQ	137
<i>M. leprae</i> _ClpP2	IQTVCLGQAASAAAVLLAAGTPGKRMALPNARVLHQ	137
	: * : * * . . : : * * * : * : : * * *	
<i>E. coli</i> _ClpP	TERDRFLSAPEAVEYGLVDSILTHRN-----	207
<i>B. subtilis</i> _ClpP	TDRDNFKSAEEALEYGLIDKILTHTEDKK----	197
<i>H. sapiens</i> _ClpP	MERDRYMSPEAEQEFGLDKVLVHPPQDGEDEP	256
<i>M. musculus</i> _ClpP	MERDRYMSPEAEQEFGLDKVLVHPPQDGEDEP	252
<i>C. elegans</i> _ClpP	LDRDRFMSAHEALKFGLVDQIETHN-----	200
<i>A. thaliana</i> _ClpP1	MDRDHFMTPEEAKAFGIIDEVIDERPLELVKD-	229
<i>D. melanogaster</i> _CG5045	MERDHFMTPEEAKVLGIIDHVLEHPPETVSETG	227
<i>A. thaliana</i> _nClpP1	TDRDFMSAKEAEYGLIDGVIMNPLKALQPLA	296
<i>L. lactis</i> _ClpP	AERDHWMDAKETLEYGFIDEIMEN--NSLK---	199
<i>M. tuberculosis</i> _ClpP	SDRDRWFTAAEALYGFVDHIITRAHVNGEAQ-	200
<i>M. leprae</i> _ClpP	SDRDRWFTASEALYGFVDHIITRAHVNGEVQ-	224
<i>A. thaliana</i> _nClpP3	TDRDNFLNPWEAEYGLIDAVIDDGKPLIAP I	267
<i>A. thaliana</i> _nClpP4	IDRDRYMSPIEAVEYGLIDGVIDGDSIIPLEPV	261
<i>M. tuberculosis</i> _ClpP2	TDRDKILTAEAKDYGIIDTVLEYRKLQAQTA-	214
<i>M. leprae</i> _ClpP2	TDRDKILTAEAKDYGIIDTVLEYRKLQAQTV-	214
	: * * . * : * * * : * : * :	

Fig. 3. Alignment of sequences flanking the active site residues in ClpP orthologs. Multiple sequence alignment of mouse ClpP orthologs surrounding the three catalytic residues (bold) from representative organisms was performed using ClustalW. The residues important for activity are shown in bold. Accession numbers of the corresponding orthologs are as follows: *A. thaliana*_ClpP1, BAC43126; *A. thaliana*_nClpP1, BAA82065; *A. thaliana*_nClpP3, AAL66941; *A. thaliana*_nClpP4, AAK68772; *B. subtilis*_ClpP, AAC46381; *C. elegans*_ClpP, NP_496215; *D. melanogaster*_CG5045, NP_609388; *E. coli*_ClpP, AAC73540; *H. sapiens*_ClpP, AAH02956; *L. lactis*_ClpP, NP_266829; *M. leprae*_ClpP, CAC30430; *M. leprae*_ClpP2, CAC30429; *M. musculus*_ClpP, NP_059089 (annotated mitochondrial localization); *M. tuberculosis*_ClpP, CAA16038; *M. tuberculosis*_ClpP2, CAA16037.

motility and sporulation [31]. ClpP orthologs are essential in *C. reinhardtii* [29], the cyanobacterium *Synechococcus* [34] and tobacco [25]. Although both mice and humans possess a ClpP ortholog (Table 1), the function of this ClpP has not been identified.

ClpYQ (HslUV) was identified as a heat shock operon with similarity to Clp ATPases and 20S proteasomal β subunits. ClpY and ClpQ can exist separately and ATP is required for oligomerization of the ClpYQ complex and protease activity. The crystal structures of ClpYQ [37] revealed a close resemblance with 20S proteasomes with a similar subunit structure and intersubunit interactions. However, 20S proteasomes are composed of heptameric rings, whereas ClpYQ are composed of hexameric rings (Fig. 1b). The unique I-domain in ClpY is thought to extend outward from the complex and assist in the binding of substrates; however, ClpY lacking the I-domain can form intact complexes that are indistinguishable from wild type ClpYQ and display proteolytic activity [26]. Further studies are required to investigate the functional roles of I domain-lacking ClpYQ complexes in vivo. The N-terminal catalytic Thr-1 of ClpQ, similar to the 20S proteasomal β subunit, is the active site nucleophile [50]. Quite surprisingly, ClpYQ of *B. subtilis* (CodWX) utilizes a serine residue present prior to

the double threonines as the active site residue. In fact, this enzyme is the only known N-terminal serine protease and mutation of this serine residue to alanine or even threonine abolishes the peptidase activity, although the ATPase activity remains intact. Interestingly, mutation of double threonines resulted in inability to form the tetradecamer, resulting in loss of peptidase activity as well as ATPase activity [18]. Biochemical studies and electron micrographs have revealed that ClpYQ from *B. subtilis* is distinct from its *E. coli* counterpart [19].

Of the representative organisms selected for the present analysis, only *E. coli* and *B. subtilis* encoded *clpQ* sequences (Table 1). It was thought that 20S proteasomes and ClpYQ are mutually exclusive; however, genomes of several microbes are now known to harbor both 20S proteasomes and ClpYQ or lack both these proteases [11]. Significantly, ClpYQ overexpression can rescue *lon*-deficient cells by hydrolyzing most Lon substrates, including SulA, RcsA and RpoH [48]. Together, Clp proteases possess specialized functions in various organisms. However, most studies on Clp complexes have been performed on lower organisms. Further studies are required to identify cellular substrates and functional roles of ClpP in organelles of plants and animals.

4. Conclusions

The current genomic survey and review of the literature has revealed certain interesting features: (i) It had been argued previously that common genes present in two small bacterial genomes, *Mycoplasma* and *H. influenzae*, represent genes important in cellular metabolism [32]. Lon appears to be the main ATP-dependent enzyme in *Mycoplasma*. However, *H. influenzae* lacks Lon but contains both ClpXP and ClpYQ (data not shown). In our analysis, *S. solfataricus* is the only organism lacking both ClpP and Lon orthologs. An AAA-ATPase (NP 343775) from *Sulfolobus* displays significant similarity (E value-1e-05) with Lon, although the active site residues could not be aligned. It is possible that this potential Lon ortholog has diverged and further functional characterization of this enzyme is required to ascertain whether it can function as a protease. However, proteasome-activating nucleotidase (PAN), an AAA-ATPase family member, and 20S proteasomes are both present in *S. solfataricus* (data not shown) and may play an important role during ATP-dependent cytosolic protein degradation in this organism. This clearly indicates that, in principle, at least one ATP-dependent enzyme is encoded in genomes from different kingdoms. (ii) Both chaperones and proteases are known to be induced in *E. coli* during heat shock. Further genetic and biochemical studies are required to identify cellular protein substrates and decipher the roles of individual protease orthologs during different stress conditions. (iii) *E. coli* proteases (Clp and Lon), unlike their eukaryotic counterparts (26S proteasomes), are not essential. However, these proteases may be essential under some conditions or in some organisms (discussed above). *E. coli* lacking ClpP and Lon are unable to survive the transition from growth in rich media to minimal media, a form of stress known as nutritional downshift [24]. *E. coli* lacking Lon is impaired in degradation of puromycyl peptides. However, this effect was not observed in *M. smegmatis* lacking Lon and 20S proteasomes [23]. The relevance of two ClpP paralogs in pathogens, e.g., *M. tuberculosis* and *M. leprae*, both lacking Lon (Table 1; Fig. 3), remains to be evaluated. Further functional studies are required to understand the roles of protease orthologs in different organisms. (iv) One of the interesting aspects that have emerged recently is the role of enzymes involved in cytosolic protein degradation in pathogenesis. *clpP* and *clpX*-deficient *S. typhimurium* strains are unable to form colonies in vivo due to impaired ability to grow intracellularly in macrophages [49]. Also, ClpP is required for virulence in *Listeria monocytogenes* [10]. *S. typhimurium* lacking *lon* express high levels of proteins encoded by *Salmonella* pathogenicity island (SPI)-I genes, demonstrating that Lon is a negative regulator [3,39]. The failure of *lon* deficient *S. typhimurium* to downmodulate SPI-I and upregulate SPI-II genes is probably responsible for the inability of these cells to cause systemic infection in a mouse model of infection [40]. These results have im-

plicated Lon- or ClpP-lacking strains as candidate vaccine targets [10,30].

In summary, the complete genome sequences of different organisms offer an opportunity to experimentally explore the roles of newer enzymes and their orthologs in the field of cytosolic protein degradation. As ATP-dependent proteases play key roles in initiating the process of cytosolic protein degradation, further experimental verification using a systematic deletion strategy needs to evaluate the roles of protease orthologs in growth (in vitro and in vivo), different stress conditions and pathogenesis. Studies on these enzymes will reveal their distinct functions in different organelles and organisms and will consequently enrich this area of research.

Acknowledgements

We thank Anujith Kumar and Pankaj Tahiliani for their help with preparation of this manuscript. The comments of researchers in the field of *E. coli* cytosolic protein degradation are greatly appreciated. Only selected references were incorporated due to limitations of space and we apologize to researchers whose contributions were not cited. This study has been supported by a grant from the Department of Science and Technology, Government of India.

References

- [1] D. Biran, E. Gur, L. Gollan, E.Z. Ron, Control of methionine biosynthesis in *Escherichia coli* by proteolysis, *Mol. Microbiol.* 37 (2000) 1436–1443.
- [2] C. Birghan, E. Mundt, A.E. Gorbalenya, A non-canonical Lon proteinase lacking the ATPase domain employs the Ser–Lys catalytic dyad to exercise broad control over the life cycle of a double-stranded RNA virus, *EMBO J.* 19 (2000) 114–123.
- [3] J.D. Boddicker, B.D. Jones, Lon protease activity causes down-regulation of *Salmonella* pathogenicity island 1 invasion gene expression after infection of epithelial cells, *Infect. Immun.* 72 (2004) 2002–2013.
- [4] I. Botos, E.E. Melnikov, S. Cherry, J.E. Tropea, A.G. Khalatova, F. Rasulova, Z. Dauter, M.R. Maurizi, T.V. Rotanova, A. Wlodawer, A. Gustchina, The catalytic domain of *E. coli* Lon protease has a unique fold and a Ser–Lys dyad in the active site, *J. Biol. Chem.* 279 (2004) 8140–8148.
- [5] D. Chandu, D. Nandi, From proteins to peptides to amino acids: Comparative genomics of enzymes involved in downstream processing events during cytosolic protein degradation, *Appl. Gen. Proteom.* 1 (2002) 235–252.
- [6] S.K. Christensen, G. Maenhaut-Michel, N. Mine, S. Gottesman, K. Gerdes, L. Van Melderen, Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: Involvement of the yefM–yoeB toxin–antitoxin system, *Mol. Microbiol.* 51 (2004) 1705–1717.
- [7] D.A. Dougan, A. Mogk, B. Bukau, Protein folding and degradation in bacteria: To degrade or not to degrade? That is the question, *Cell Mol. Life Sci.* 59 (2002) 1607–1616.
- [8] J.M. Flynn, S.B. Neher, Y.I. Kim, R.T. Sauer, T.A. Baker, Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals, *Mol. Cell* 11 (2003) 671–683.

- [9] T. Fukui, T. Eguchi, H. Atomi, T. Imanaka, A membrane-bound archaeal Lon protease displays ATP-independent proteolytic activity towards unfolded proteins and ATP-dependent activity for folded proteins, *J. Bacteriol.* 184 (2002) 3689–3698.
- [10] O. Gaillot, E. Pellegrini, S. Bregenholt, S. Nair, P. Berche, The ClpP serine protease is essential for the intracellular parasitism and virulence of *Listeria monocytogenes*, *Mol. Microbiol.* 35 (2000) 1286–1294.
- [11] C. Gille, A. Goede, C. Schloetelburg, R. Preissner, P.M. Klotzel, U.B. Gobel, C. Frommel, A comprehensive view on proteasomal sequences: Implications for the evolution of the proteasome, *J. Mol. Biol.* 326 (2003) 1437–1448.
- [12] R. Gillet, B. Felden, Emerging views on tmRNA-mediated protein tagging and ribosome rescue, *Mol. Microbiol.* 42 (2001) 879–885.
- [13] S. Gottesman, E. Roche, Y.N. Zhou, R.T. Sauer, The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system, *Gen. Dev.* 12 (1998) 1338–1347.
- [14] J. Ortega, H.S. Lee, M.R. Maurizi, A.C. Steven, ClpA and ClpX ATPases bind simultaneously to opposite ends of ClpP peptidase to form active hybrid complexes, *J. Struct. Biol.* 146 (2004) 217–226.
- [15] T. Ishikawa, F. Beuron, M. Kessel, S. Wickner, M.R. Maurizi, A.C. Steven, Translocation pathway of protein substrates in ClpAP protease, *Proc. Natl. Acad. Sci. USA* 98 (2001) 4328–4333.
- [16] U. Jenal, T. Fuchs, An essential protease involved in bacterial cell-cycle control, *EMBO J.* 17 (1998) 5658–5669.
- [17] U. Jenal, R. Hengge-Aronis, Regulation of proteolysis in bacterial cells, *Curr. Opin. Microbiol.* 6 (2003) 163–172.
- [18] M.S. Kang, B.K. Lim, I.S. Seong, J.H. Seol, N. Tanahashi, K. Tanaka, C.H. Chung, The ATP-dependent CodWX (HslVU) protease in *Bacillus subtilis* is an N-terminal serine protease, *EMBO J.* 20 (2001) 734–742.
- [19] M.S. Kang, S.R. Kim, P. Kwack, B.K. Lim, S.W. Ahn, Y.M. Rho, I.S. Seong, S.C. Park, S.H. Eom, G.W. Cheong, C.H. Chung, Molecular architecture of the ATP dependent CodWX protease having an N-terminal serine active site, *EMBO J.* 22 (2003) 2893–2902.
- [20] J.A. Kenniston, T.A. Baker, J.M. Fernandez, R.T. Sauer, Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine, *Cell* 114 (2003) 511–520.
- [21] Y.I. Kim, R.E. Burton, B.M. Burton, R.T. Sauer, T.A. Baker, Dynamics of substrate denaturation and translocation by the ClpXP degradation machine, *Mol. Cell* 5 (2000) 639–648.
- [22] Y.I. Kim, I. Levchenko, K. Fraczkowska, R.V. Woodruff, R.T. Sauer, T.A. Baker, Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase, *Nat. Struct. Biol.* 8 (2001) 230–233.
- [23] N. Knipfer, A. Seth, S.G. Roudiak, T.E. Shrader, Species variation in ATP-dependent protein degradation: Protease profiles differ between mycobacteria and protease functions differ between *Mycobacterium smegmatis* and *Escherichia coli*, *Gene* 231 (1999) 95–104.
- [24] A. Kuroda, K. Nomura, R. Ohtomo, J. Kata, T. Ikeda, N. Takiguchi, H. Ohtake, A. Kornberg, Role of inorganic phosphate in promoting ribosomal protein degradation by the lon protease in *E. coli*, *Science* 293 (2001) 705–708.
- [25] H. Kuroda, P. Maliga, The plastid clpP1 protease gene is essential for plant development, *Nature* 425 (2003) 86–89.
- [26] A.R. Kwon, B.M. Kessler, H.S. Overkleeft, D.B. McKay, Structure and reactivity of an asymmetric complex between HslV and I-domain deleted HslU, a prokaryotic homologue of the eukaryotic proteasome, *J. Mol. Biol.* 330 (2003) 185–195.
- [27] C. Lee, M.P. Schwartz, S. Prakash, M. Iwakura, A. Matouschek, ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal, *Mol. Cell* 7 (2001) 627–637.
- [28] K. Luciakova, B. Sokolikova, M. Chloupkova, B.D. Nelson, Enhanced mitochondrial biogenesis is associated with increased expression of the mitochondrial ATP-dependent Lon protease, *FEBS Lett.* 444 (1999) 186–188.
- [29] C. Huang, S. Wang, L. Chen, C. Lemieux, C. Otis, M. Turmel, X.Q. Liu, The *Chlamydomonas* chloroplast clpP gene contains translated large insertion sequences and is essential for cell growth, *Mol. Gen. Genet.* 244 (1994) 151–159.
- [30] H. Matsui, M. Suzuki, Y. Isshiki, C. Kodama, M. Eguchi, Y. Kikuchi, K. Motokawa, A. Takaya, T. Tomoyasu, T. Yamamoto, Oral immunization with ATP-dependent protease-deficient mutants protects mice against subsequent oral challenge with virulent *Salmonella enterica* serovar *typhimurium*, *Infect. Immun.* 71 (2003) 30–39.
- [31] T. Msadek, V. Dartois, F. Kunst, M.L. Herbaud, F. Denizot, G. Rapoport, ClpP of *Bacillus subtilis* is required for competence development, motility, degradative enzyme synthesis, growth at high temperature and sporulation, *Mol. Microbiol.* 27 (1998) 899–914.
- [32] A.R. Mushegian, E.V. Koonin, A minimal gene set for cellular life derived by comparison of complete bacterial genomes, *Proc. Natl. Acad. Sci. USA* 93 (1996) 10268–10273.
- [33] J.-B. Peltier, D.R. Ripoll, G. Friso, A. Rudella, Y. Cai, J. Ytterber, L. Giacomelli, J. Pillardy, K.J. van Wijk, 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 (2004) 4768–4781.
- [34] J. Schelin, F. Lindmark, A.K. Clarke, The *clpP* multigene family for the ATP-dependent Clp protease in the cyanobacterium *Synechococcus*, *Microbiology* 148 (2002) 2255–2265.
- [35] R. Rosen, D. Biran, E. Gur, D. Becher, M. Hecker, E.Z. Ron, Protein aggregation in *Escherichia coli*: Role of proteases, *FEMS Microbiol. Lett.* 207 (2002) 9–12.
- [36] S.K. Singh, R. Grimaud, J.R. Hoskins, S. Wickner, M.R. Maurizi, Unfolding and internalization of proteins by the ATP-dependent proteases ClpXP and ClpAP, *Proc. Natl. Acad. Sci. USA* 97 (2000) 8898–8903.
- [37] M.C. Sousa, C.B. Trame, H. Tsuruta, S.M. Wilbanks, V.S. Reddy, D.B. McKay, Crystal and solution structures of an HslUV protease-chaperone complex, *Cell* 103 (2000) 633–643.
- [38] H. Stahlberg, E. Kutejova, K. Suda, B. Wolpensinger, A. Lustig, G. Schatz, A. Engel, C.K. Suzuki, Mitochondrial Lon of *Saccharomyces cerevisiae* is a ring-shaped protease with seven flexible subunits, *Proc. Natl. Acad. Sci. USA* 96 (1999) 6787–6790.
- [39] A. Takaya, T. Tomoyasu, A. Tokumitsu, M. Morioka, T. Yamamoto, The ATP-dependent lon protease of *Salmonella enterica* serovar *Typhimurium* regulates invasion and expression of genes carried on *Salmonella* pathogenicity island 1, *J. Bacteriol.* 184 (2002) 224–232.
- [40] A. Takaya, M. Suzuki, H. Matsui, T. Tomoyasu, H. Sashinami, A. Nakane, T. Yamamoto, Lon, a stress-induced ATP-dependent protease, is critically important for systemic *Salmonella enterica* serovar *typhimurium* infection of mice, *Infect. Immun.* 71 (2003) 690–696.
- [41] N. Tojo, S. Inouye, T. Komano, Cloning and nucleotide sequence of the *Myxococcus xanthus* lon gene: Indispensability of lon for vegetative growth, *J. Bacteriol.* 175 (1993) 2271–2277.
- [42] N. Tojo, S. Inouye, T. Komano, The *lonD* gene is homologous to the lon gene encoding an ATP-dependent protease and is essential for the development of *Myxococcus xanthus*, *J. Bacteriol.* 175 (1993) 4545–4549.
- [43] T. Tomoyasu, A. Mogk, H. Langen, P. Goloubinoff, B. Bukau, Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol, *Mol. Microbiol.* 40 (2001) 397–413.
- [44] L. Van Melderen, S. Gottesman, Substrate sequestration by a proteolytically inactive Lon mutant, *Proc. Natl. Acad. Sci. USA* 96 (1999) 6064–6071.
- [45] J. Wang, J.A. Hartling, J.M. Flanagan, The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis, *Cell* 91 (1997) 447–456.
- [46] D. Weichart, N. Querfurth, M. Dreger, R. Hengge-Aronis, Global role for ClpP-containing proteases in stationary-phase adaptation of *Escherichia coli*, *J. Bacteriol.* 185 (2003) 115–125.

- [47] R. Wright, C. Stephens, G. Zweiger, L. Shapiro, M.R. Alley, *Caulobacter* Lon protease has a critical role in cell-cycle control of DNA methylation, *Gen. Dev.* 10 (1996) 1532–1542.
- [48] W.F. Wu, Y. Zhou, S. Gottesman, Redundant in vivo proteolytic activities of *Escherichia coli* Lon and the ClpYQ (HslUV) protease, *J. Bacteriol.* 181 (1999) 3681–3687.
- [49] T. Yamamoto, H. Sashinami, A. Takaya, T. Tomoyasu, H. Matsui, Y. Kikuchi, T. Hanawa, S. Kamiya, A. Nakane, Disruption of the genes for ClpXP protease in *Salmonella enterica* serovar *Typhimurium* results in persistent infection in mice, and development of persistence requires endogenous gamma interferon and tumor necrosis factor alpha, *Infect. Immun.* 69 (2001) 3164–3174.
- [50] S.J. Yoo, Y.K. Shim, I.S. Seong, J.H. Seol, M.S. Kang, C.H. Chung, Mutagenesis of two N-terminal Thr and five Ser residues in HslV, the proteolytic component of the ATP-dependent HslVU protease, *FEBS Lett.* 412 (1997) 57–60.