Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains

Tatyana V. Rotanova[1], Edward E. Melnikov[1], Anna G. Khalatova[1], Oksana V. Makhovskaya[1], Istvan Botos[2], Alexander Wlodawer[2] and Alla Gustchina[2]

[1]*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia;* [2]*Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, MD, USA*

ATP-dependent Lon proteases belong to the superfamily of AAA$^+$ proteins. Until recently, the identity of the residues involved in their proteolytic active sites was not elucidated. However, the putative catalytic Ser–Lys dyad was recently suggested through sequence comparison of more than 100 Lon proteases from various sources. The presence of the catalytic dyad was experimentally confirmed by site-directed mutagenesis of the *Escherichia coli* Lon protease and by determination of the crystal structure of its proteolytic domain. Furthermore, this extensive sequence analysis allowed the definition of two subfamilies of Lon proteases, LonA and LonB, based on the consensus sequences in the active sites of their proteolytic domains. These differences strictly associate with the specific characteristics of their AAA$^+$ modules, as well as with the presence or absence of an N-terminal domain.

Keywords: AAA$^+$ proteins; Lon proteases; proteolytic site; LonA and LonB subfamilies; Ser–Lys dyad.

ATP-dependent proteases assigned to the Lon family are key enzymes responsible for intracellular selective proteolysis, which controls protein quality and maintains cellular homeostasis. These enzymes eliminate mutant and abnormal proteins and play an important role in the rapid turnover of short-lived regulatory proteins [1–5]. Lon proteases are conserved in prokaryotes and in eukaryotic organelles such as mitochondria. Lon and all other known ATP-dependent proteases (FtsH, ClpAP, ClpXP, and HslVU) belong to the AAA$^+$ protein superfamily (ATPases associated with diverse cellular activities) [6–14]. Besides selective proteolysis, AAA$^+$ proteins are involved in many other cellular processes, including cell-cycle regulation, protein transport, organelle biogenesis, and microtubule severing.

The structural core of the AAA$^+$ proteins is represented by the so-called AAA$^+$ modules consisting of 220–250 residues [6,12], which occur either singly or as repeats. Although in the majority of AAA$^+$ proteins the AAA$^+$ modules are located within a separate subunit of the protein, in some, including Lon, such modules can form domains within a single polypeptide chain.

The AAA$^+$ modules consist of two domains: a larger N-terminal nucleotide-binding domain (or α/β domain) and a smaller C-terminal helical domain (α domain). The sequences of the α/β domains contain some conserved motifs, including Walker A and B as well as sensor-1, which take part in nucleotide binding [6]. The α domains also contain some conserved motifs, in particular sensor-2, with an Arg or Lys residue involved in ATP hydrolysis [6,7]. These AAA$^+$ modules participate in target selection and regulation of the functional component activity of AAA$^+$ proteins [1,6–15], and their α domains appear to mediate the transmission of free energy of ATP hydrolysis by AAA$^+$ proteins to their functional subunits and substrates [7,8].

E. coli Lon protease was the first ATP-dependent protease to be discovered [16,17], its sequence being deciphered about 15 years ago [18,19]. This protease is a cytosolic, homooligomeric enzyme and its subunit (784 amino acids) consists of three functional domains [19,20]: the N-terminal domain (N, also referred to as LAN [7]) which, possibly together with the AAA$^+$ module, can selectively interact with target proteins [7,9,21–23]; the central ATPase (AAA$^+$ module or A domain) described above; and the C-terminal proteolytic (P) domain. The identity of the catalytically active Ser679 residue in the P domain was first predicted based on sequence comparisons of serine proteases [19] and later confirmed by site-directed mutagenesis [20]. The proteolytic domain of Lon protease showed no sequence homology to any known serine proteases containing the classical catalytic Ser–His–Asp triad [17–20].

The existence of the Lon family, then consisting of ≈ 20 representatives, including enzymes from evolutionarily distant sources, was described in the late 1990s [24]. Detailed comparison of their sequences led to attempts to define other residues that could form, together with

Ser679, the catalytic site of *E. coli* Lon. Experimental verification of the role of different residues led to the preparation of a series of mutants of amino acids in *E. coli* Lon that were found to be conserved in the other Lon proteases [25], including His665, His667, and Asp676. These mutants lost their ATP-dependent proteolytic activity, leaving open the possibility of their involvement in the creation of a functional Ser–His–Asp triad. However, these residues were all located within the fragment HVHVPEGATPKDGPS(665–679), a stretch of only 15 amino acids preceding and including the catalytic Ser679. Their proximal location in the sequence did not correspond to the topology of the catalytic triad in any known subfamily of 'classical' serine proteases. At about the same time, functional catalytic hydroxyl/amine dyads were described in the active sites of some peptide hydrolases [26]. We hypothesized that a possible functional catalytic Ser–Lys dyad might also be present in the active site of Lon protease [25].

It should also be noted that the presence of a Ser–Lys dyad was reported in viral Vp4 proteases from different sources [27,28]. Vp4 and its homologues were considered to represent a unique branch of the Lon family whose P domain was not associated with an AAA$^+$ module [27]. It was also concluded that the mechanism of proteolysis utilized by Vp4 should also be conserved across the ATP-dependent Lon proteases.

In this study we follow up and expand the recent observations [29] by presenting a comparative analysis of the amino acid sequences of the majority of the currently known Lon proteases. The results of site-directed mutagenesis of *E. coli* Lon protease and insights from the crystal structure of its proteolytic domain [30] were also taken into account. This analysis proved our hypothesis about the presence of a catalytic dyad and concluded with the identification of two subfamilies of Lon proteases.

Materials and methods

Site-directed mutagenesis of *E. coli* Lon protease

Strains BL21 and HB101 (Stratagene, La Jolla, CA, USA) of *E. coli* were utilized in this study. Standard procedures were used in all DNA manipulations utilized for cloning [31]. Site-directed mutagenesis was performed using the polymerase chain reaction/splicing by overlapping extension (SOE) method [32]. Expression plasmid pBR327-*lon* [18] was used as the matrix in the first PCR step. The structure of the mutagenic primers that encode both the mutation K722Q and an additional recognition site of *Pvu*II restriction endonuclease were 5′-GGTTTGAA AGAACAG CTGCTGGCAGCG-3′ (direct primer) and 5′-ATGCGC TGCCAGCAGCTGTTCTTTCAA-3′ (reverse primer), where mismatched nucleotides are underlined. The target wild-type fragment of the *lon* gene, cloned in pBR327 vector, was replaced by the mutant PCR fragment using *Bam*HI and *Sph*I restriction sites. Plasmids isolated from transformed HB101 cells were used for restriction analysis and were tested for expression. The structure of the subcloned PCR fragment was verified by DNA sequencing.

Expression of the *lon* gene and purification of Lon protease and its mutant Lon-K722Q

Wild-type Lon protease and the mutant Lon-K722Q were expressed in *E. coli lon*-deficient strain BL21 and isolated as described previously [33]. Protein concentrations were determined by the method of Bradford [Bio-Rad (Hercules, CA, USA) protein assay] [34] using bovine serum albumin as a standard. Protein purification was monitored by SDS/PAGE by the method of Laemmli [35].

Activity assays

The proteolytic activity of the enzymes was detected through hydrolysis of β-casein using 12% SDS/PAGE. The peptidase activity was assayed by the hydrolysis of Suc-Phe-Leu-Phe-SBzl [36,37]. ATPase activity was determined as described by Bencini *et al.* [38] in the presence or absence of a protein substrate [39].

Results and Discussion

The recent availability of a large number of genomic sequences has significantly increased the number of identifiable analogs of *E. coli* Lon and prompted a reanalysis of the active sites of this family of proteases. The alignment of the proteolytic domains derived from the sequences of > 100 Lon proteases from a variety of sources provided several major insights.

Lon does not utilize a classical catalytic triad

The proteolytic domains of Lon lack strictly conserved histidine and aspartic acid residues; thus His665, His667, and Asp676 (the numbering corresponds to the sequence of *E. coli* Lon), earlier considered to be possible participants in the classical catalytic triad [25], are not conserved among all members of the Lon family. Successful determination of the crystal structure of the proteolytic domain of *E. coli* Lon [30] allowed us to explain the loss of proteolytic activity of the mutants at these sites [25]. These three residues were all found to be involved in important intra- or intermolecular interactions (Fig. 1). The side chain of Asp676 is located directly above the N-terminus of α helix 1, thus making electrostatic interactions with its positive charge and forming two hydrogen bonds with the amide nitrogens of Val633 and Met634 from this helix (not shown). His665 and His667 are located on the surface of the molecule, within an oligomeric interface of the hexameric rings of P domains. The side chains of these two residues are involved in extensive interactions with Leu709 and Thr643 of a neighboring subunit. At the same time, His667 also forms an ion pair with Glu614 belonging to its own subunit. The latter residue, in turn, is hydrogen bonded (N–O distance of 2.7 Å) to the amide nitrogen of Leu709 from the second molecule. The orientation of the side chain of His667 is also maintained due to the proximity of the negative charge of the side chain of Glu706 from the neighboring subunit. The mutation of these residues might interfere with the oligomerization required for the proteolytic activity of Lon. This analysis shows that Lon proteases do not utilize any His or Asp residues to create their active sites, eliminating the
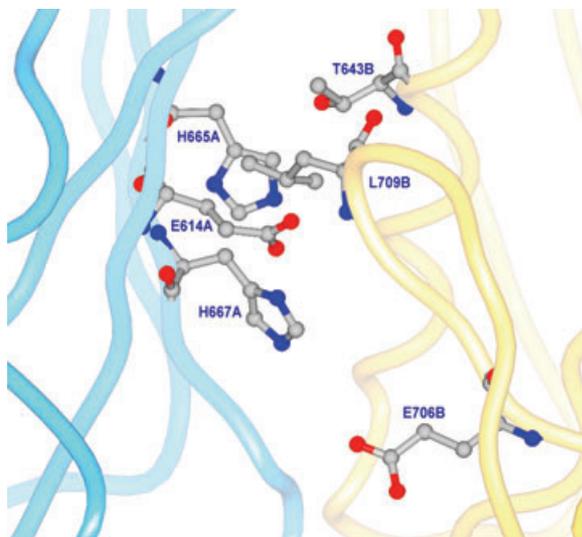
Fig. 1. Interactions of residues located within the oligomeric interface of two proteolytic domains of *E. coli* Lon provide a structural basis explaining the loss of catalytic activity of their mutants. The interacting residues, Glu614, His665, and His667 in molecule A and Thr643, Glu706, and Leu709 in molecule B, are shown in a ball-and-stick representation, whereas the main chains of the two domains are color-coded. The figure was created using the program spock [47], with coordinates from the Protein Data Bank, accession code 1rre.

Table 1. Relative enzymatic activities of *E. coli* Lon protease (Lon-wild-type) and its mutant forms Lon-S679A and Lon-K722Q. Activities were measured in 50 mM Tris/HCl buffer, pH 8.0, 0.1 M NaCl, 37 °C. Concentrations of enzymes were 1 μM for β-casein hydrolysis and 0.1 μM for Suc-Phe-Leu-Phe-SBzl hydrolysis; those of the substrates were 0.03 mM for β-casein and 0.1 mM for Suc-Phe-Leu-Phe-SBzl; ATP concentration was 2.5–5.0 mM and $MgCl_2$ 20 mM.

| | Substrate | | | |
| | β-casein | | Suc-Phe-Leu-Phe-SBzl | |
Enzyme	−ATP	+ATP	−ATP	+ATP
Lon-wt	0	100	30	100
Lon-S679A	0	0	0	0
Lon-K722Q	0	0	0	0

possibility of the presence of classical serine protease catalytic triad.

The Ser–Lys catalytic dyad

All Lon proteolytic domains contain a single conserved lysine, located 43 residues beyond the catalytic serine (Ser679 and Lys722 in *E. coli* Lon). To elucidate the role of this residue and to verify the hypothesis of the possible presence of a catalytic Ser–Lys dyad [25] we performed site-directed mutagenesis of Lys722 and investigated the effects of its mutation on the enzymatic properties of the *E. coli* Lon. Guided by data showing that glutamine is the most common replacement for a lysine in the sequences of naturally occurring proteins [40] and assuming that such a replacement is unlikely to affect gross structure of the protein while changing the charge of the residue, we mutated Lys722 to glutamine. This mutation did not change such properties of the protein as solubility, although the small amount of the expressed protein precluded its detailed structural characterization.

The mutant K722Q completely lost its hydrolytic activity for the protein (β-casein) and the small thioester (Suc-Phe-Leu-Phe-SBzl) substrates, despite the presence of ATP and magnesium ions in the reaction mixture (Table 1). The K722Q mutant has similar properties to the S679A mutant, shown previously to be proteolytically inactive [20] (Table 1). These results emphasize the important role played by Lys722 in the activity of Lon and, together with the sequence alignment data for the Lon family, can be used to infer the presence of a functional Ser–Lys dyad in the proteolytic site.

The crystal structure of the proteolytic domain of *E. coli* Lon provided the final verification of the existence of the Ser–Lys dyad. Ala679, which replaced Ser679 in the inactive mutant that was the subject of the crystallographic analysis, was located in the immediate vicinity of Lys722, with no other potential catalytic chains nearby [30]. A model of the active enzyme could be easily deduced [30], and its analysis showed that the two residues of the putative catalytic dyad could make hydrogen-bonded contacts without any rearrangements of their vicinity. We have recently determined the structure of the proteolytic domain of wild-type Lon, which does not exhibit any gross conformational changes compared with the mutant (I. Botos, unpublished data). Thus sequence analysis, site-directed mutagenesis, and crystal structure all independently support the presence of a Ser–Lys catalytic dyad in the active site of Lon protease.

The tertiary structure of the Lon proteolytic domain also represented a unique, previously unreported protein fold. Based on these observations, the *E. coli* Lon protease became the founding member of a newly introduced clan SJ in the MEROPS classification of proteolytic enzymes [41].

Identification and structural characteristics of two Lon subfamilies

In the majority of Lon proteases the residues immediately adjacent to the catalytic Ser are located in the previously described conserved fragment PKDGP_S_AG [20]. New extensive sequence analysis of the Lon protease family reveals significant differences in the 72-residue-long consensus fragments that include the catalytic Ser and Lys residues (Fig. 2). A different consensus sequence, XΦ(E/D)GD_S_A(S/T) (Φ = hydrophobic amino acid), was found in some other members of the family [29]. The two template sequences described above have corresponding consensus sequences around the catalytic Lys722: (K/R)X_K_XΦ and (T/N)X_K_ΦE, respectively. Based on this, we can suggest a division of the Lon protease family into two subfamilies: LonA and LonB.

In LonA subfamily these 72-residue fragments contain 21 strictly conserved residues, whereas 18 residues are conserved in the equivalent fragments of LonB subfamily. Only 11 residues remain conserved between the two
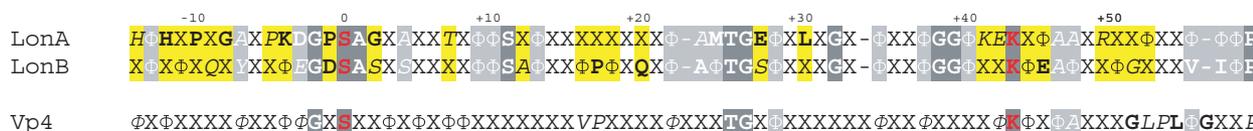
Fig. 2. Consensus sequences for fragments of LonA, LonB, and Vp4 proteases that include the catalytically active Ser and Lys residues. Catalytically active Ser (position 0) and Lys (position +43) residues are marked in red. Strictly conserved residues are in bold; residues conserved in > 90% of the sequences are shown in italics. Residues conserved in both Lon subfamilies are highlighted in dark gray, whereas similar residues are highlighted in gray and different residues in yellow. Residues present in the sequence of Vp4 that are conserved or similar to the corresponding residues in the Lon family are also highlighted. Residues marked by X may represent deletions in the structure of Vp4 only.

subfamilies. In addition to the catalytic Ser and Lys residues, these 11 residues include: Gly, preceding, and Ala, following the catalytic Ser (positions −2 and +1, respectively), as well as Ser (+11), Thr (+25), four Gly residues (+26, +32, +38 and +39), and Pro (+58) (Fig. 2). Moreover, similar residues were found in another 18 positions; thus, the overall combined identity and similarity for this fragment is about 40%. The residue variation in 26 of the remaining 43 positions of the 72-residue fragment (Fig. 2, residues marked in yellow) may lead to significant differences in the architecture of the proteolytic sites of the two subfamilies.

The most significant difference between the two sub-families is the presence of 10 strictly conserved residues specific only to the LonA subfamily (positions −12, −10, −8, −4, −3, −1, +2, +24, +27, and +30) and five conserved residues found only in the LonB subfamily (positions −1, +17, +20, +23 and +45) (Fig. 2). Substitutions close to the catalytically active residues [Pro → Asp (position −1), Lys → hydrophobic amino acid (position −4), and hydrophobic amino acid → Glu (position +45)] might lead to differences in the activity and specificity towards peptide substrates of these two subfamilies of Lon proteases.

Division of the Lon family into two subfamilies, based primarily on the characteristics of their catalytic sites, is in agreement with the differences in the respective consensus sequences of their AAA$^+$ modules. In the LonA subfamily, the Walker A and B motifs are located in the conserved fragments GPPGVGKTS and PΦ$_4$DEIDK, whereas in the LonB subfamily these motifs are represented by the sequences GXPGXGKSΦ and GΦ$_4$DEIXX, respectively. The sequences in the vicinity of the conserved sensor-1,

arginine finger, and sensor-2 residues (Asn473, Arg484, and Arg542 in *E. coli* LonA protease) are also notably different in LonA and LonB proteases. The other very important differences between the two subfamilies of Lon proteases are the absence of N-terminal domain and the presence of transmembrane fragment in LonB proteases (Fig. 3; also see below).

Evolutionary classification and structural variation of Lon subfamilies

According to the evolutionary classification of the AAA$^+$ ATPases [7,9], Lon family belongs to the HslU/ClpX/Lon/ClpAB-C clade and consists of two distinct branches, bacterial and archaeal Lon, on the basis of the differences in their AAA$^+$ modules. Our assignment of the two subfamilies agrees with both the above and the *MEROPS* [41] classification of Lon family proteases that is based on differences between their proteolytic domains.

The LonA subfamily consists mainly of bacterial and eukaryotic enzymes (*MEROPS*, clan SJ, ID: S16.001–16.004, S16.006 and partially S16.00X, Table 2), accounting for > 80% of the presently known Lon proteases. The LonA subfamily members mimic the 'classical' Lon protease from *E. coli* and they all contain the N and P domains that flank the AAA$^+$ module (Fig. 3). The overall length of LonA proteases ranges from 772 (*Oceanobacillus iheyensis*) to 1133 (*Saccharomyces cerevisiae*) amino acid residues (Table 2). The N domains are found to be the most variable, both in their length (220–510 amino acids) and in their amino acid sequences. The P domains of LonA proteases have similar lengths (188–224 amino acids) and are highly
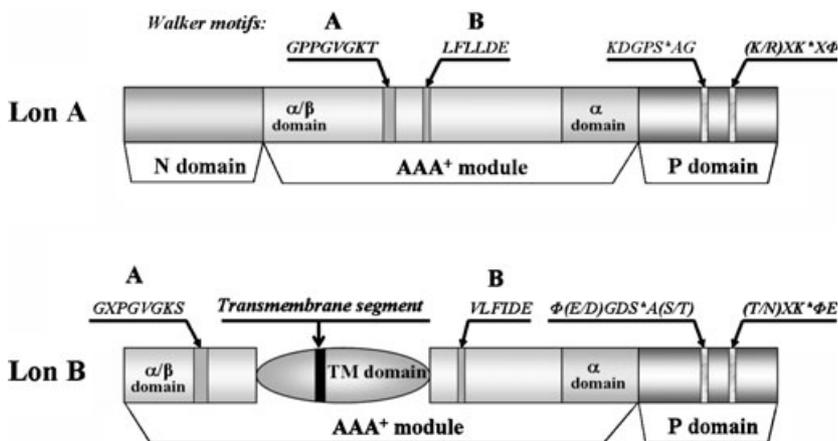
Fig. 3. Schematic representation of the LonA and LonB subfamilies outlining the domain structures with the important consensus sequences. See text for the definition of the domains. The locations and sequences of the Walker A and B motifs (AAA$^+$ module) and of fragments of the proteolytic domains including catalytically active serine (**S***) and lysine (**K***) residues are marked. The intein insertions that might be located just after the TM domains in some LonB proteases are not shown.

Table 2. Comparison of the sizes of LonA and LonB subunits and their putative domains. The sizes result from data obtained by limited proteolysis of *E. coli* LonA protease by chymotrypsin. Nucleotide-binding (NB) domains contain Walker A and B motifs, as well as SRH motif bearing sensor-1 and Arg finger residues. For LonA, NB domain corresponds to α/β domain; for LonB, NB domain is conventionally represented by two parts and corresponds to the α/β domain without transmembrane (TM) domain and intein. Differences in NB domain sizes are mostly due to the differences of their N-terminal fragments.

Subfamily	Representative number	MEROPS classification S 16	Representative number	Number of amino acid residues							
				N domain	AAA+ module — α/β domain as a whole				α domain	P domain	Total in subunits
					NB domain	TM domain	Intein	Total			
LonA	80	001	52	230–260	255–278	–	–	255–278	88–126	188–224	772–848
		002	14	249–510	252–260	–	–	252–260	93–175	193–221	819–1133
		003	4	285–286	258	–	–	258	137–140	191–205	875–888
		004	3	244–257	256–257	–	–	256–257	93–97	188–194	791–795
		006	2	0; 253	239; 257	–	–	239; 257	133	209	581; 852
		00X	5	220–445	254–267	–	–	254–267	94–143	188–217	779–1063
LonB	21	005	3	–	186–203	112	333–474	655–786	118	211–232	998–1127
		005	11	–	181–260	108–128	–	305–375	118 (139)[a]	205–231	621–702
		00X	7	?	?	?	–	?	?	233–261	586–817

[a] The α domain size of LonB protease from *Methanocaldococcus jannaschii* is listed in parentheses.

homologous. LonA AAA+ modules show very high homology for their nucleotide binding α/β domains, whereas their α-helical domains vary significantly due to C-terminal insertions or extensions (Table 2).

ATP-dependent enzymes from the LonB subfamily (< 20% of known Lon proteases) are found only in archaebacteria (*MEROPS*, ID: S16.005). LonB-like proteins with homologous proteolytic domains but no clearly defined AAA+ domains are also found in other bacteria (ID: S16.00X, partially). The subunit architecture of archaeal LonB proteases is significantly different from that of LonA proteases. LonB enzymes (621–1127 amino acids) consist of AAA+ modules and proteolytic domains (205–232 amino acids), but lack the N (LAN) domains [7,42]. These proteins are membrane bound via one or two potential transmembrane (TM) segments that may be part of additional TM domains. The putative TM domains are inserted within the nucleotide-binding domains (α/β), between the Walker A and B motifs (Fig. 3). Thus, the architecture of the LonB AAA+ module is similar to the HslU subunit of HslUV protease with an insertion domain (I domain) between its Walker motifs [43]. We have noticed that some *lonB* genes (e.g. from *Pyrococcus* sp.) contain self-splicing elements that encode polypeptides (inteins, 333–474 amino acids), also located between the Walker A and B motifs and following the TM domains. The α domain of archaeal LonB proteases typically consists of 118 residues, except for *Methanocaldococcus jannaschii* LonB, which has 139 residues in its α domain. Archaeal LonB proteases are highly homologous except for their transmembrane segments.

The first membrane-bound LonB protease to be purified was recently isolated from *Thermococcus kodakarensis* [44]. LonB proteases are expected to bear the functions of the only bacterial membrane-bound ATP-dependent protease, FtsH (*MEROPS*, ID: M41.001), because the latter enzymes are not present in Archaea [42]. However, one should not postulate that Archaea contain solely LonB proteases, because the *Methanosarcinacae* genomes are known to encode both LonA and LonB proteases. A number of bacterial genomes (e.g., *E. coli*, *Thermotoga maritima*, *Vibrio cholerae*) encode not only LonA proteases, but also LonB-like proteases. The P domains of the latter (232–260 amino acids) are highly homologous to archaeal LonB P domains. However, the canonical conserved fragments such as sensor-1, sensor-2, and Walker motifs are not found in the sequence fragments (340–557 amino acids) that precede their P domains, raising a possibility that these are not ATP-dependent enzymes. Thus, the metabolic role and biochemical specificity of these bacterial LonB-like proteases are still obscure.

Lon-like proteases

Birnavirus Vp4 proteases, which are included in the *MEROPS* database as a separate family (S50) in the SJ clan, and some other proteins that lack AAA+ modules and are present in the genomes of Archaea and *Caenorhabditis elegans*, have been identified as having proteolytic fragments homologous with Lon proteases [27]. It was pointed out that a common core, composed of ≈ 80 amino acids conserved across Lon/Vp4 proteases [27], includes six

invariant residues: Gly677, Ser679, Thr704, Gly705, Lys722 and Pro737 of *E. coli* LonA (positions −2, 0, +25, +26, +43 and +58 in Fig. 2). However, we note that a series of residues conserved in LonA and LonB subfamilies are altered in Lon-like protein fragments, including the vicinity of the catalytic Ser and Lys residues (Fig. 2). In particular, in contrast to Lon family proteases, Lon-like enzymes have a number of different residues in positions (−1) and (+1) relative to the catalytic Ser, and there is a 37–43-residue variable spacing between their catalytic Ser and Lys residues. The above-mentioned differences make it clear that Lon-like proteases cannot be characterized as clearly belonging to either the LonA or LonB subfamilies.

Residue conservation in LonA and LonB subfamilies

Although several residues are conserved between LonA and LonB subfamilies, only those that were identified by us either on the basis of mutagenesis experiments or the crystal structures to be significant for the function will be discussed below. The *E. coli* LonA protease has been previously characterized as a sulfhydryl-dependent enzyme [17]. Each of its subunits contains six cysteine residues: one located in the N domain, one in each of the α/β and α domains of the AAA$^+$ module, and three in the P domain. The majority of LonA proteases contain between 1 and 11 Cys residues, although ≈ 2% of these proteases do not have any cysteines at all. The most highly conserved Cys residue is present in > 90% of LonA proteases. It is located in the α/β domain, on the P loop preceding the Walker A motif. Sequence alignment suggests that < 10% of LonA proteases may contain a disulfide bond equivalent to Cys617–Cys691, identified in the structure of the *E. coli* Lon protease P domain [30]. This is a very unusual, surface-exposed disulfide bond, and it is still unclear to what extent its presence might influence the structure and function of LonA.

Archaeal LonB proteases contain a total of one to six cysteine residues (not taking into account the Cys residues of inteins), and more than half of these enzymes do not contain any Cys residues in their P domains. The only strictly conserved cysteine is located in the C terminal part of the α/β domain following the Walker B motif. Bacterial LonB enzymes have between 2 and 10 Cys residues. However, none of the Cys residues conserved within the LonA or LonB subfamily are conserved across the entire Lon family.

Several residues conserved in both subfamilies of Lon proteases have either structural or functional importance. For example, the conserved Gly677 (located at position −2 with respect to the catalytic Ser) is also present in a vast majority of serine proteases, utilizing either a catalytic triad or a dyad in their active sites. The torsion angles of this residue are unusual and accessible only to a glycine, thus imposing a conformation of the main chain for a stretch of residues that are involved in the interactions with the substrate. A similar role may also be assigned to that residue in Lon proteases.

Tyr493, located at the N-terminus of the α domain of *E. coli* Lon, may also play an important role in both the LonA and LonB subfamilies. We have previously found that the phenylalanine substitution leads to a 2.5-fold increase in the ATPase activity of the mutant LonA, making it as active as the wild-type enzyme activated by protein substrate [45]. This result, as well as the analysis of the three-dimensional structure of the α domain of *E. coli* Lon [46], suggest that Tyr493 may participate both in the transfer of a conformational change signal from the ATPase site to the proteolytic site and also in interaction with bound nucleotides.

Conclusions

This analysis of the available Lon sequences suggested that: (a) the hypothesis about the absence of the classical catalytic triad Ser–His–Asp in their active sites [25] is correct; (b) the conserved Lys residue is a member of the catalytic Ser–Lys dyad; and (c) two Lon subfamilies, named LonA and LonB, can be identified. LonA, LonB, and Lon-like proteases exhibit different proteolytic site sequences, although only two clearly identifiable motifs are inherent in true ATP-dependent Lon proteases. Further structural studies of other Lon family members are necessary in order to clarify the relationship between their different architecture and function.

Acknowledgements

References

1. Wickner, S., Maurizi, M.R. & Gottesman, S. (1999) Posttranslational quality control: folding, refolding, and degrading proteins. *Science* **286**, 1888–1893.
2. Goldberg, A.L. (1992) The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. *Eur. J. Biochem.* **203**, 12029–12034.
3. Gottesman, S. & Maurizi, M.R. (1992) Regulation by proteolysis: energy-dependent proteases and their targets. *Microbiol. Rev.* **56**, 592–621.
4. Gottesman, S. (1996) Proteases and their targets in *Escherichia coli. Annu. Rev. Genet.* **30**, 465–506.
5. Maurizi, M.R. (1992) Proteases and protein degradation in *Escherichia coli. Experientia* **48**, 178–201.
6. Neuwald, A.F., Aravind, L., Spouge, J.L. & Koonin, E.V. (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43.
7. Iyer, L.M., Leipe, D.D., Koonin, E.V. & Aravind, L. (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146**, 11–31.
8. Ogura, T. & Wilkinson, A.J. (2001) AAA+ superfamily ATPases: common structure – diverse function. *Genes Cells* **6**, 575–597.
9. Lupas, A.N. & Martin, J. (2002) AAA proteins. *Curr. Opin. Struct. Biol.* **12**, 746–753.
10. Maurizi, M.R. & Li, C.C.H. (2001) AAA proteins: in search of a common molecular basis. *EMBO Report* **2**, 980–985.
11. Maupin-Furlow, J.A., Wilson, H.L., Kaczowka, S.J. & Ou, M.S. (2000) Proteasomes in the archaea: from structure to function. *Front. Biosci.* **5**, D837–D865.
12. Patel, S. & Latterich, M. (1998) The AAA team: related ATPases with diverse functions. *Trends Cell. Biol.* **8**, 65–71.
13. Langer, T. (2000) AAA proteases: cellular machines for degrading membrane proteins. *Trends Biochem. Sci.* **25**, 247–251.

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

15. Guo, F., Maurizi, M., Esser, L. & Xia, D. (2002) Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease. *J. Biol. Chem.* **277**, 46743–46752.

16. Swamy, K.H. & Goldberg, A.L. (1981) *E. coli* contains eight soluble proteolytic activities, one being ATP dependent. *Nature* **292**, 652.

17. Goldberg, A.L., Moerschell, R.P., Chung, C.H. & Maurizi, M.R. (1994) ATP-dependent protease La (lon) from *Escherichia coli*. *Methods Enzymol.* **244**, 350–375.

18. Amerik, A.Yu, Chistyakova, L.G., Ostroumova, N.I., Gurevich, A.I. & Antonov, V.K. (1988) Cloning, expression and structure of the functionally active shortened *lon* gene in *Escherichia coli*. *Bioorg. Khim.* **14**, 408–411.

19. Amerik, A.Yu, Antonov, V.K., Ostroumova, N.I., Rotanova, T.V. & Chistyakova, L.G. (1990) Cloning, structure and expression of the full-size *lon* gene in *Escherichia coli* coding for ATP-dependent La-proteinase. *Bioorg. Khim.* **16**, 869–880.

20. Amerik, A.Yu, Antonov, V.K., Gorbalenya, A.E., Kotova, S.A., Rotanova, T.V. & Shimbarevich, E.V. (1991) Site-directed mutagenesis of La protease. A catalytically active serine residue. *FEBS Lett.* **287**, 211–214.

21. Ebel, W., Skinner, M.M., Dierksen, K.P., Scott, J.M. & Trempy, J.E. (1999) A conserved domain in *Escherichia coli* Lon protease is involved in substrate discriminator activity. *J. Bacteriol.* **181**, 2236–2243.

22. Frickey, T. & Lupas, A.N. (2004) Phylogenetic analysis of AAA proteins. *J. Struct. Biol.* **146**, 2–10.

23. Mogk, A., Dougan, D., Weibezahn, J., Schlieker, C., Turgay, K. & Bukau, B. (2004) Broad yet high substrate specificity: the challenge of AAA+ proteins. *J. Struct. Biol.* **146**, 90–98.

24. Rotanova, T.V. (1999) Structural and functional characteristics of ATP-dependent Lon protease from *Escherichia coli*. *Bioorgan. Khim.* **25**, 883–891.

25. Starkova, N.N., Koroleva, E.P., Rumsh, L.D., Ginodman, L.M. & Rotanova, T.V. (1998) Mutations in the proteolytic domain of *Escherichia coli* protease Lon impair the ATPase activity of the enzyme. *FEBS Lett.* **422**, 218–220.

26. Paetzel, M. & Dalbey, R.E. (1997) Catalytic hydroxyl/amine dyads within serine proteases. *Trends Biochem. Sci.* **22**, 28–31.

27. Birghan, C., Mundt, E. & Gorbalenya, A.E. (2000) 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**, 114–123.

28. Lejal, N., Da Costa, B., Huet, J.C. & Delmas, B. (2000) Role of Ser-652 and Lys-692 in the protease activity of infectious bursal disease virus VP4 and identification of its substrate cleavage sites. *J. General Virol.* **81**, 983–992.

29. Rotanova, T.V., Melnikov, E.E. & Tsirulnikov, K.B. (2003) A catalytic Ser – Lys dyad in the active site of the ATP-dependent Lon protease from *Escherichia coli*. *Bioorgan. Khim.* **29**, 97–99.

30. Botos, I., Melnikov, E.E., Cherry, S., Tropea, J.E., Khalatova, A.G., Rasulova, F., Dauter, Z., Maurizi, M.R., Rotanova, T.V., Wlodawer, A. & Gustchina, A. (2004) The catalytic domain of *Escherichia coli* Lon protease has a unique fold and a Ser-Lys dyad in the active site. *J. Biol. Chem.* **279**, 8140–8148.

31. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

32. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59.

33. Rotanova, T.V. & Kotova, S.A. (1994) Amerik, A.Yu., Lykov, I.P., Ginodman, L.M. & Antonov, V.K. ATP-dependent proteinase La from *Escherichia coli*. *Bioorgan. Khim.* **20**, 114–125.

34. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.

35. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

36. Melnikov, E.E., Tsirulnikov, K.B., Rasulova, F.S., Ginodman, L.M. & Rotanova, T.V. (1998) Suc-Phe-Leu-Phe-SBzl, a new substrate for functional study of *Escherichia coli* ATP-dependent Lon-proteinase and its modified forms. *Bioorgan. Khim.* **24**, 638–640.

37. Melnikov, E.E., Tsirulnikov, K.B. & Rotanova, T.V. (2001) Coupling of proteolysis and hydrolysis of ATP upon functioning of Lon proteinase of *Escherichia coli*. II. Hydrolysis of ATP and activity of peptide hydrolase sites of the enzyme. *Bioorgan. Khim.* **27**, 120–129.

38. Bencini, D.A., Wild, J.R. & O'Donovan, G.A. (1983) Linear one-step assay for the determination of orthophosphate. *Anal. Biochem.* **132**, 254–258.

39. Melnikov, E.E., Tsirulnikov, K.B. & Rotanova, T.V. (2000) Coupling of proteolysis with ATP hydrolysis by *Escherichia coli* Lon proteinase. I. Kinetic aspects of ATP hydrolysis. *Bioorgan. Khim.* **26**, 530–538.

40. Dayhoff, M.O. (1972) Atlas of protein sequence and structure. Natl. Biom. Res. Found. Washington DC.

41. Barrett, A.J., Rawlings, N.D. & O'Brien, E.A. (2001) The *MEROPS* database as a protease information system. *J. Struct. Biol.* **134**, 95–102.

42. Ward, D.E., Shockley, K.R., Chang, L.S., Levy, R.D., Michel, J.K., Conners, S.B. & Kelly, R.M. (2002) Proteolysis in hyperthermophilic microorganisms. *Archaea* **1**, 63–74.

43. Dougan, D.A., Mogk, A. & Bukau, B. (2002) Protein folding and degradation in bacteria: to degrade or not to degrade? That is the question. *Cell. Mol. Life Sci.* **59**, 1607–1616.

44. Fukui, T., Eguchi, T., Atomi, H. & Imanaka, T. (2002) A membrane-bound archaeal Lon protease displays ATP-independent proteolytic activity towards unfolded proteins and ATP-dependent activity for folded proteins. *J. Bacteriol.* **184**, 3689–3698.

45. Melnikov, E.E., Tsirulnikov, K.B., Ginodman, L.M. & Rotanova, T.V. (1998) *In vitro* coupling of ATP hydrolysis to proteolysis of ATP site mutant forms of Lon proteinase from. *E. Coli. Bioorg. Khim.* **24**, 293–299.

46. Botos, I., Melnikov, E.E., Cherry, S., Khalatova, A.G., Rasulova, F.S., Tropea, J.E., Maurizi, M.R., Rotanova, T.V., Gustchina, A. & Wlodawer, A. (2004) Crystal structure of the AAA+ α domain of *E. coli* Lon protease at 1.9 Å resolution. *J. Struct. Biol.* **146**, 113–122.

47. Christopher, J.A. (1998) *SPOCK: the Structural Properties Observation and Calculation Kit*. The Center for Macromolecular Design, Texas A & M University, College Station, TX.