The Active Site of a Lon Protease from *Methanococcus jannaschii* Distinctly Differs from the Canonical Catalytic Dyad of Lon Proteases*

Young Jun Im‡, Young Na‡, Gil Bu Kang‡, Seong-Hwan Rho‡, Mun-Kyoung Kim‡, Jun Hyuck Lee‡, Chin Ha Chung§, and Soo Hyun Eom‡¶

From the ‡Department of Life Science, Gwangju Institute of Science and Technology, Gwangju 500-712 and the §National Research Laboratory of Protein Biochemistry, School of Biological Sciences, Seoul National University, Seoul 151-742, Korea

ATP-dependent Lon proteases catalyze the degradation of various regulatory proteins and abnormal proteins within cells. *Methanococcus jannaschii* Lon (Mj-Lon) is a homologue of *Escherichia coli* Lon (Ec-Lon) but has two transmembrane helices within its N-terminal ATPase domain. We solved the crystal structure of the proteolytic domain of Mj-Lon using multiwavelength anomalous dispersion, refining it to 1.9-Å resolution. The structure displays an overall fold conserved in the proteolytic domain of Ec-Lon; however, the active site shows uniquely configured catalytic Ser-Lys-Asp residues that are not seen in Ec-Lon, which contains a catalytic dyad. In Mj-Lon, the C-terminal half of the $\beta4$-$\alpha2$ segment is an α-helix, whereas it is a β-strand in Ec-Lon. Consequently, the configurations of the active sites differ due to the formation of a salt bridge between Asp-547 and Lys-593 in Mj-Lon. Moreover, unlike Ec-Lon, Mj-Lon has a buried cavity in the region of the active site containing three water molecules, one of which is hydrogen-bonded to catalytic Ser-550. The geometry and environment of the active site residues in Mj-Lon suggest that the charged Lys-593 assists in lowering the pK_a of the Ser-550 hydroxyl group via its electrostatic potential, and the water in the cavity acts as a proton acceptor during catalysis. Extensive sequence alignment and comparison of the structures of the proteolytic domains clearly indicate that Lon proteases can be classified into two groups depending on active site configuration and the presence of DGPSA or (D/E)GDSA consensus sequences, as represented by Ec-Lon and Mj-Lon.

In all cells, energy-dependent proteolysis plays a key role in the rapid turnover of short-lived regulatory proteins and in the elimination of defective and denatured proteins (1). Bacterial cells possess a number of ATP-dependent proteases, which are complex enzymes containing both ATPase and proteolytic activity as separate domains within a single polypeptide or as individual subunits within complex assemblies. *Escherichia coli*, for example, express five different ATP-dependent proteases: Lon, ClpAP, ClpXP, HslUV (ClpYQ), and FtSH (2). Homologous proteases have also been identified in archaea and eubacteria, as well as in numerous eukaryotes. Some archaeal Lons have one or two putative transmembrane regions, suggesting that they are membrane-associated (3). The proteolytic components of ATP-dependent proteases include several different types of active sites. For instance, ClpP is a classical serine protease (4), whereas HslV has a catalytic N-terminal Thr residue (5).

Lon was the first ATP-dependent protease to be described (6). Similar to the molecular chaperon, Lon recognizes a broad range of proteins and mediates their turnover of abnormal and short-lived normal proteins. Indeed, through degradation of various specialized proteins, Lon is involved in the regulation of a number of biological functions (7). Moreover, it also reportedly acts as a DNA-binding protein, influencing the regulation of DNA replication and gene expression (8). *E. coli* Lon (Ec-Lon)[1] is an 87-kDa protein containing N-terminal ATPase and C-terminal protease domains on a single polypeptide chain that is found as a homooligomer (9), although the oligomeric states of intact Lon were not reported consistently. Still, recent observations suggest that Lon isoforms from several sources self-associate into hexameric or heptameric rings (10–12). Sequence comparisons suggest that Lon contains a catalytic Ser-Lys dyad, and the first crystal structure of the proteolytic domain from Ec-Lon confirmed the presence of a catalytic Ser-Lys dyad within a unique structural fold, distinct from that of the classical serine proteases (13, 14). In the present study, however, we found that the Mj-Lon proteolytic domain employs a unique catalytic Ser-Lys-Asp triad. Extensive sequence alignment and comparison of the structures of their proteolytic domains clearly indicate that Lon proteases can be classified into two groups depending on the configuration of the catalytic residues in the active site, as represented by Ec-Lon and Mj-Lon.

MATERIALS AND METHODS

Protein Expression and Purification—Full-length Lon protease was cloned by PCR using a *Methanococcus jannaschii* genomic library as a template. The resultant product was cloned into the NcoI and XhoI sites of pET28b vector (Novagen). The coding regions for the proteolytic domain of Mj-Lon (residues 456–649) were then amplified from a plasmid harboring the full-length *lon* gene and cloned into the NdeI and XhoI sites of pET28b. The resultant construct provides for an N-terminal His$_6$ tag separated from the protein by a thrombin protease recognition site (LVP(R/G)S). The fusion protein was overexpressed in *E. coli*

[1] The abbreviations used are: Ec-Lon, *E. coli* Lon; Mj-Lon, *M. jannaschii* Lon; MES, 4-morpholine ethane sulfonic acid; Se-Met, selenomethionine; r.m.s., root mean square.

TABLE I
Data collection and refinement statistics

Data set	SeMet-MAD			Native
Crystal form	$P4_22_12$			$P4_22_12$
Cell parameter a, c (Å)	92.16, 101.68			89.85, 100.61
X-ray source	PF BL-18B			PF AR-NW12
Wavelength	$\lambda 1 = 0.9780$	$\lambda 2 = 0.9785$	$\lambda 3 = 0.9940$	1.000
Resolution (Å)	50–2.3			50–1.9
Mean $I/\sigma(I)$	27.6 (7.6)	27.8 (7.8)	28.0 (7.6)	29.8 (3.1)
R_{sym}[a] (%)	9.6 (34.3)	9.3 (33.4)	9.3 (34.1)	5.3 (35.7)
Completeness [b] (%)	98.9 (98.6)	98.9 (98.8)	98.9 (97.8)	99.9 (98.4)
Phasing statistics				
Number of selenium site	2/388 amino acids			
Mean FOM (SOLVE)	0.40			
Overall FOM (RESOLVE)	0.56			
Refinement				
R_{cryst}[c] total (%)				22.6
R_{free}[d] total (%)				26.3
r.m.s. bond length (Å)				0.005
r.m.s. bond angle (°)				1.3
Average B value (Å^2)				25.7

[a] $R_{sym} = \sum |\langle I \rangle - I|/\sum \langle I \rangle$.
[b] Values in parentheses relate to the highest resolution shell.
[c] $R_{cryst} = \sum \|F_o| - |F_c\|/\sum |F_o|$.
[d] R_{free} calculated with 5% of all reflections excluded from the refinement stages.

BL21(DE3). The cells were grown in LB at 37 °C to an A_{600} ~0.8, at which time expression was induced using 1 mM isopropyl-1-thio-β-D-galactopyranoside. Growth was continued at 30 °C for 6 h, after which the cells were harvested, resuspended in binding buffer (500 mM NaCl and 50 mM NaH_2PO_4 (pH 7.5)), and broken by sonication. The lysate was then run on an immobilized nickel affinity column, after which the column was washed with binding buffer, and the fusion protein was eluted in buffer comprised of the binding buffer plus 500 mM imidazole. After concentration, the tag was cleaved from the protein by treating it with thrombin protease at room temperature overnight, and the cleaved protein was subjected to size exclusion chromatography on a Superdex 200 column (Amersham Biosciences) equilibrated with 10 mM Tris-HCl (pH 8.0) plus 50 mM NaCl. The fractions containing the recombinant protein were pooled and concentrated to 9 mg/ml. To prepare the Se-Met-enriched protein, the protease domain of Mj-Lon was expressed in *E. coli* BL21(DE3) using saturation of the Met biosynthetic pathways protocols due to poor growth of the Met auxotroph strain *E. coli* B834(DE3) in supplemented M9 medium (15).

Crystallization—The proteolytic domain of Mj-Lon was crystallized at room temperature (20 ± 1 °C) using the hanging drop vapor diffusion method. Initial seeds of multiple twinned crystals were grown on a siliconized coverslip by equilibrating a mixture containing 2 μl of protein solution (4.5 mg/ml protein in 50 mM NaCl and 10 mM MES-NaOH (pH 8.0)) and an equal volume of well solution (2.4 M ammonium sulfate and 100 mM MES-NaOH (pH 7.0)) against 1.0 ml of well solution. Under the same crystallization conditions, crystals were grown by microseeding with or without subsequent macroseeding to increase the size of single crystals. Within 1 week, single crystals grew to dimensions of $0.1 \times 0.2 \times 1.0$ mm and were flash-frozen by direct transfer to Paratone-N cryoprotectant solution (Hampton Research).

Crystallographic Analysis—Se-Met multiwavelength anomalous diffraction data (Table I) were collected to 2.3-Å resolution from a single frozen crystal with an ADSC Quantum 4R CCD detector at beamline BL-18B at the Photon Factory, Tsukuba, Japan. The data set was processed and scaled using HKL2000 packages (16) and then handled with the CCP4 program suite (17). A second native data set from a Se-Met-labeled crystal was collected to 1.9-Å resolution with an ADSC Quantum Q210 CCD detector at beamline AR-NW12. Multiwavelength anomalous diffraction phasing was carried out using the programs SOLVE and RESOLVE (18). Despite the low proportion of anomalous scatter and the presence of only two Se-Mets among 388 residues, the resultant electron density map, showing a dimer in an asymmetric unit, was readily interpretable. Automatic model building was then carried out using the program MAID (19), with which 88% of the structure was modeled. A partial model containing 342 residues was refined using the program CNS (20) for several cycles of B-factor refinement and simulated annealing. Model phases from the partial model were then applied to the 1.9-Å data sets. Subsequent density modification of CNS yielded an electron density map of excellent quality showing a clear trace of all the residues except for 9 residues at the C termini. The remainder of the model was built manually into the density-modified map using the

program O (21). The refinement of the native structure was completed with CNS to a final crystallographic R-factor of 22.6% and an R_{free} of 26.3%. The final model covers residues 456–640. The stereochemistry of the model was analyzed using PROCHECK (22), which showed no residue to be in a disallowed region.

RESULTS AND DISCUSSION

Description of the Structure—Like other Lon proteases from various sources, the 71.9-kDa Mj-Lon contains an N-terminal ATPase domain and a C-terminal protease domain. The ATPase domain of Mj-Lon containing two transmembrane helices shows low overall similarity to its bacterial and eukaryotic counterparts. As compared with the ATPase domains, the proteolytic domains are well conserved over all Lon families with 40% of the minimal pairwise homology. The proteolytic domain of Mj-Lon shares 29% identity and 49% similarity with the proteolytic domain of Ec-Lon over 193 amino acids. In the present study, we determined the crystal structure of the proteolytic domain of Mj-Lon (residues 456–649) using multiwavelength anomalous dispersion and refined it to 1.9-Å resolution. The structure of the proteolytic domain consists of five α-helices and nine β-strands (Fig. 1A). The N-terminal β1 strand and antiparallel β2 strand form a long β-hairpin loop. The parallel β3 and β4 strands, which are connected by the longest helix (α1), form the first large β-sheet with the β1 and β2 strands. The subsequent helix α2 is kinked at Ser-550, which is a catalytic residue in this enzyme. Helices α1 and α2 interact with the first β-sheet, forming a compact N-terminal subdomain. The other subdomain, which is connected by a random loop (α2-β5), is composed of three parallel β-strands (β5, β8, and β9) surrounded by three α-helices. A β-hairpin loop composed of strands β6 and β7 connects strand β5 with helix α3, and subsequent strands and helices alternate along the primary structure.

The crystal structure of Mj-Lon showed that each asymmetric unit contains two molecules forming a dimer with a noncrystallographic 2-fold axis, presumably corresponding to the dimer in solution during purification. The dimeric structure of the proteolytic domain of Mj-Lon resembles a pair of lungs when viewed from the vertical with respect to the 2-fold axis (Fig. 1B). The dimer interface consists of a β-hairpin loop connecting strand β5, helices α3 and α4, the C-terminal end of helix α3, and 5 residues in the N-terminal subdomain from each monomer. The surface area buried upon dimer formation
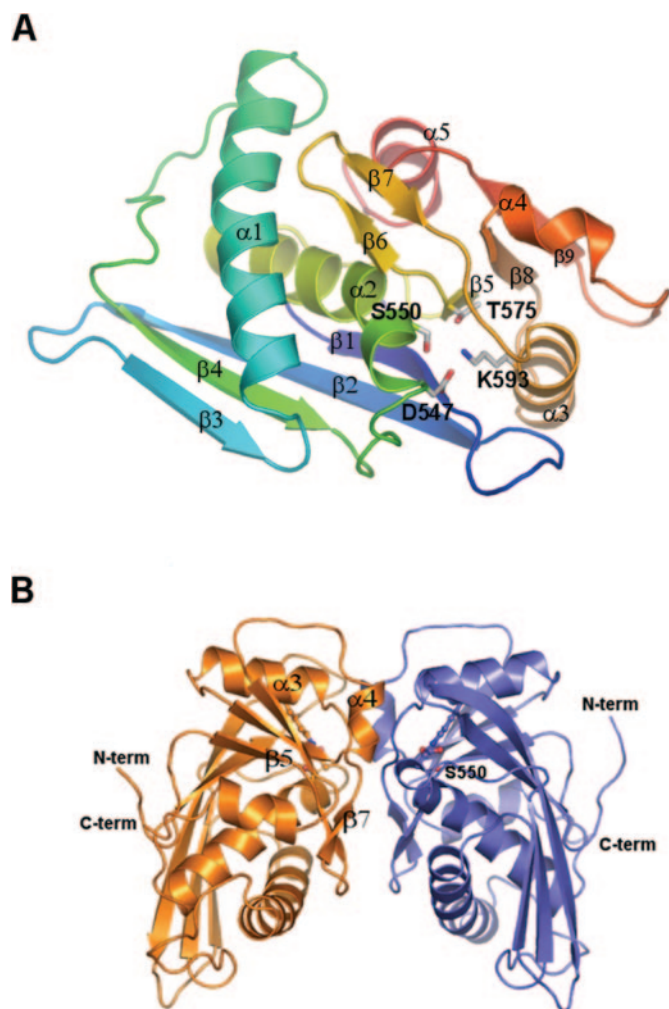
FIG. 1. **Structure of the proteolytic domain of *M. jannaschii* Lon.** *A*, ribbon diagram with the catalytic residues shown in *ball-and-stick*. *B*, ribbon diagram of the dimeric structure of Mj-Lon proteolytic domain observed in the asymmetric unit of the crystal. The catalytic residues are shown in *ball-and-stick*. *N-term*, N terminus; *C-term*, C terminus.

is 655 Å^2, or 7.5% of the total surface area of each monomer. The interface is composed of 64.0% nonpolar atoms and 35.0% polar atoms, and within the dimer, the monomers are held together by two hydrogen bonds and numerous van der Waals interactions. It is questionable, however, whether the crystal structure of the dimer is consistent with the assembly of proteolytic domains in the intact Lon protease. This is because the active sites of the two monomers are both completely buried within the interface, preventing exposure of catalytic residues to the solvent unless the interface undergoes a marked conformational change. Considering that there are hydrophobic patches around active sites, it may be that dimerization of these proteolytic domains through hydrophobic interactions occurs in the absence of the ATPase domains. Also, given that obstruction of the active sites of the Mj-Lon proteolytic domains by dimerization is not reasonable *in vivo*, it seems likely that the observed dimeric structure is artifactual. Therefore, despite the presence of intact active site in the proteolytic domain, the dimerization might lead to loss of proteolytic activity. The truncation of N-terminal domain in *Brevibacillus thermoruber* Lon protease resulted in the failure of oligomerization and led to the inactivation of proteolytic and ATPase activities, indicating that N-terminal domain is essential for the correct oligomerization of the protein (23). Intact Lon protease is report-

edly active as a homooligomer comprised of 4–8 subunits. Lon proteases from several sources, including bovine (12), yeast (24), and *Thermus thermophilus* (25), self-associate into hexameric rings. Moreover, Botos *et al.* (14) reported that the recombinant catalytic domain of Ec-Lon protease assembled into the hexameric ring structure within crystal, suggesting a hexameric configuration of the holoenzyme.

Comparison of the Mj-Lon and Ec-lon Protease Domains—The overall fold of the Mj-Lon proteolytic domain is similar to that of Ec-Lon; the r.m.s. deviation is 1.22 Å for 140 equivalent Cα atoms. The configurations of the secondary structure elements are almost identical in the two structures, except for the segment between strand β4 and helix α2 (Fig. 2*A*). The residues extending from the middle of helix α2 to helix α3 (residues 549–594 in *M. jannaschii*), which contain 2 catalytic residues, are structurally well conserved, with a Cα r.m.s. deviation of 0.77 Å, and the catalytic residues, Ser-550 and Lys-593, are located at almost identical positions within each protease. The structural differences mainly originate from the loops connecting the secondary structure elements. The loop β1-β2 near the active site shows the highest conformational difference among the connecting loops. The most significant difference is in the N-terminal portion of helix α2, where the catalytic Ser residue is located. Mj-Lon protease has two additional turns of α-helix bent by 35 degrees at the N-terminal end of helix α2. In Ec-Lon, by contrast, the equivalent segment is a β-strand spanning 6 residues. Sequence comparison clearly shows that Mj-Lon has helix-forming residues in the segment between residues 542 and 550, whereas Ec-Lon has 2 Pro residues hindering the formation of α-helix (Fig. 2*B*). Therefore, the conserved variation in the sequence of the β4-α2 segment results in a significant difference of the configuration of the active site in the two groups of Lon proteases represented by Mj-Lon and Ec-Lon, respectively. Although both proteases have a conserved Asp residue within the β4-α2 segment, only Asp-547 of Mj-Lon is located within salt bridge distance of the catalytic Lys. Consequently, the α-helical structure of the β4-α2 segment in Mj-Lon not only changes the shape of the active site but also significantly alters the properties of the catalytic Lys through the formation of a salt bridge.

Another striking feature of the Mj-Lon proteolytic domain is the presence of a cavity in each protomer containing three water molecules aligned consecutively from within hydrogen-bonding distance of the catalytic Ser-550 toward the interior of the protein (Fig. 3*A*). This cavity is completely buried and inaccessible from the surface of a monomeric protease domain. With a volume of 65 Å^3, it contacts Ser-550 and is situated in a cleft between helix α2 and strands β1 and β5. The Ec-Lon proteolytic domain has a small cavity with a volume of 13 Å^3, which contains one water molecule at a position corresponding to Wat3 of Mj-Lon. There is no contact with the active site residues, however. The known structure of the Ec-Lon proteolytic domain is that of an active site mutant (S679A), which does not faithfully represent the active site structure of the wild-type enzyme due to disruption of the hydrogen-bonding network caused by replacing the catalytic Ser-679 with an Ala. Still, the presence of additional water molecules in the cavity of wild-type Ec-Lon is not probable because the C-terminal half of the β1 strand of Ec-Lon is closer to helix α2 than that of Mj-Lon, resulting in shrinkage of the cavity surrounded by helix α2 and strands β1 and β5.

Active Site—The structure of the active site shows that Mj-Lon employs a pseudocatalytic triad comprised of Ser-550, Lys-593, and Asp-547 (Fig. 3*B*). Asp-547 and the catalytic residue Ser-550 are located in the same face of helix α2, oriented toward Lys-593 in helix α3. Superposition of the structures of
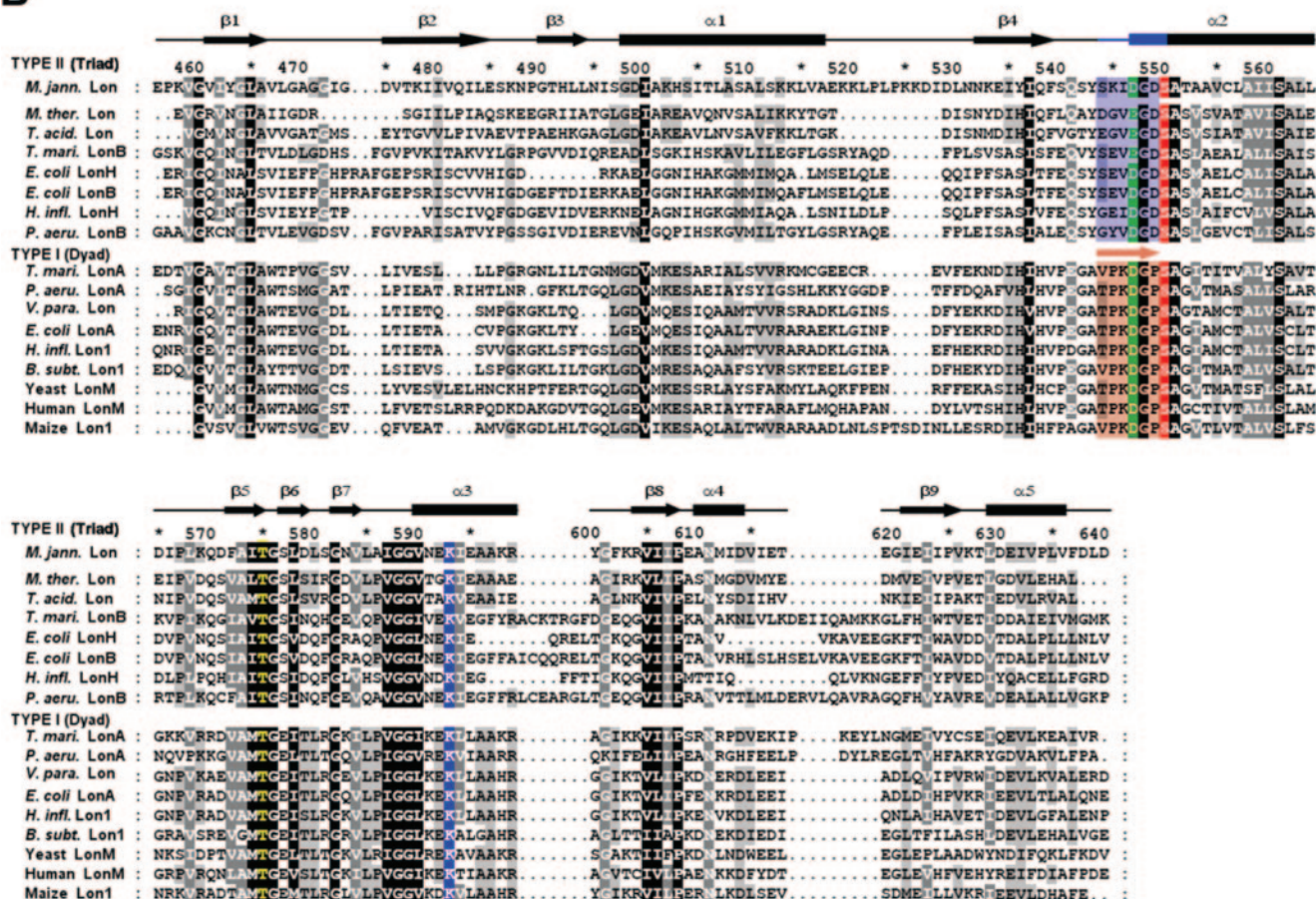
A
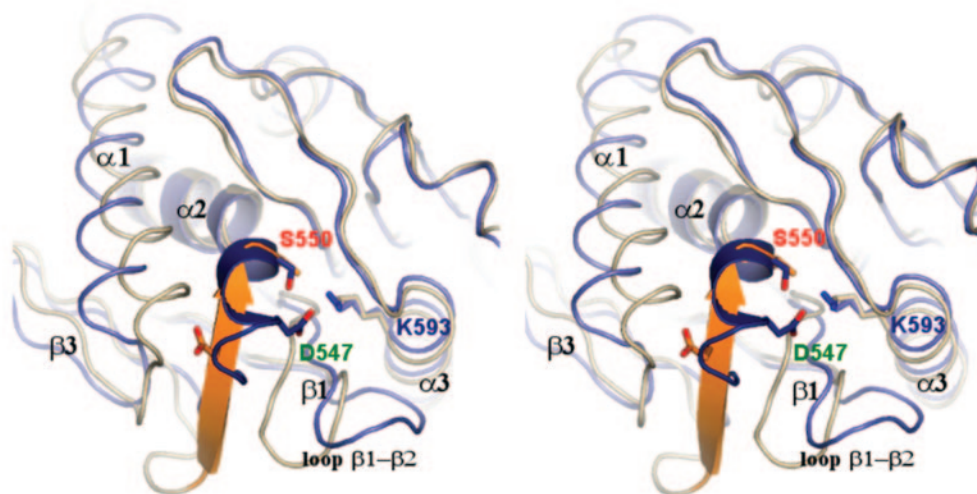
B

FIG. 2. **Structure and sequence comparison of Lon proteolytic domains.** *A*, superposition of the proteolytic domains of Mj- and Ec-Lon; Mj-Lon is shown in *slate blue*, and Ec-Lon (Protein Data Bank code: 1RR9) is shown in *wheat*. The segment between β4 and α2 is colored in *dark blue* for Mj-Lon and in *orange* for Ec-Lon to highlight the different configurations of the respective active sites. Three conserved residues, Asp-547, Ser-550, and Lys-593, and the corresponding Ec-Lon residues are *highlighted in ball-and-stick*. *B*, sequence alignment of the proteolytic domains of 17 Lon family members. Type II Lon proteases, which have a catalytic triad, are listed in the *upper block*; type I Lon proteases, which have a catalytic dyad, are listed in the *lower block*. Highly conserved residues are shaded in *black* and *gray*, and the 6 residues between β4 and α2 are shaded in *blue* and *orange* for group II and group I, respectively. The secondary structure elements of Mj-Lon are shown as *arrows* (β-strands), *bars* (α-helices), and *lines* (connecting loops). The 3 catalytic residues are in different colors that match with the colors of the residue labels in *panel A*. The sequences were aligned using the program ClustalX (35).
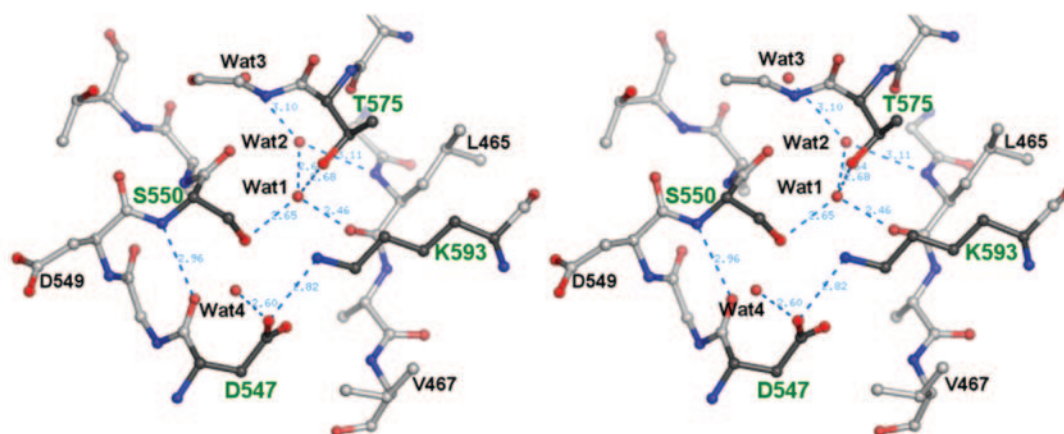
the Ec- and Mj-Lon proteolytic domains shows that Ser-550 and Lys-593 share almost identical positions in the two enzymes, but Mj-Lon has an additional residue, Asp-547, that is

located in the N-terminal end of helix α2 and interacts with the catalytic residues (Fig. 2*A*). The carboxyl group of Asp-547 is located at the first turn of helix α2 and makes a salt bridge with
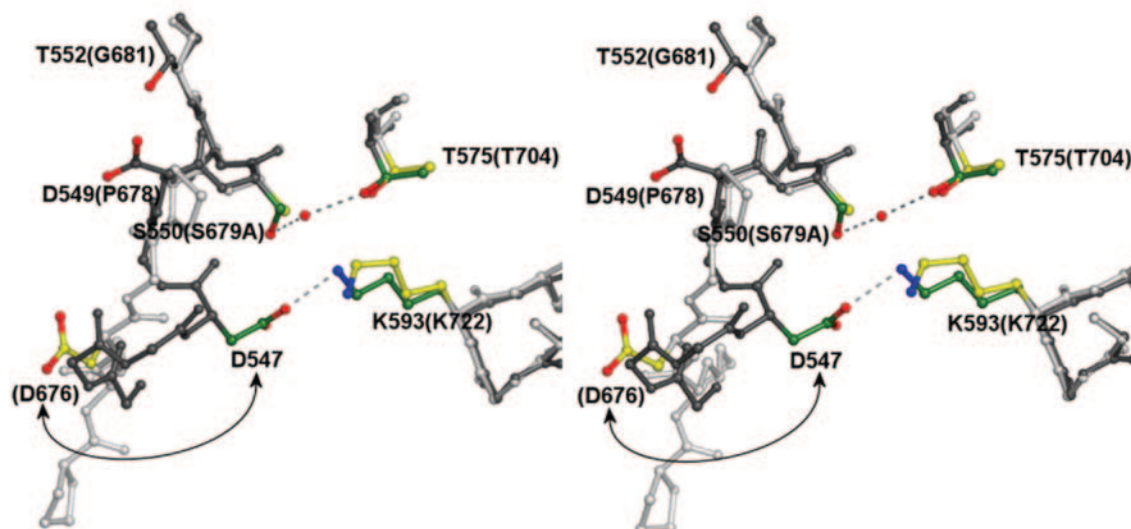
A
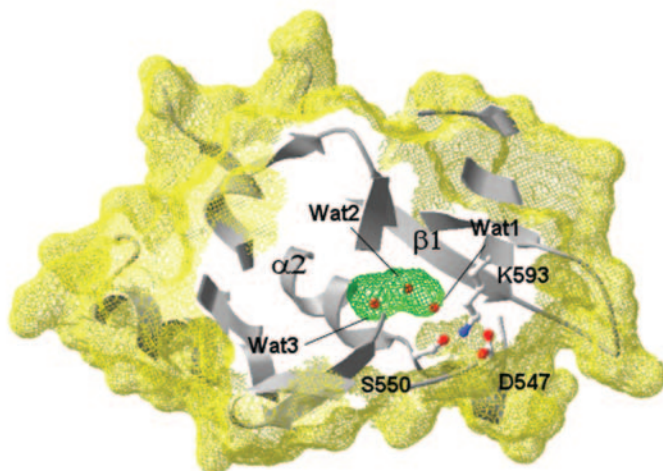
B

C

FIG. 3. **Active site of *M. jannaschii* Lon protease.** *A*, molecular surface of the Mj-Lon proteolytic domain protomer. The secondary structure elements inside the surface are shown. The cavity containing three water molecules is shown in *green mesh*. The 3 catalytic residues and the water molecules in the cavity are *highlighted* in *ball-and-stick*. *B*, stereo diagram of active site shown in *ball-and-stick*. *C*, superposition of the active sites of the Mj- and Ec-Lon proteolytic domains. Mj-Lon is colored in *gray*, and Ec-Lon is colored in *white*. The side chains of 4 conserved residues are highlighted in *green* for Mj-Lon and in *yellow* for Ec-Lon. The residue numbers in the Ec-Lon proteolytic domain are shown in *parentheses*. The *arrow* indicates equivalent Asp residues in both proteases.
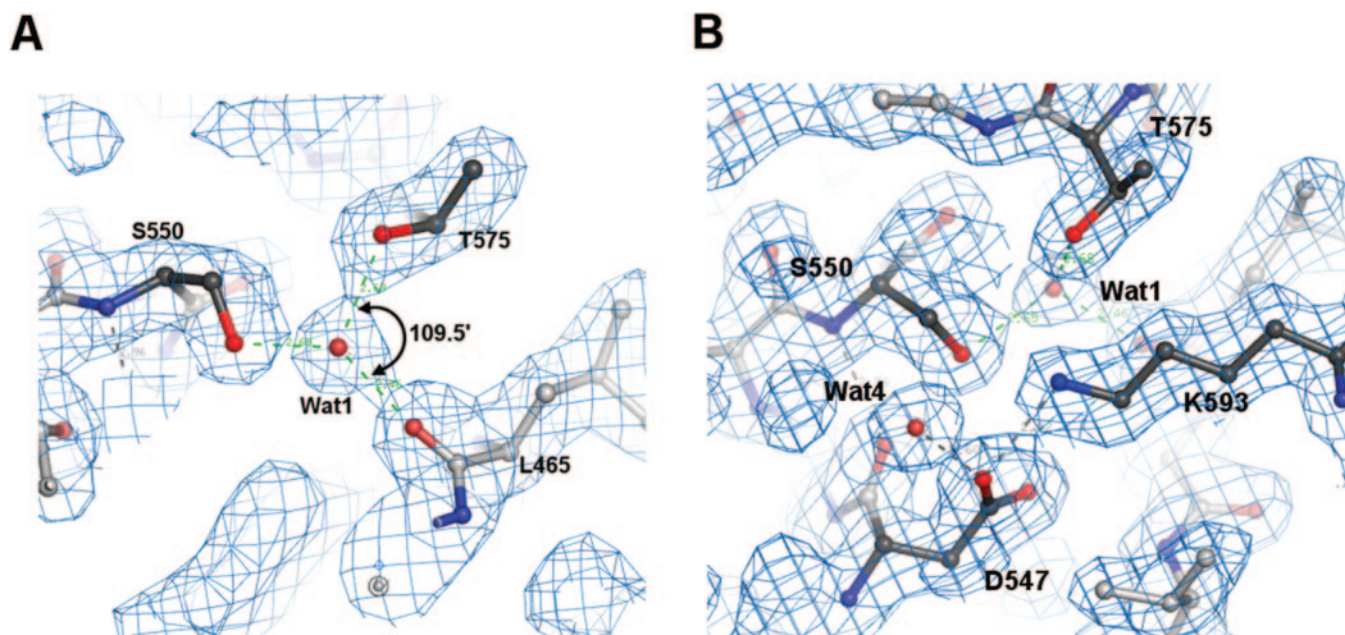
Fig. 4. *A and B, **ball-and-stick** model of the active site with a 2F_o − F_c electron density map superimposed to show the coordination of a water molecule by hydrogen bonds.* All figures were prepared using the program PyMOL (www.pymol.org).

Lys-593 and a hydrogen bond with a water molecule. In Ec-Lon, the Asp residue is also conserved at the sequence level, but it is exposed to the solvent and not involved with the active site residues because the segment corresponding to the N-terminal end of helix α2 is a β-strand, which puts the Asp residue at a position distant from the active site (Fig. 3*C*).

The catalytic mechanism of the classical Ser-His-Asp triad is known to begin with the polarization of the Ser residue by the His residue, which abstracts the hydroxyl proton of the Ser residue. The Asp stabilizes the charged imidazole of the His, promoting its role as a base (26). Conceivably, the Lys in the Ser-Lys dyad may serve the same function as the His in serine proteases. Although Ec-Lon protease has no sequence or structural similarity to signal peptidase or the LexA family of serine proteases, the Ec-Lon proteolytic domain is believed to utilize a Ser-Lys dyad mechanism similar to that used by signal peptidase, in that it has conserved geometries of the Ser-Lys dyad and a Thr residue assisting the dyad (14, 27). The structures of signal peptidases suggest that the hydrophobic environment surrounding the catalytic Lys ϵ-amino group is likely essential for lowering its pK_a so that it can reside in the deprotonated state required for its role as a general base (28, 29). It is supposed from the structural comparison that Lys-722 in Ec-Lon is not charged and acts as a general base as there is no acidic residue in the vicinity of the catalytic Lys and the side chain Oγ of the conserved Thr might be within hydrogen-bonding distance of the catalytic Lys in the wild-type enzyme (14). It appears that the Ser-Lys dyad mechanism requires a third residue (Thr/Ser Oγ) for optimal activity, possibly serving to help align the general base Lys with the nucleophilic Ser hydroxyl via a hydrogen-bonding network. The role of such a Thr (Thr-704 in Ec-Lon) might be similar to the Asp present in the classic catalytic triad of serine proteases (30).

The same mechanism is not applicable to the Ser-Lys-Asp triad of Mj-Lon because the chemical properties of the Lys-Asp pair are quite different from those of a deprotonated Lys. In the environment of the active site of Mj-Lon, Lys-593 forms a salt bridge with Asp-547 and, presumably, is therefore positively charged. Moreover, unlike Ec-Lon, the conserved Thr-575 is not hydrogen-bonded to Lys-593 but to a water molecule (Wat1). Consequently, Lys-593 is unsuitable to function as a proton

acceptor to abstract a proton from Ser-550. It is expected that the role of Lys-593 is to lower the pK_a of the hydroxyl group of Ser-550 via its electrostatic potential so that the Oγ of Ser-550 can act as a nucleophile during proteolysis, as is seen in the structures of 20 S proteasomes (31).

The most probable proton acceptor is Wat1, which is located in the cavity and hydrogen-bonds with Ser-550. Wat1 is ideally positioned to act as the general base and promote the abstraction of a proton from the Ser-550 hydroxyl group, initiating nucleophilic attack on the carbonyl carbon of the peptide bond. The oxygen atom of Wat1 points directly at the hydroxyl of Ser-550 as if ready to deprotonate it, whereas the two hydrogen atoms of Wat1 are firmly hydrogen-bonded to the carbonyl oxygen of Leu-465 and hydroxyl group of Thr-575, as is shown in the 2F_o − F_c electron density map (Fig. 4*A*, 4*B*). Coordination of Wat1 by carbonyl and hydroxyl oxygens suggests that it might act as a proton acceptor in the initial polarization of Ser-550 during catalysis. The presence of an additional carbonyl oxygen of Ser-550 near Wat1 may stabilize the protonated water molecule. The conserved Thr residue (Thr-575 in Mj-Lon), which assists the catalytic dyad in Ec-Lon by hydrogen bonding with the deprotonated Lys, hydrogen-bonds instead with Wat1 in Mj-Lon, implying that its function differs from that in Ec-Lon.

An important component of the catalytic machinery in serine proteases is the oxyanion hole that works by neutralizing the developing negative charge on the scissile carbonyl oxygen during the formation of the tetrahedral intermediates. Typical oxyanion holes are formed by two main-chain amide hydrogens that serve as hydrogen bond donors to the developing oxyanion (32). The oxyanion hole in Mj-Lon is formed by two main chain amide hydrogens, one originating from the catalytic Ser-550 and the other from the preceding residue, Asp-549.

Despite the similar geometries of Ser and Lys residues and the similarity of the oxyanion holes in signal peptidases and the proteolytic domain of Mj-Lon, it is conceivable that Mj-Lon does not employ a Ser-Lys dyad because the hydrogen-bonding environment for the catalytic Lys is quite different. Instead, a structural comparison suggests that the mechanism of Mj-Lon protease is comparable with the autolytic mechanism of prosequences in the 20 S proteasomes, which belong to the N-termi-

nal threonine protease family. In 20 S proteasomes, a conserved Lys charged by the salt bridge with a nearby Asp lowers the pK_a of the Thr-1 Oγ via its electrostatic potential, and a base close to Thr-1 plays a role in the proton transfer (33). Because the N-terminal amino group is not available to serve as a proton acceptor before activation of the protease through cleavage of the prosequence, a water molecule close to Thr-1 Oγ presumably serves as the proton acceptor in the initial autolysis step (31). Although the folds of the 20 S proteasome and Mj-Lon proteolytic domains are completely different overall, the presence of equivalent catalytic residues (Ser, Lys, and Asp) suggests that they share a similar proteolytic mechanism. However, the exact catalytic mechanism of Mj-Lon cannot be deduced from its structure. Further biochemical and structural analysis will be needed to reveal the precise mechanism.

Classification of Proteolytic Domains of Lon Proteases—A search using the proteolytic domain of Mj-Lon against the Swiss-Prot data base with the BLAST algorithm identified over 100 different Lon proteases from 90 different organisms. Extensive sequence alignment of homologous proteins showed that the segment containing the catalytic residues (residues extending from helix α2 to α3, residues 545–596 in *M. jannaschii*) are well conserved among all Lon protease families; moreover, the catalytic residues, Ser-550 and Lys-593 in *M. jannaschii*, are strictly conserved, without exception. On the other hand, the residues preceding the catalytic Ser are clearly divided into two groups: Asp (Asp-549 in *M. jannaschii*) preceded by α-helix-forming residues and Pro (Pro-678 in *E. coli*) preceded by β-strand-forming residues (Fig. 2*B*). A third putative residue in the active site of Mj-Lon, Asp-547, is also well conserved at the sequence level as either an Asp or a Glu among all proteases but has a different structural configuration for each group of proteases as described earlier. Thus, the shape of the active site pocket will significantly differ, depending on whether a Pro or an Asp precedes the catalytic Ser. Consequently, the protease domains of Lon can be classified into two groups with consensus sequences of DGPSA or (D/E)GDSA in the active site, which are represented by the proteolytic domains of Ec- and Mj-Lon in the present study. Hereafter, these two types of Lon proteases will be referred as type I and type II proteases, respectively.

Comparison of the structures of the Mj- and Ec-Lon proteolytic domains revealed that these enzymes likely have different catalytic mechanisms, although they share the same overall fold and conserved dyad residues (Ser-550 and Lys-593). The difference in the shapes of active sites also implies that the substrate specificity and the catalytic activity of these proteases differ. Most type II Lon proteases are archaeal, but some have been identified in eubacteria (*e.g. E. coli* and *Haemophilus ducreyi*). Some microorganisms (*e.g. E. coli*, *Bacillus subtilis*, and *Pseudomonas aeruginosa*) express both Lon types. Lon homologues located in a second open reading frame found in various bacteria have been designated LonB and are not necessarily type II Lon proteases. It is clear that LonA (*e.g.* Ec-Lon and Mj-Lon) and LonB are distinct as the latter contains an active site region (the catalytic Ser residue) but not an ATP-

binding site (34). The presence of different types of Lon homologues within a cell suggests that they have distinct metabolic functions and substate specificities. Type II Lon is not found in eukaryotes (*e.g.* yeast, mouse, human, and other higher organisms), implying that dyad proteases are selected evolutionarily for eukaryotic cellular function.

REFERENCES

1. Wickner, S., Maurizi, M. R., and Gottesman, S. (1999) *Science* **286,** 1888–1893
2. Gottesman, S. (1999) *Curr. Opin. Microbiol.* **2,** 142–147
3. Fukui, T., Eguchi, T., Atomi, H., and Imanaka, T. (2002) *J. Bacteriol.* **184,** 3689–3698
4. Wang, J., Hartling, J. A., and Flanagan, J. M. (1997) *Cell* **91,** 447–456
5. Yoo, S. J., Shim, Y. K., Seong, I. S., Seol, J. H., Kang, M. S., and Chung, C. H. (1997) *FEBS Lett.* **412,** 57–60
6. Swamy, K. H., and Goldberg, A. L. (1981) *Nature* **292,** 652–654
7. Gottesman, S. (1996) *Annu. Rev. Genet.* **30,** 465–506
8. Fu, G. K., Smith, M. J., and Markovitz, D. M. (1997) *J. Biol. Chem.* **272,** 534–538
9. Goldberg, A. L., Moerschell, R. P., Chung, C. H., and Maurizi, M. R. (1994) *Methods Enzymol.* **244,** 350–375
10. Lee, A. Y., Tsay, S. S., Chen, M. Y., and Wu, S. H. (2004) *Eur. J. Biochem.* **271,** 834–844
11. Stahlberg, H., Kutejova, E., Suda, K., Wolpensinger, B., Lustig, A., Schatz, G., Engel, A., and Suzuki, C. K. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6787–6790
12. Watabe, S., and Kimura, T. (1985) *J. Biol. Chem.* **260,** 14498–14504
13. Birghan, C., Mundt, E., and Gorbalenya, A. E. (2000) *EMBO J.* **19,** 114–123
14. 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., and Gustchina, A. (2004) *J. Biol. Chem.* **279,** 8140–8148
15. Doublie, S. (1997) *Methods Enzymol.* **276,** 523–530
16. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
17. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
18. Terwilliger, T. C. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 1763–1775
19. Levitt, D. G. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 1013–1019
20. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
21. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
22. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) *Proteins* **12,** 345–364
23. Lee, A. Y., Hsu, C. H., and Wu, S. H. (2004) *J. Biol. Chem.*
24. van Dijl, J. M., Kutejova, E., Suda, K., Perecko, D., Schatz, G., and Suzuki, C. K. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 10584–10589
25. Watanabe, S., Muramatsu, T., Ao, H., Hirayama, Y., Takahashi, K., Tanokura, M., and Kuchino, Y. (1999) *Eur. J. Biochem.* **266,** 811–819
26. Fersht, A. R. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, pp. 472–482, W. H. Freeman and Co., New York
27. Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (2002) *J. Biol. Chem.* **277,** 9512–9519
28. Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (1998) *Nature* **396,** 186–190
29. Paetzel, M., Strynadka, N. C., Tschantz, W. R., Casareno, R., Bullinger, P. R., and Dalbey, R. E. (1997) *J. Biol. Chem.* **272,** 9994–10003
30. Dodson, G., and Wlodawer, A. (1998) *Trends Biochem. Sci.* **23,** 347–352
31. Groll, M., and Huber, R. (2003) *Int. J. Biochem. Cell Biol.* **35,** 606–616
32. Menard, R., and Storer, A. C. (1992) *Biol. Chem. Hoppe-Seyler* **373,** 393–400
33. Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H. D., and Huber, R. (1997) *Nature* **386,** 463–471
34. Ward, D. E., Shockley, K. R., Chang, L. S., Levy, R. D., Michel, J. K., Conners, S. B., and Kelly, R. M. (2002) *Archaea* **1,** 63–74
35. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882