

Atomic-resolution Crystal Structure of the Proteolytic Domain of *Archaeoglobus fulgidus* Lon Reveals the Conformational Variability in the Active Sites of Lon Proteases

Istvan Botos¹, Edward E. Melnikov², Scott Cherry¹, Serguei Kozlov³
Oksana V. Makhovskaya², Joseph E. Tropea¹, Alla Gustchina¹
Tatyana V. Rotanova² and Alexander Wlodawer^{1*}

¹Macromolecular
Crystallography Laboratory
National Cancer Institute at
Frederick, Frederick, MD
21702-1201, USA

²Shemyakin-Ovchinnikov
Institute of Bioorganic
Chemistry, Russian Academy of
Sciences, Moscow 117997
Russian Federation

³Cancer and Developmental
Biology Laboratory, National
Cancer Institute at Frederick
Frederick, MD 21702-1201
USA

The atomic-resolution crystal structure of the proteolytic domain (P-domain, residues 415–621) of *Archaeoglobus fulgidus* B-type Lon protease (wtAfLonB) and the structures of several mutants have revealed significant differences in the conformation of the active-site residues when compared to other known Lon P-domains, despite the conservation of the overall fold. The catalytic Ser509 is facing the solvent and is distant from Lys552, the other member of the catalytic dyad. Instead, the adjacent Asp508 forms an ion pair with the catalytic lysine residue. Glu506, an analog of the putative third catalytic residue from a related *Methanococcus jannaschii* LonB, also faces the solvent and does not interact with the catalytic dyad. We have established that full-length wtAfLonB is proteolytically active in an ATP-dependent manner. The loss of enzymatic activity of the S509A mutant confirms the functional significance of this residue, while retention of considerable level of activity by the D508A and E506A mutants rules out their critical involvement in catalysis. In contrast to the full-length enzymes, all individually purified P-domains (wild-type and mutants) were inactive, and the mutations had no influence on the active-site structure. These findings raise the possibility that, although isolated proteolytic domains of both AfLonB and *E. coli* LonA are able to assemble into expected functional hexamers, the presence of the other domains, as well as substrate binding, may be needed to stabilize the productive conformation of their active sites. Thus, the observed conformational variability may reflect the differences in the stability of active-site structures for the proteolytic counterparts of single-chain Lon versus independently folded proteolytic subunits of two-chain AAA⁺ proteases.

Published by Elsevier Ltd.

Keywords: active site; ATP-dependent proteases; atomic resolution; catalytic dyad; structure comparisons

*Corresponding author

Introduction

Energy-dependent proteases of the Lon family (MEROPS:¹ clan SJ, ID S16) are peptide hydrolases

Abbreviations used: AAA⁺, ATPase associated with various cellular activities; DDM, *n*-dodecyl- β -D-maltopyranoside; MBP, maltose-binding protein; SAD, single-wavelength anomalous dispersion; TEV, tobacco etch virus.

E-mail address of the corresponding author:
wlodawer@ncifcrf.gov

with a catalytic Ser-Lys dyad that participate in the rapid turnover of short-lived regulatory proteins and control protein quality through eliminating mutant and abnormal proteins.^{2–4} Canonical Lon proteases all include AAA⁺-type ATPase domains (AAA⁺ modules) that precede their proteolytic domains, and are found in all three domains of life, Bacteria, Archaea and Eucarya.^{5–7} Recently, it became clear that the Lon family can be subdivided into two subfamilies, LonA and LonB, with different consensus sequences around the catalytic residues of their proteolytic domains.^{2,4}

Furthermore, these differences seem to be strictly associated with the characteristics of their AAA⁺ modules, as well as with the presence or absence of additional N-terminal domains.⁴⁻⁷

The LonA subfamily, typified by the “classical” Lon protease from *Escherichia coli*, includes mainly cytosolic bacterial and mitochondrial eukaryotic enzymes, accounting for about 80% of the presently known Lon proteases.⁴ All members of the LonA subfamily contain the successively linked N-terminal, AAA⁺, and proteolytic domains (N, A and P-domains, correspondingly).⁸ Enzymes belonging to the LonB subfamily (the remaining 20% of all Lon proteases) are found only in Archaea.⁴⁻⁷ The architecture of the LonB proteases is significantly different from that of LonA proteases, since they consist of AAA⁺ modules and proteolytic domains, but lack the N-terminal domains. In addition, LonB proteases are membrane-bound through transmembrane segments that are inserted within their AAA⁺ modules.⁴⁻⁷

LonA protease from *E. coli* (*EcLonA*) was the first ATP-dependent protease to be discovered.^{9,10} Subsequently, a number of other A-type enzymes from bacterial and eukaryotic sources have been isolated and described.¹¹⁻¹⁹ Still, *EcLonA* is the most widely studied enzyme of the LonA subfamily and it has become a model enzyme for the study of the ATP-dependent protease function.²⁰⁻²⁴

No three-dimensional structure of any full-length Lon protease has been reported to date. Recently, we have determined the crystal structures of two *EcLonA* protease fragments identified by limited proteolysis; the P-domain³ and the α -domain (the C-terminal part of the AAA⁺ module).²⁵ The

P-domain exhibited a unique fold and formed a ring-shaped hexamer in the crystals, whereas the monomeric α -domain had a fold typical for other AAA⁺ protein α -domains.

Much less is known about the properties of the membrane-bound LonB enzymes. The enzymatic properties of two LonB proteases, isolated from *Thermococcus kodakaraensis* (*TkLonB*) and *Thermoplasma acidophilum* (*TaLonB*), were described recently, and they were found to be noticeably different from those of *EcLonA*.²⁶⁻²⁸ In addition, the recently published crystal structure of the proteolytic domain of LonB protease from *Methanococcus jannaschii* (*MjLonB*) showed that, whereas the overall fold of the catalytic domain of *MjLonB* resembled that of *EcLonA*, significant differences were found in the active site.²⁹ This observation led to suggestion of a different catalytic mechanism for LonB than that of LonA,²⁹ although only very limited biochemical characterization of *MjLonB* was presented. The significant differences between the two Lon subfamilies indicate that further work on defining their properties is needed.

For the crystallographic and enzymatic studies presented here, we have selected Lon protease from the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus* (*AfLonB*) as a representative of the LonB subfamily. In common with *EcLonA*, *AfLonB* displays ATP-dependent proteolytic activity, but only at elevated temperatures. In order to correlate the biochemical properties of *AfLonB* with its structure, we have solved the crystal structure of its proteolytic domain at high resolution and compared it to the previously described structures of the *EcLonA* and *MjLonB*

Table 1. Statistics of data collection

	Hexagonal				Orthorhombic ^a	Monoclinic ^a
	Se-Met	Native	D508A	E506A	Native	Native
Construct	AFPR2	AFPR2	AFPR2	AFPR2	AFPP	AFPR1
Space group	<i>P</i> ₆₅	<i>P</i> ₆₅	<i>P</i> ₆₅	<i>P</i> ₆₅	<i>P</i> ₂₁ <i>2</i> ₁ <i>2</i> ₁	<i>P</i> ₂₁
Molecules/asym.unit	1	1	1	1	6	6
Wavelength (Å)	0.97926	0.99997	1.0	1.0	1.54	1.54
Unit cell parameters						
<i>a</i> (Å)	83.74	84.47	81.96	83.26	86.65	48.68
<i>b</i> (Å)	83.74	84.47	81.96	83.26	88.69	86.11
<i>c</i> (Å)	41.23	41.50	41.50	41.17	147.24	135.61
β (deg.)						94.7
Resolution (Å)	50–1.7	50–1.15	30–1.55	50–1.55	20–2.3	50–2.05
Total reflections	204,757	598,283	259,230	188,231	186,661	300,097
Unique reflections	35,303	52,481	23,366	23,792	50,449	68,204
Completeness ^b (%)	99.2 (100.0)	96.5 (94.4)	100.0 (100.0)	99.9 (100.0)	97.1 (95.7)	96.5 (76.8)
$\langle I \rangle / \langle \sigma(I) \rangle$	25.8 (2.3)	24.5 (3.1)	41.4 (2.8)	34.4 (2.8)	9.3 (2.1)	12.6 (1.3)
R_{merge}^c (%)	7.2 (63.5)	7.0 (68.3)	5.6 (67.3)	6.3 (45.7)	12.9 (48.2)	13.2 (58.2)
Phasing statistics (30–1.70 Å)						
Number of Se sites	5					
Anomalous phasing power (acentric)	2.06					
Anomalous Cullis \bar{R} (acentric)	0.57					
Figure of merit (acentric)	0.41					
Correlation after solvent flattening	0.65					

^a These data were collected from twinned crystals and represent reflections isolated for one component of the twin only. The high values of the merging R -factors are largely due to the contamination of intensities by the contribution of the other twin.

^b The values in parentheses relate to the highest-resolution shell.

^c $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity obtained from multiple observations.

proteolytic domains. Since the isolated proteolytic domains of Lon proteases do not exhibit detectable enzymatic activity, we have analyzed the enzymatic activity of the full-length wild-type *AfLonB* along with several mutant variants, establishing the relative importance of selected residues in the active site. Structural studies of these mutants in the context of the isolated proteolytic domains complemented the biochemical data.

Results

Structure solution

The crystal structure of the *AfLonB* P-domain was solved by the single-wavelength anomalous dispersion (SAD) technique using the data (Table 1) obtained from a Se-Met derivative of a hexagonal crystal form of the D415-V621 variant of the recombinant P-domain (AFPR2, see Materials and Methods for definition). After solvent flattening, the experimental electron density map could be traced automatically for almost all the residues. Partially refined coordinates of the derivative were subsequently used to solve and refine the native structure at the atomic resolution of 1.2 Å. The structure has been refined utilizing riding hydrogen atoms, individual anisotropic temperature factors, and a model of bulk solvent. The final model consists of residues 416–621 of *AfLonB* (numbering corresponding to the full-length protein), since the first residue of the construct used to prepare the AFPR2 variant of the protein was not observed in the electron density. In addition, three residues

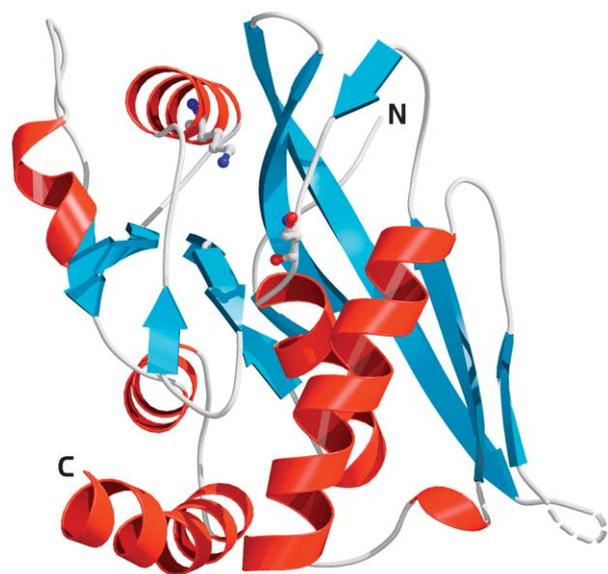


Figure 1. Structure of the proteolytic domain of *AfLonB*. The protein is shown as a ribbon diagram, with the catalytic dyad residues highlighted in ball-and-stick representation. Residues not seen in the electron density map are dashed. The Figure was generated with Bobscrip⁵⁹ and Raster3D.⁶⁰

(454–456) located at the tip of a loop are not visible in the electron density and were not included in the model. The final model includes 228 water molecules modeled as single oxygen atoms, and two calcium cations (Figure 1). The side-chains of nine residues (Ser478, Ile480, Asn491, Ser509, Ser511, Ile522, Ser536, Val564, and Met610) have been modeled in double conformation. The carboxyl terminus of the enzyme is very well ordered due to anchoring interactions with one of the calcium ions. The protein is densely packed in the crystal lattice, with the value of the Matthews coefficient V_M of only 1.94 Å³/Da,³⁰ and the calculated solvent content of 36.5% (v/v). This dense packing correlates with the relatively limited number of solvent molecules that were modeled. The Ramachandran plot as implemented in PROCHECK³¹ indicated that 96% of the main-chain torsion angles for residues other than proline and glycine were found in the most-favored regions and 4% in the additionally allowed regions (98.5% of residues in favored regions according to MolProbity³²). The low values of the final *R*-factors, very good geometry of the refined model, and excellent quality of the resulting electron density map (Figure 2), all support the correctness of this structure. The structure discussed here in detail corresponds to this high-resolution model, unless explicitly indicated otherwise.

The hexagonal crystals of the D508A mutant appeared isomorphous to the wild-type, but in reality represented considerably different packing of the molecules.³³ This necessitated solution of the structure by molecular replacement, using the model discussed above as a starting point. The structure was refined at 1.55 Å resolution (Table 2) and showed that, in this crystal form, the molecule is rotated by ~23° compared to the wild-type enzyme. The calcium cations that stabilize the conformation of the C terminus of the wild-type protein are not seen and thus this region of the mutant is less well ordered. The mutation did not affect the overall structure of the enzyme, although a stretch of residues encompassing the mutation site (498–510) was less well ordered than in the wild-type structure. This was reflected by higher temperature factors, while the conformation of this segment of the structure was not affected.

The E506A mutant was isomorphous with the wild-type enzyme and its structure was not modified in any significant way, beyond the removal of the side-chain of residue 506. The calcium ions were present and the C terminus was well ordered.

Two additional structures of the *AfLonB* P-domain were solved using crystals grown from protein obtained either by chymotrypsin digestion (AFPP) or from the initially prepared, shorter clone (AFPR1). Both variants of the protein produce crystals that can be described as very unusual twins that contain superimposed orthorhombic and monoclinic lattices.³³ The structure of AFPP P-domain was refined with the orthorhombic data

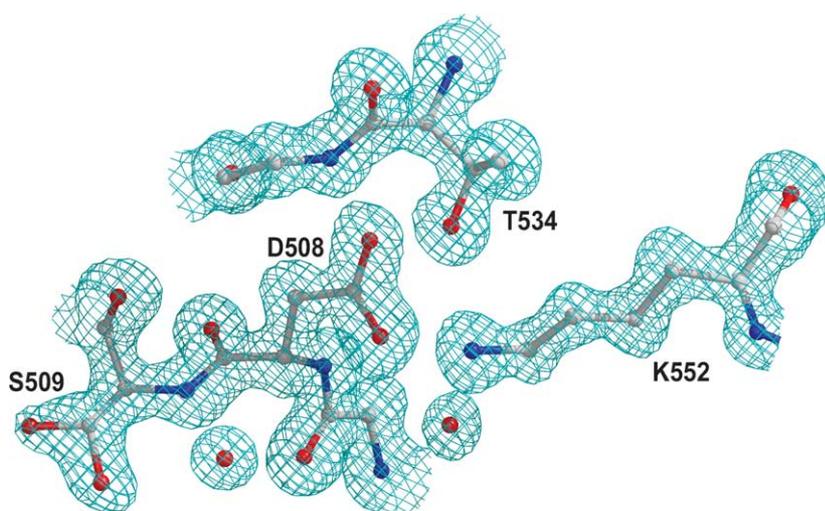


Figure 2. An electron density map of the *AfLonB* P-domain active site. The $2F_o - F_c$ map was calculated at 1.2 Å resolution after the final refinement cycle and was contoured at 1.4σ . The map clearly supports double orientation of the side-chain of Ser509 and its positioning away from Lys552.

extending to 2.3 Å, whereas the structure of AFPR1 P-domain was refined with the monoclinic data extending to 2.05 Å. The crystallographic *R*-factors for both structures are comparatively high (Table 2), most likely due to difficulties encountered in separating diffraction from the two lattices, but there is no doubt about their overall correctness. The packing of these crystals is slightly less dense, with the value of V_M equal 2.14 Å³/Da, corresponding to 42.4% (v/v) solvent content.

Description of the structure

The overall fold of the proteolytic domain of *AfLonB* is very similar to the folds of the other two known Lon proteolytic domains. The r.m.s. deviations between the C^α coordinates of *AfLonB* and *EcLonA*³ are 1.72 Å for 164 atom pairs, and 1.46 Å for 163 pairs for *MjLonB*.²⁹ Since the secondary structure of *EcLonA* was the first one to be described in detail, it will be used for reference. The N terminus of the *AfLonB* P-domain forms a

large loop that includes the first very short strand S0 (residues 423–425, Figure 3(a)), not present in *EcLonA*. This loop is followed by strand S1 (residues 428–435) and an antiparallel strand S2 (439–451) that together form a long β -hairpin loop. The loop and the parallel strands S3 (459–461) and S4 (492–498), separated by helix H1 (467–485) and the short $_3$ ₁₀ helix H1A (489–491), form the first large β sheet (Figure 1). Strand S5 (505–506) is perpendicular to the plane of this β sheet and is connected to helix H2 (512–523) by a loop that contains the catalytic Ser509. Helices H1 and H2 interact with the first β sheet, forming a relatively compact subdomain.

Following helix H2, a random coil (524–526) forms a bridge to the second subdomain. Two short strands, S5A (527–528) and S6 (531–533), lead into another β loop (535–544) formed by antiparallel strands S7 and S8, followed by helix H3 (549–559). Helix H3 runs nearly parallel along the distal surface and carries the second catalytic residue, Lys552. Strand S9 (563–567) returns along helix H3,

Table 2. Refinement statistics for the final structures

	Hexagonal			Orthorhombic	Monoclinic
	Native	D508A	E506A	Native	Native
Resolution (Å)	1.2	1.55	1.55	2.3	2.05
<i>R</i> (%) ^a	13.7 (13.0) ^b	21.3	17.9	19.5	21.3
<i>R</i> _{free} (%) ^c	18.0 (17.1) ^b	26.4	20.7	32.8	29.9
r.m.s.d.					
Bond lengths (Å)	0.013	0.014	0.013	0.045	0.033
Bond angle distances (Å) or angles (deg) ^d	0.031	2.10	1.47	3.24	2.49
Temperature factors					
Protein (Å ³)	21.9	38.0	31.6	12.7	37.7
Solvent (Å ³)	38.3	56.7	53.1	24.1	44.4
Number of protein atoms	1540	1518	1519	8778	8742
No. of calcium ions	2	0	2	0	0
No. solvent molecules	228	173	211	1095	552
PDB code	1z0w	1z0c	1z0b	1z0g	1z0e

^a $R = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively, calculated for all data.

^b Values in parentheses show the statistics from SHELXL for reflections with $F > 4\sigma(F)$.

^c R_{free} was as defined.³⁴

^d Distances from SHELXL or angles from REFMAC5.

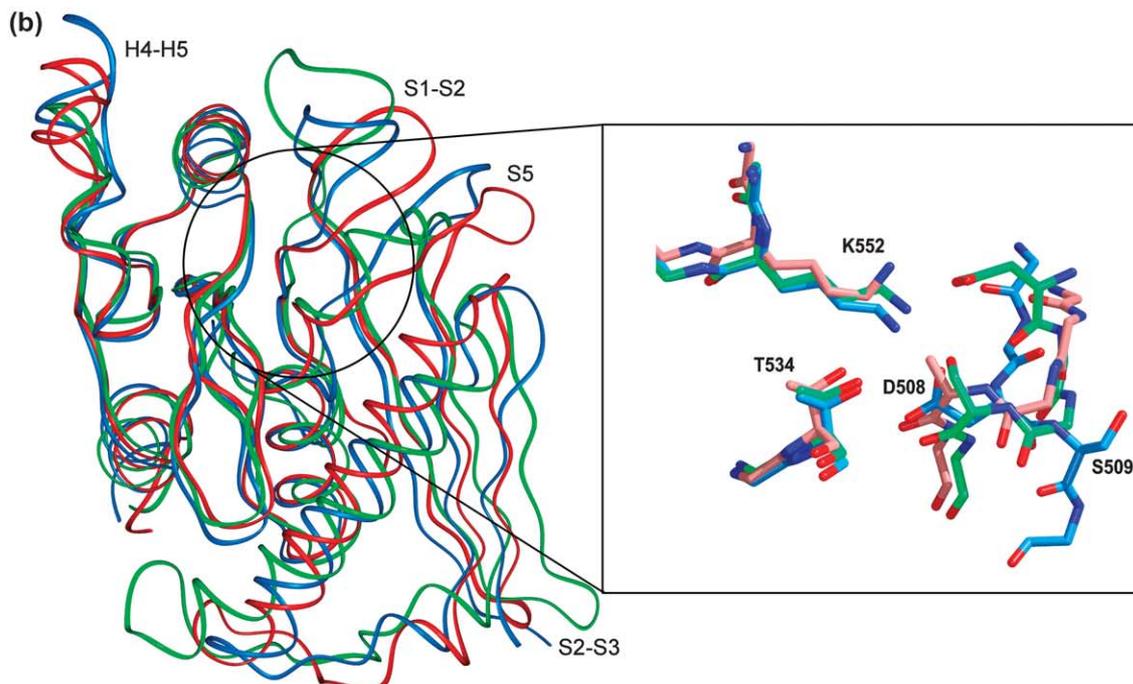
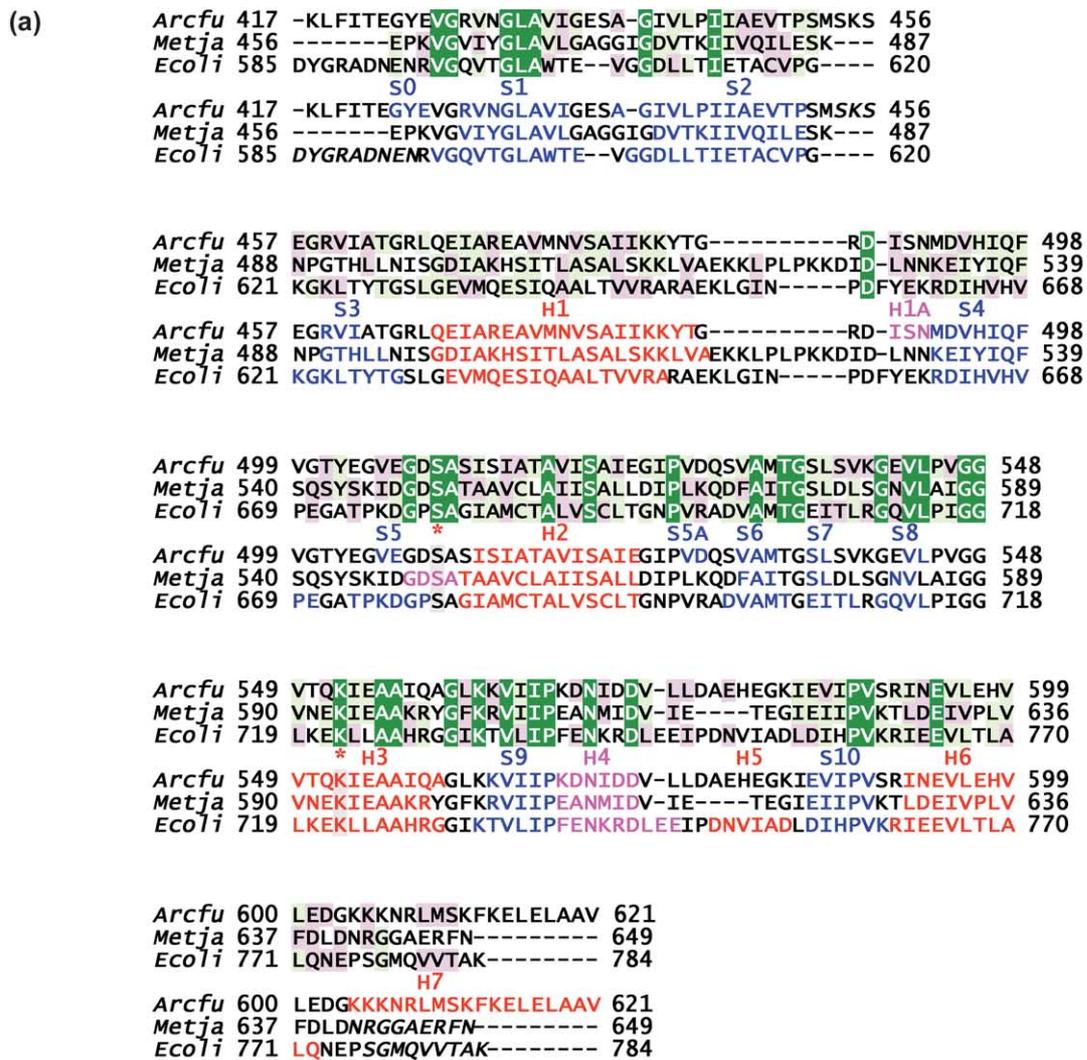


Figure 3. Comparison of Lon P-domains. (a) Structure-based sequence alignment of the P-domains of *A*LonB (MER03868), *Mj*LonB (MER03359, PDB code 1xhk), and *Ec*LonA (MER00485, PDB code 1rre). The top block highlights

followed by a 3_{10} helix 4 (568–573) and parallel strand S10 (585–589). Strands S6, S9, and S10 form a third, small β sheet, sandwiched by helices H3 and H6 (592–599). A short loop connects helix H6 to the long C-terminal helix H7 (604–620), not present in the other two Lon P-domain structures. The end of this helix is unwound from residue 615 in the orthorhombic and monoclinic crystals, and could not be modeled in these structures.

One N-terminal residue and three residues in the 454–456 loop are not visible in the electron density of the hexagonal crystal form, while four N-terminal residues are disordered in the orthorhombic and monoclinic crystal forms. The side-chain of Tyr416 is anchored firmly in the hexagonal structure of the wild-type enzyme, but is not seen in the other two. The loop region that is not visible in the electron density in all three crystal forms (broken lines in Figure 1) is located at the apical end of the molecule. Some biological importance might be attributed to the flexibility of this loop, since all other residues are very well defined in the structure. In limited proteolytic cleavage experiments, one of the lower molecular mass species has its N terminus beginning at this loop. Given its apical position, the loop is probably interacting with the distal side of the A-domain hexameric ring (see below).

Active site

Very well defined electron density in the area that corresponds to the active site of the wild-type *AfLonB* (Figure 2) revealed an unexpected mutual disposition of the catalytic residues. The side-chain of Ser509 points out into the solvent and its O^γ group occupies two positions, one within 2.5 Å of a water molecule, and another 3.27 Å away from $O^{\delta 1}$ of Glu472. Lys552 is very well ordered, as indicated by the temperature factor of its side-chain, which is below the average for the protein. Its N^ζ group makes three hydrogen bonds, being 2.77 Å away from the carbonyl oxygen atom of Gly547, 2.88 Å away from $O^{\gamma 1}$ of Thr534, and 2.69 Å away from $O^{\delta 1}$ of Asp508. The $O^{\delta 2}$ group of Asp508 makes another hydrogen bond to $O^{\gamma 1}$ of Thr534 (2.64 Å) and accepts a bond from the main-chain amide nitrogen atom of Gly535 (2.99 Å). Removal of the side-chain of Asp508 in the D508A mutant does not change significantly the positions of either Lys552 (which is still well ordered and makes hydrogen bonds to

Thr534 and Gly547), or Ser509 (for which only a single orientation, hydrogen bonded to a water molecule, is observed in this case). As discussed below, the observed conformation of the active site, although determined unambiguously, cannot represent an active enzyme.

Oligomeric structure

The asymmetric unit of both the orthorhombic and monoclinic crystals consists of six *AfLonB* P-domains arranged as a ring with approximate non-crystallographic 6-fold symmetry (Figure 4(a)). The hexagonal crystal form has a single molecule in the asymmetric unit, with the molecules following a helical packing pattern along the 6_5 axis. Translation of the molecules in steps of ~ 6.5 Å along the hexagonal axis yields a hexameric ring that superimposes perfectly on the hexameric rings present in the orthorhombic and monoclinic crystal forms. The helical packing of the molecules is identical in the crystals of wild-type and D508A mutant, although the whole assembly is rotated by $\sim 23^\circ$ along the crystallographic c axis between these two non-isomorphous structures, leading to very different intermolecular contacts. These observations suggest that the hexameric ring is not an artifact of crystal packing, but rather a biologically significant unit, since it is present in different crystal types. Although the P-domain of *AfLonB* is monomeric in solution at low concentration (unpublished data), the assembly into the observed hexameric rings may precede crystallization at a high concentration of protein. The diameter of the hexamer is $\sim 80/100$ Å (narrowest/widest section), with a thickness of ~ 47 Å. The diameter of the central pore formed by the hexameric ring is ~ 20 Å, with a pore surface area of ~ 3420 Å². The present hexamer is very similar to that reported for the protease domain of *EcLonA*,³ whereas only a dimer with head-to-head orientation was seen in the crystals of *MjLonB*.²⁹ In particular, charge distribution on the surface of the hexamer of *AfLonB* P-domain (Figure 4(b)) has characteristics very similar to the distribution described for *EcLonA*.³ The residues involved in the oligomerization interface are located on strands S2–S4 on one side of the monomer and helices H1, H4 and strands S5, S7, and S8 on the opposite side of the molecule. The sheet formed by strands S2–S4 provides a relatively flat interaction surface. The total accessible surface area of the hexameric ring is

the sequence similarity (magenta) and identity (green), whereas the bottom part emphasizes the secondary structural elements: α -helices (H) are colored red, 3_{10} helices (H) are magenta and strands (S) are blue; the catalytic dyad is shown in gray highlight; the residues that are not visible in the electron density are italicized. The numbering of the secondary structure elements follows the pattern previously introduced for *EcLonA*.³ (b) The structure of the *AfLonB* P-domain (blue) compared with *EcLonA* (red) and *MjLonB* (green). The C-terminal helix H7, present in *AfLonB* only, was omitted for clarity. The secondary structure elements of the regions exhibiting the highest temperature factors in all three structures are marked. The inset shows the superposition of the active-site residues, with analogous color coding (numbering corresponds to *AfLonB*). The active site of *EcLonA* represents an inactive mutant in which the catalytic serine residue was replaced by alanine.³

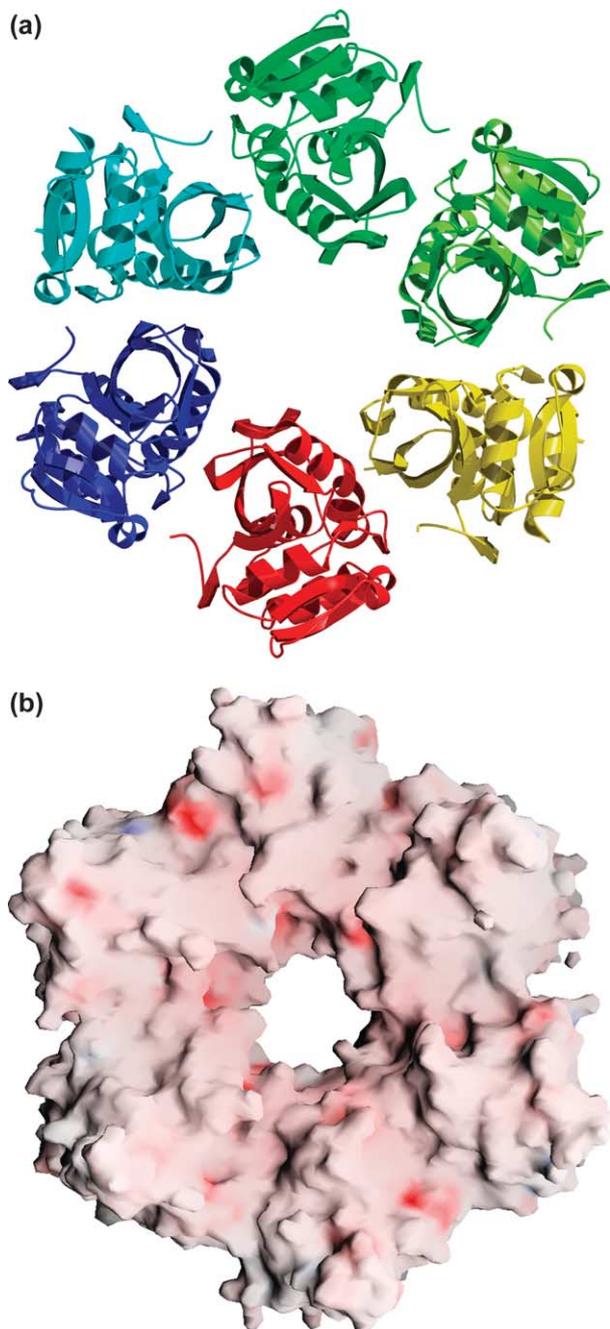


Figure 4. Oligomeric structure of the P-domain of *AfLonB*. (a) A hexamer as seen in the monoclinic crystal form. Each of the crystallographically independent molecules is colored differently. (b) Distal view of the hexamer molecular surface with the surface charge distribution (the Figure was generated with program GRASP: <http://trantor.bioc.columbia.edu/grasp/>). The depicted surface contains the active sites of the enzyme and is comparable directly to the *EcLonA* charge distribution from Figure 2(a) of the previous paper.³

50,869 Å², which is 3245 Å² more than that of the hexamer of the *EcLonA* P-domain. The surface area buried upon oligomerization is 1555 Å² per monomer, which is 164 Å² more than in *EcLonA*.

Comparison with proteolytic domains of other Lon proteases

The proteolytic domains of *AfLonB*, *EcLonA*³ and *MjLonB*²⁹ were superimposed in order to compare their structural and functional characteristics. A structure-based sequence alignment was created on the basis of this superposition (Figure 3(a)), which allowed us to indicate the distribution of the secondary structure elements along the sequences in a comparative way, and at the same time to reveal the level of sequence similarity between the three proteins with respect to their structural assignment. Despite the high overall level of sequence similarity within the family of Lon proteases (Table 3), *AfLonB* and *EcLonA* clearly display consensus sequence fragments typical for their subfamilies, especially in the vicinity of the catalytic Ser and Lys residues.⁴ The level of sequence similarity between either *Af*- or *MjLonB* and *EcLonA* P-domains is ~54%. However, within the LonB family, the *MjLonB* sequence differs significantly from any other enzyme. These differences include, for example, the presence of Asp in position (-3)Ser* and Glu in position (-1)Lys*; a ten amino acid residue insertion between Gly486 and Arg487 (*AfLonB* numbering); and a number of substitutions among the residues that are strictly conserved in LonB subfamily, such as Pro444, Gly464, Ser511, Thr516, Glu523, Ser538, Thr550, Glu506, and others. That affects the level of sequence similarity between *AfLonB* and *MjLonB*, which is only ~60% versus 75–77% for any other two members of the subfamily (*MjLonB* excluded).

Superposition of the C α traces of the P-domain of *AfLonB* onto *EcLonA* and *MjLonB* shows an identical overall fold, with r.m.s. deviations of 1.71 Å and 1.46 Å, respectively (Figure 3(b)). The P-domain conformations are significantly different in four regions, whose locations correlate with the highest temperature factors in all three structures (Figure 3(b)). Three of them are loops connecting secondary structure elements: strands S2 and S3, helix H1 and strand S4, and helices H4 and H5. Structure-based sequence alignment (Figure 3(a)) indicates that all of these loops contain deletions and insertions. The largest insertion is in the loop H1-S4 of *MjLonB*, which is the shortest in *AfLonB*. Strands S2 and part of S3 (Figure 3(a)), have some of the highest temperature factors in all three structures, suggesting considerable flexibility. Because of that, residues 454–456 on the tip of the loop S2–S3 in *AfLonB* are disordered and could not be traced. Helix H5 is present only in *EcLonA*, while in the two other structures only a helical turn is formed at the corresponding position.

Surprisingly, the largest differences are located around the active-site region, which includes the loop connecting strands S1 and S2, as well as the strand S5, preceding the catalytic Ser509. Both the loop S1–S2, which has a different conformation and location in all three structures, and strand S5 (corresponding to part of helix H2 in *MjLon*) seem

Table 3. Sequence homology between P-domains of Lon

	Identity	Similarity	Identity + similarity
All three	35	44	79 (35 + 44)
<i>AfLonB</i> / <i>MjLonB</i>	66 (35 + 31) ^a	56 (44 + 12)	122 (66 + 56)
<i>AfLonB</i> / <i>EcLonA</i>	57 (35 + 22)	53 (44 + 9)	110 (57 + 53)
<i>MjLonB</i> / <i>EcLonA</i>	54 (35 + 19)	57 (44 + 13)	111 (54 + 57)

The homology is evaluated by the number of identical/similar residues.

^a Sum of identities in all three enzymes, plus in this pair only.

to be very flexible in all three structures. It is known that the enzymatic activity of the proteolytic domain is regulated in an allosteric manner by conformational changes of the ATPase domain upon ATP binding and hydrolysis.^{34,35} Since strand S2 is located quite close to the N terminus of the P-domain, it might move as a result of the conformational change in the ATPase domain. The motion of this loop can affect the conformation of the strand S5 and the orientation of the catalytic serine and other adjacent residues. This hypothesis is consistent with the observation that the full-length enzymes show ATP-dependent proteolytic activity (Figure 5), while isolated proteolytic domains are inactive (data not shown). The lack of activity may result from non-productive conformation of the strands S2 and S5 when the loop S1–S2 is not connected to the ATPase domain.

The positions of the catalytic Lys552 and of the adjacent Thr534 (whose function is not yet clear) are very well conserved among the three structures (Figure 3(b)). However, the loop containing Ser509 adopts a completely different conformation compared to the other Lon structures, being flipped away from Lys552. The remote position of the catalytic serine residue suggests that the enzyme in

this conformation must be proteolytically inactive. The expected position of the catalytic serine residue is occupied by Asp508, which participates in several interactions with Lys552 and Thr534. Removal of the functional group in the D508A mutant does not change the local structure (although it leads to increased flexibility of this region), showing that these interactions are not required for adopting this conformation. In the *MjLonB* structure, the catalytic serine residue (Ser550) is in its expected position but the side-chain of another aspartate residue (Asp547) points into the active site, forming interactions with the catalytic lysine residue and an active-site water molecule. This observation led to postulating the presence of a catalytic triad rather than dyad in *MjLonB*.²⁹ However, it should be pointed out that Asp547 of *MjLonB* is not universally conserved in LonB sequences and corresponds to Glu506 in the *AfLonB*.

Enzymatic activity

The P-domain of *AfLonB* exhibited no detectable activity towards protein substrates under any conditions that we tested (different temperature, pH and reaction time; data not shown), so the observations described above could not be correlated directly with the enzymatic activity. However, full-length wild-type *AfLonB* exhibited considerable ATP-dependent proteolytic activity when assayed at 70 °C (Figure 5). Site-directed mutagenesis of a number of amino acid residues located in the vicinity of the active site was carried out in order to study the role of these residues in the enzymatic mechanism and to explain the structural results obtained for the isolated P-domain. The catalytic residue Ser509, as well as the adjacent Glu506 and Asp508 were mutated in turn to alanine. The results of the ATP-dependent proteolytic activity assays (Figure 5) show clearly that only the S509A mutant is completely inactive. It exhibited no detectable proteolytic activity at 70 °C, or after decreasing the temperature to 37 °C, or even upon significant extension of the reaction time (data not shown). These results are in agreement with the lack of both proteolytic and peptidase activities reported for the equivalent mutant of *TaLonB*, as well as for a Lys552-equivalent mutant.²⁸ On the other hand, the mutants D508A and E506A retained considerable level of activity (Figure 5). Some decrease of ATP-dependent proteolytic activity of *AfLonB*(E506A) may have

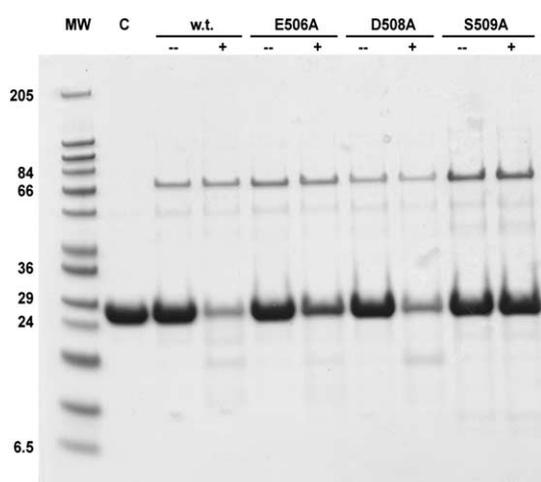


Figure 5. Proteolytic activity of *AfLonB* and its mutants. The assay was carried out at pH 8.0 in the presence of 1 mg/ml of β -casein (C); 0.1 mg/ml of *AfLonB* (wild-type, E506A, D508A) or 0.2 mg/ml of *AfLonB* (S509A) at 70 °C for 40 minutes, in the presence (+) or in the absence (–) of 5 mM ATP (see Materials and Methods). MW denotes a lane corresponding to wide-range molecular mass marker kit (Sigma).

resulted from impaired activation of the protease domain by the ATPase, and/or from a decrease of thermostability of the P-domain. In view of the interactions between Asp508 and Lys552 observed in the structure of the wild-type P-domain, it is interesting to note that mutation of Asp508 to alanine did not affect the activity of the full-length enzyme in any noticeable way.

Discussion

The crystal structure of the proteolytic domain of *AfLonB* has revealed similar hexameric arrangement of the proteolytic domains of *AfLonB* and *EcLonA*, indicating that isolated P-domains of single-chain Lon proteases are capable of assembling a functional oligomeric state. Despite the similarity of the overall fold, significant differences were found in the conformation of the active-site residues when compared to those in two other structures of Lon P-domains, *MjLonB* and *EcLonA* (Figure 3(b)). In *AfLonB* the polypeptide chain makes a sharp turn in the immediate vicinity of Ser509, which orients it away from Lys552 and prevents formation of the putative Ser-Lys dyad. Instead, the adjacent Asp508, brought by this turn into the position occupied by an equivalent Ser residue in *MjLonB* and *EcLonA*, is hydrogen bonded to Lys552 and Thr534, forming a "triad-like" structure similar to that postulated for *EcLonA*,³ with the difference that Asp508 substitutes for the catalytic serine residue. In the active site of *MjLonB*, the putative triad includes both the catalytic serine and lysine residues, as in *EcLonA*, but the third counterpart is Asp547, structurally equivalent not to Asp508, but to Glu506 in *AfLonB* (Figure 3(a)). Taken at face value, this result could indicate that despite the overall identical fold, the catalytic mechanism of *AfLonB* is completely different from both mechanisms proposed for the other Lon enzymes, yet for reasons presented below, we are unwilling to accept such a conclusion.

Although the structures presented and discussed here are clearly correct in crystallographic terms, the possibility that the conformation of the active site might not correspond to the productive state during catalysis should not be disregarded. While such phenomena are not very common, they have been well documented for several enzymes. In some cases, the origin of an "inactive" conformation can be traced to covalent modifications with components of the crystallization buffers, such as cacodylate, which was responsible for the inability to bind divalent cations by the catalytic domain of HIV integrase.^{36,37} In some other cases, the reasons for rearrangements of the active site might be more subtle. An illuminating example is provided by several crystal structures of β -carbonic anhydrase, in which the catalytic Zn^{2+} is coordinated by three active-site side-chains of the enzyme, leaving the fourth coordination site accessible to the substrate.^{38,39} However, structures of closely related

enzymes from other sources indicated the presence of a fourth coordinating residue, an aspartate, leading to a postulate of a significantly different mechanism of action.⁴⁰ As was noted subsequently, that structure, together with another very similar one, might not describe the catalytically competent enzyme.⁴¹ The additional interaction of the fourth residue with the metal ion was caused by a flip of the polypeptide chain, which brought a residue that pointed away in the two other structures to the immediate vicinity of the catalytic zinc cation and forced an adjacent serine to point away. It was suggested that these structural differences could be related to pH-dependent autoinactivation of the enzyme.⁴¹

The differences in the interactions of the catalytic residues in the active sites of *AfLonB* and the other two Lon proteases are connected primarily to the variable conformational state of the segment that precedes Ser509. It appears that in isolated P-domains this segment does not have a stable conformation that could maintain a proper structure of the active site. This was not obvious before the three structures of Lon P-domains became available, since Lon is the first structurally characterized single-chain ATP-dependent protease. All other ATP-dependent proteases with known structures, such as, for instance, HslUV or ClpAP, are two-chain enzymes, and in their independently folded proteolytic subunits, HslV⁴² or ClpP,⁴³ respectively, the catalytic residues are in appropriate positions. It seems justified to suggest that the interactions with other domains (the ATPase domain in particular), as well as ligand binding, might lead to rearrangements in Lon P-domains. Similarly to carbonic anhydrase, a simple reverse turn of a polypeptide chain could regenerate the "active" conformations of the catalytic residues in *AfLonB*. It should be pointed out that both *AfLonB* and *MjLonB* are thermophilic enzymes, while their crystals were grown at ambient temperature. A temperature-dependent conformational switch, analogous to that reported for the chaperone/protease DegP (HtrA),⁴⁴ should not be ruled out.

The putative catalytic triad in the active site of *MjLonB*, the finding of which led to the consideration of a catalytic mechanism quite distinct from that of *EcLonA*,²⁹ might be formed due to an even more subtle modification of the active site. The presence of an ion pair between Lys593 and Asp547 in a triad makes it difficult to explain the ability of these residues to activate the serine hydroxyl group for nucleophilic attack. Mutagenesis data on *AfLonB* presented here indicate that the presence of a carboxylate group in the residue equivalent in sequence (but not in structure) to Asp547 of *MjLonB* is not necessary for the activity of *AfLonB*. In the absence of convincing data to the contrary, it is safest to assume that enzymes that belong to a single family should employ a similar mechanism of action.

The results of activity assays of full-length

AfLonB support the presence of a catalytic dyad (Ser509 and Lys552) in this enzyme, similar to that described in the A-type Lon, whereas they do not support any significant involvement of Asp508 or Glu506 in the catalytic activity. The question of whether the observed, presumably inactive, conformation of an isolated P-domain was due to crystallization conditions, or is present in solution, can be answered by utilizing a different technique such as NMR, or by crystallizing a different construct of AfLonB. The conditions under which a functional Ser-Lys dyad could be formed still have to be elucidated, and it is necessary to identify other residues that might be involved directly in catalysis.

Materials and Methods

Cloning and mutagenesis of the *Af-lon* gene

Full-length cDNA encoding AfLonB has been cloned by a PCR approach into expression plasmid pET24(+). A modification of the QuikChange[®] site-directed mutagenesis procedure developed by Stratagene (La Jolla, CA) was employed to introduce mutations E506A, D508A, and S509A. Briefly, the plasmid containing the wild-type *Af-lon* gene was subjected to 25 rounds of PCR amplification using *PfuI* Turbo polymerase (Stratagene) with the following oligonucleotide primer pairs (for, forward; rev, reverse; nucleotides altered to introduce a desired change of amino acid are underlined):

E506A.for: 5'-ACCTACGAGGGTGTTCCGGGCGACT
CTGCAAGC-3'
E506A.rev: 5'-TGCTTGCAGAGTCGCCCGCAACA
CCCTCGTAGG-3'
D508A.for: 5'-GAGGGTGTTCAGGGCGCCTCTG
CAAGCATAAGC-3'
D508A.rev: 5'-TGCTTATGCTTGCAGAGGCGCCCT
CAACACCCTCG-3'
S509A.for: 5'-GGTGTTCAGGGCGACGCTGCAAG
CATAAGCATTGC-3'
S509A.rev: 5'-AATGCTTATGCTTGCAGCGTCGCCC
TCAACACC-3'

The reaction mixture was treated with DpnI restriction endonuclease to reduce the background of parental plasmid followed by transformation into competent *E. coli* cells. A C-terminal PmaCI/HpaI fragment encompassing all five targeted amino acids and harboring a specific mutation was subsequently swapped with the analogous wild-type fragment in the parental vector to reduce the possibility of errors caused by PCR amplification. All final mutant vectors were sequenced, confirming the substitutions and correct amino acid exchange.

Expression and purification of full-length AfLonB protease and its mutant forms

The full-length proteins were expressed in the *E. coli* strain Rosetta(DE3)pLysS (Novagen, Madison, WI), using plasmid constructs pET24a(+)-lonAf carrying genes encoding either wild-type enzyme or its mutants listed above. Transformed cells were grown overnight at 37 °C in LB medium containing 50 µg/ml of kanamycin and 30 µg/ml of chloramphenicol, re-passed to a larger volume of medium and induced with 1 mM IPTG at

A₆₀₀ ~0.5. Cells were harvested by centrifugation after two hours of induction and stored at -80 °C.

Thawed cells (15 g of paste obtained from a 5 l culture) were suspended in 135 ml of buffer A (50 mM potassium phosphate (pH 7.0), 1 mM EDTA, 1 mM DTT). The cell suspension was divided into five portions of approximately 30 ml each, disrupted by sonication and the homogenate was centrifuged at 40,000g for two hours at 4 °C. The supernatant was filtered through a 0.45 µm pore-size cellulose acetate membrane and subjected to chromatography with P-11 cellulose phosphate (Whatman Inc., Clifton, NJ).

The P-11 resin was prepared according to the manufacturer's protocol. The cell-free extract was diluted twofold with buffer A1 (buffer A containing 10% (v/v) glycerol) and applied at 1 ml/minute to a 100 ml P-11 column pre-equilibrated with buffer A1. The column was washed with 1 l of buffer A1, and the protein was eluted with 400 ml of buffer containing 300 mM potassium phosphate (pH 8.0), 10% (v/v) glycerol, 1 mM EDTA and 1 mM DTT.

The protein solution was diluted fourfold with buffer B (50 mM Tris-HCl (pH 7.5), 10% (v/v) glycerol, 1 mM DTT), filtered through a 0.2 µm pore-size cellulose acetate membrane and loaded onto three 5 ml HiTrap Q-Sepharose HP columns (Amersham Biosciences, Piscataway, NJ) connected in series and pre-equilibrated with buffer B. Columns were washed with 150 ml of buffer B1 (buffer B containing 1 mM *n*-dodecyl-β-D-maltopyranoside (DDM)) and the protein was eluted with a 225 ml 0 M-1 M linear gradient of NaCl in buffer B1. Fractions eluted within 0.3 M-0.6 M NaCl were pooled, diluted twofold with buffer B1 and loaded onto a 5 ml HiTrap heparin-Sepharose HP column (Amersham Biosciences, Piscataway, NJ) equilibrated with buffer B1 containing 200 mM NaCl. The column was washed with 50 ml of the equilibration buffer and the protein was eluted with a 75 ml linear 0.2 M-1 M gradient of NaCl in buffer B1. Fractions eluted within 0.3 M-0.6 M NaCl were pooled, concentrated on a YM10 filter (Millipore, Bedford, MA) and used for size-exclusion chromatography on a HiLoad 16/60 Superdex 200 pg column (Amersham Biosciences, Piscataway, NJ) equilibrated with buffer C (150 mM potassium phosphate (pH 8.0), 1 mM DDM) or with a buffer containing 20 mM Tris-HCl (pH 7.5), 200 mM NaCl and 1 mM DDM.

Limited proteolysis of intact AfLonB and purification of the P-domain

Purified AfLonB (50 mg) was cleaved with 1 mg of bovine α-chymotrypsin (Sigma, St Louis, MO) in 50 ml of buffer C at 30 °C. After two hours of incubation, the reaction was stopped by adding β-mercaptoethanol to 1% (v/v) and phenylmethylsulfonyl fluoride (PMSF) to 1 mM. The reaction solution was cooled to 4 °C, diluted fourfold with buffer D (50 mM Hepes, pH 7.0), filtered through a 0.45 µm pore-size cellulose acetate membrane and loaded onto a 5 ml HiTrap heparin HP column equilibrated with buffer D. The unbound protein (flow-through) was diluted again fivefold with the same buffer, concentrated on a YM10 membrane to 200 ml and loaded onto a 5-ml HiTrap Q-Sepharose column equilibrated with buffer D. The column was washed and the protein was eluted with a 100 ml linear 0 M-0.5 M gradient of NaCl in buffer D. Fractions eluted within 0.1 M-0.2 M NaCl were pooled, concentrated on a YM10 membrane and applied onto a HiLoad 16/60 Superdex 75 pg column

(Amersham Biosciences, Piscataway, NJ) equilibrated with 20 mM Tris-HCl (pH 7.5), 0.2 M NaCl.

All protein isolation and purification procedures were performed at 4 °C. Samples were monitored for purity by SDS-PAGE on 4–12% NuPage gels (Invitrogen, Carlsbad, CA). Protein concentration was estimated with a Bio-Rad Protein Assay (Bio-Rad Laboratories GmbH, Germany) using bovine serum albumin as standard. The purity and homogeneity of the target fragments were verified by N-terminal sequencing and electrospray mass spectrometry (Agilent 1100). The P-domain (AFPP) was found to comprise the C-terminal residues K417–V621 of the full-length AfLonB.

Cloning, expression, and purification of the P-domain

Two versions of the P-domain of AfLonB, K417-V621 (AfLonB P domain, recombinant 1, abbreviated AFPR1) and D415-V621 (AfLonB P domain, recombinant 2, abbreviated AFPR2), were amplified from the full-length wild-type gene by PCR using the following oligonucleotide primers:

LonK.for: 5'-GAGAACCTGTACTTCCAGAAGCTCTT CATCACGGAAGGATAC-3'

Lon.rev: 5'-GGGGACCACCTTTGTACAAGAAAGCTG GGTATTAGACTGCGCAAGCTCAAGCTC-3'

for K417-V621, and

LonD.for: 5'-GAGAACCTGTACTTCCAGGACTATA AGCTCTTCATCACG-3'

and Lon.rev for D415-V621. The PCR amplicons were used subsequently as templates for a second PCR reaction with the following primers:

Lon2.for: 5'-GGGGACAAGTTTGTACAAAAAAGC AGGCTCGGAGAACCTGTACTTCCAG-3'

and Lon.rev.

The amplicons from the second PCR were inserted by recombinational cloning into the entry vector pDONR201 (Invitrogen, Carlsbad, CA), and the nucleotide sequences were confirmed experimentally. The open reading frame encoding the P-domains and containing a recognition site for tobacco etch virus (TEV) protease at their N terminus (ENLYFQ/K for K417-V621 and ENLYFQ/D for D415-V621), was moved by recombinational cloning into the destination vector pDEST-HisMBP to produce plasmids pJT3 and pJT4. pJT3 and pJT4 direct the expression of the K417-V621 and the D415-V621 P-domains of AfLonB, respectively, as N-terminal fusions with *E. coli* maltose-binding protein (MBP) and with an intervening TEV protease recognition site. The MBP contained an N-terminal hexahistidine tag for affinity purification by immobilized metal affinity chromatography. Two other constructs, pJT11 and pJT12, that express the AFPR2(E506A) and AFPR2(D508A) P-domains as MBP fusion proteins, respectively, were prepared in the same way as AFPR2 using full-length AfLon(E506A) or AfLon(D508A) as templates. The fusion proteins were expressed in the *lon*-deficient *E. coli* strain BL21(DE3) pRIL (Stratagene, La Jolla, CA). Cells containing the expression plasmid were grown to mid-log phase ($A_{600} \sim 0.5$) at 37 °C in LB medium containing 100 µg/ml of ampicillin, 35 µg/ml of chloramphenicol and 0.2% (w/v) glucose. Overproduction of the fusion proteins was induced with IPTG at a final concentration of 1 mM for four hours at 30 °C. The cells were pelleted by centrifugation and stored at -80 °C. Selenomethionine-substituted Lon P-domain was produced using the saturation of the methionine biosynthetic pathway.⁴⁵

Cell paste was suspended in buffer E (25 mM Hepes (pH 8), 100 mM NaCl, 25 mM imidazole), and disrupted

with an APV Gaulin model G1000 homogenizer at 10,000 psi (1 psi \approx 6.9 kPa). The homogenate was centrifuged at 30,000g for 30 minutes, the supernatant filtered through a 0.45 µm pore-size cellulose acetate membrane and applied onto a 25 ml Ni-NTA Superflow column (Qiagen, Valencia, CA) equilibrated with buffer E. The column was washed and eluted with a linear 25–200 mM gradient of imidazole. Fractions containing the recombinant fusion proteins were pooled and incubated overnight at 4 °C with hexahistidine-tagged TEV protease. The digest was diluted fourfold with 25 mM Hepes (pH 8), 100 mM NaCl and applied onto a 50-ml Ni-NTA Superflow column equilibrated with buffer E. The column effluent containing recombinant Lon proteolytic domain was concentrated using a YM10 membrane (Millipore Corporation, Bedford, MA) and applied onto a HiPrep 26/60 Sephacryl S-100 HR column (Amersham Biosciences, Piscataway, NJ) equilibrated with 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. Fractions containing the Lon P-domain were pooled and concentrated as described above.

In all cases, the final product was judged to be >95% pure on the basis of silver staining after SDS-PAGE. Protein concentrations were estimated spectrophotometrically at 280 nm using molar extinction coefficients of 3840 M⁻¹ cm⁻¹ for AFPR1 and 5120 M⁻¹ cm⁻¹ for AFPR2, AFPR2(E506A), and AFPR2(D508A). The molecular masses of the recombinant Lon P-domains were confirmed by electrospray mass spectrometry. The mass spectra confirmed that for the selenomethionine-containing samples, the Se-Met substitution was over 99%.

Enzymatic assay

Proteolytic activities of the full-length, wild-type AfLonB and of its mutants towards protein substrate (β -casein) were analyzed by SDS-PAGE. β -Casein (100 µg; Sigma) was incubated with the enzyme (10–50 µg) in 100 µl of BTP (50 mM Bis-Tris-propane, pH 6.5–9.5 at 25 °C), 0.1 M NaCl in the presence or in the absence of 5 mM ATP and 20 mM MgCl₂ at 70 °C for various times. Aliquots (10 µl) of the reaction mixture were analyzed by SDS-PAGE after staining with Coomassie brilliant blue R. The actual pH value at high temperature was calculated by using the formula $\Delta pK_a / \Delta T$ (°C) = -0.015 for BTP.

Protein crystallization

Initial screening for crystallization conditions⁴⁶ of the native proteolytically produced P-domain (AfLonB P-domain proteolytic, abbreviated AFPP) was carried out by the sitting-drop, vapor-diffusion method, using the Hampton (Hampton Research, Laguna Niguel, CA) and Wizard (Emerald Biostructures, Bainbridge Island, WA) screens with a HydraII-Plus One (Apogent Discoveries, Hudson, NH) crystallization robot. The native P-domain was concentrated to 12 mg/ml and the Se-Met form to 16 mg/ml. A single crystallization condition was identified and optimized in a larger-scale, hanging-drop, vapor-diffusion experiment. Native Lon, whether obtained proteolytically (AFPP) or recombinant (AFPR1), crystallized under very similar conditions (20% PEG 8000, 0.2 M calcium acetate, 0.1 M Mes (pH 6.0), 22 °C). Plate-shaped crystals grew at room temperature to reach the size of 0.2 mm \times 0.2 mm \times 0.05 mm in two to three days. The crystals, characterized as unusual monoclinic/orthorhombic twins, are described in detail elsewhere.³³ The

longer recombinant form (AFPR2) and the E506A mutant crystallized under similar conditions, but yielded larger, hexagonal crystals at pH 6.5 in the presence of 40% PEG 300, 0.2 M calcium acetate and 0.1 M sodium cacodylate, with the largest crystals reaching the size of 0.8 mm × 0.2 mm × 0.2 mm in two weeks at 22 °C. The Se-Met derivative crystals of AFPR2 grew under the same conditions. The crystals of the D508A mutant were grown at pH 6.5 in 40% PEG 600, 0.2 M calcium acetate and 0.1 M sodium cacodylate. Before flash-freezing, the AFPP and AFPR1 crystals were transferred into a cryoprotecting agent consisting of 100% Paratone-N. Native, Se-Met and D508A crystals of the AFPR2 construct were frozen directly in their mother liquor.

Crystallographic procedures

X-ray diffraction data for the twinned monoclinic/orthorhombic crystals of AFPP and AFPR1 were collected on a MAR345 detector, using a Rigaku H3R rotating anode X-ray generator operated at 50 kV and 100 mA, with CuK α radiation focused by an MSC/Osmic mirror system.³³ A series of putative heavy-atom derivatives were tested, but without success. High-resolution native and Se-Met derivative data for AFPR2 were collected for the hexagonal crystals at the SER-CAT beamline 22-ID at the Advanced Photon Source (APS), Argonne, IL, on a MAR300 CCD detector. Data for the AFPR2(D508A) and AFPR2(E506A) mutants were collected on beamline 22-BM. All data were processed using the HKL2000 package (Table 1).⁴⁷ The highest-resolution data set was collected in two redundant passes and included data extending to 1.15 Å resolution, but only 1.2 Å resolution data were used in the refinement.

Structure solution and refinement

Single-wavelength anomalous dispersion (SAD) phasing⁴⁸ was carried out using the program SHARP⁴⁹ with 1.7 Å resolution data collected at the selenium absorption peak. Solvent flattening yielded an excellent map in the $P6_5$ space group and ARP/wARP⁵⁰ was able to automatically build a partial model consisting of 198 residues of the single protein molecule present in the asymmetric unit. The remaining residues were built manually into the initial map using program O.⁵¹ Initial rigid-body refinement with REFMAC5,⁵² part of the CCP4 suite,⁵³ using maximum-likelihood targets, was followed by several cycles of restrained positional refinement. The model was rebuilt and fitted into density using O. The high-resolution native structure was subsequently refined with SHELXL⁵⁴ to an R -factor of 13.7% and R_{free} ⁵⁵ of 18.0%, after addition of solvent and modeling alternate conformations for a number of side-chains (Table 2). Protein hydrogen atoms were added in riding positions and the temperature factors were refined anisotropically. Crystals of the E506A mutant were isomorphous with those of the wild-type enzyme. Although the crystallographic parameters for the D508A mutant (Table 1) suggested that the crystals should also be isomorphous with the wild-type, that was not the case, and the structure was solved by molecular replacement with AMoRe.⁵⁶ Both structures were refined with REFMAC5 (Table 2).

The crystals grown from the samples AFPP and AFPR1 exhibited an unusual mode of twinning, consisting of superimposed monoclinic and orthorhombic lattices. A detailed description of this phenomenon is provided

elsewhere.³³ The monoclinic and orthorhombic data were processed as separate data sets (Table 1). The orthorhombic structure was solved by molecular replacement with the program EPMR,⁵⁷ starting from the refined coordinates of the high-resolution hexagonal structure, whereas the monoclinic form was solved from the same starting point using AMoRe.⁵⁶ Each of these crystal forms contains six molecules in an asymmetric unit. Rigid body and positional refinement, including TLS parameters, was carried out with REFMAC5 (Table 2). The TLS parameters describe the thermal motion of a rigid body in terms of translation and libration matrices.⁵⁸

Protein Data Bank accession codes

The coordinates and structure factors have been deposited in the RCSB Protein Data Bank (PDB) with the accession codes given in Table 2.

Acknowledgements

We express our gratitude to Dr Zbyszek Dauter for help in collecting and analyzing the crystallographic data utilized in this study, and to Professor Mariusz Jaskólski for critical reading of the manuscript. This work was supported, in part, by a grant from the Russian Foundation for Basic Research (Project no. 02-04-48481) to T.V.R. and by the US Civilian Research and Development Foundation grant RB1-2505-MO-03 to T.V.R. and A.W. Part of diffraction data used in this study were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamlines 22-ID and 22-BM, located at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract W-31-109-Eng-38.

References

1. Rawlings, N. D., Tolle, D. P. & Barrett, A. J. (2004). MEROPS: the peptidase database. *Nucl. Acids Res.* **32**, D160–D164.
2. Rotanova, T. V., Melnikov, E. E. & Tsirulnikov, K. B. (2003). Catalytic dyad Ser-Lys at the active site of *Escherichia coli* ATP-dependent Lon-proteinase. *Bioorg. Khim.* **29**, 97–99.
3. Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F. *et al.* (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.
4. Rotanova, T. V., Melnikov, E. E., Khalatova, A. G., Makhovskaya, O. V., Botos, I., Wlodawer, A. & Gustchina, A. (2004). Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains. *Eur. J. Biochem.* **271**, 4865–4871.
5. 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.
6. Lupas, A. N. & Martin, J. (2002). AAA proteins. *Curr. Opin. Struct. Biol.* **12**, 746–753.
 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. Amerik, A. Y., Antonov, V. K., Ostroumova, N. I., Rotanova, T. V. & Chistyakova, L. G. (1990). Cloning, structure and expression of the full *lon* gene of *Escherichia coli* coding for ATP-dependent protease La. *Bioorg. Khim.* **16**, 869–880.
 9. Swamy, K. H. & Goldberg, A. L. (1981). *E. coli* contains eight soluble proteolytic activities, one being ATP dependent. *Nature*, **292**, 652–654.
 10. 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.
 11. Gill, R. E., Karlok, M. & Benton, D. (1993). *Mycococcus xanthus* encodes an ATP-dependent protease which is required for developmental gene transcription and intercellular signaling. *J. Bacteriol.* **175**, 4538–4544.
 12. Cloud, J. L., Marconi, R. T., Eggers, C. H., Garon, C. F., Tilly, K. & Samuels, D. S. (1997). Cloning and expression of the *Borrelia burgdorferi lon* gene. *Gene*, **194**, 137–141.
 13. Roudiak, S. G., Seth, A., Knipfer, N. & Shrader, T. E. (1998). The *lon* protease from *Mycobacterium smegmatis*: molecular cloning, sequence analysis, functional expression, and enzymatic characterization. *Biochemistry*, **37**, 377–386.
 14. Watanabe, S., Muramatsu, T., Ao, H., Hirayama, Y., Takahashi, K., Tanokura, M. & Kuchino, Y. (1999). Molecular cloning of the *lon* protease gene from *Thermus thermophilus* HB8 and characterization of its gene product. *Eur. J. Biochem.* **266**, 811–819.
 15. Robertson, G. T., Kovach, M. E., Allen, C. A., Ficht, T. A. & Roop, R. M. (2000). The *Brucella abortus lon* functions as a generalized stress response protease and is required for wild-type virulence in BALB/c mice. *Mol. Microbiol.* **35**, 577–588.
 16. Lee, A. Y., Hsu, C. H. & Wu, S. H. (2004). Functional domains of *Brevibacillus thermoruber lon* protease for oligomerization and DNA binding: role of N-terminal and sensor and substrate discrimination domains. *J. Biol. Chem.* **279**, 34903–34912.
 17. Kutejova, E., Durcova, G., Surovkova, E. & Kuzela, S. (1993). Yeast mitochondrial ATP-dependent protease: purification and comparison with the homologous rat enzyme and the bacterial ATP-dependent protease La. *FEBS Letters*, **329**, 47–50.
 18. Wang, N., Gottesman, S., Willingham, M. C., Gottesman, M. M. & Maurizi, M. R. (1993). A human mitochondrial ATP-dependent protease that is highly homologous to bacterial *lon* protease. *Proc. Natl Acad. Sci. USA*, **90**, 11247–11251.
 19. Lu, B., Liu, T., Crosby, J. A., Thomas-Wohlever, J., Lee, I. & Suzuki, C. K. (2003). The ATP-dependent *lon* protease of *Mus musculus* is a DNA-binding protein that is functionally conserved between yeast and mammals. *Gene*, **306**, 45–55.
 20. Goldberg, A. L. (1992). The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. *Eur. J. Biochem.* **203**, 9–23.
 21. Gottesman, S., Wickner, S. & Maurizi, M. R. (1997). Protein quality control: triage by chaperones and proteases. *Genes Dev.* **11**, 815–823.
 22. Rotanova, T. V. (1999). Structural and functional characteristics of ATP-dependent *lon*-protease from *Escherichia coli*. *Bioorg. Khim.* **25**, 883–891.
 23. Gottesman, S. (2003). Proteolysis in bacterial regulatory circuits. *Annu. Rev. Cell Dev. Biol.* **19**, 565–587.
 24. Rotanova, T. V. & Starkova, N. N. (2003). ATP-dependent protease *lon* from *Escherichia coli*. Structure, active site, coupling of proteolysis to ATP hydrolysis. In *Protein Structures. Kaleidoscope of Structural Properties and Functions* (Uversky, V. N., ed.), pp. 579–592, Research Signpost Publishers, Trivandrum, Kerala, India.
 25. Botos, I., Melnikov, E. E., Cherry, S., Khalatova, A. G., Rasulovala, F. S., Tropea, J. E. *et al.* (2004). Crystal structure of the AAA⁺ domain of *E. coli lon* protease at 1.9 Å resolution. *J. Struct. Biol.* **146**, 113–122.
 26. 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.
 27. Besche, H., Tamura, N., Tamura, T. & Zwickl, P. (2004). Mutational analysis of conserved AAA⁺ residues in the archaeal *lon* protease from *Thermoplasma acidophilum*. *FEBS Letters*, **574**, 161–166.
 28. Besche, H. & Zwickl, P. (2004). The *Thermoplasma acidophilum lon* protease has a Ser-Lys dyad active site. *Eur. J. Biochem.* **271**, 4361–4365.
 29. Im, Y. J., Na, Y., Kang, G. B., Rho, S. H., Kim, M. K., Lee, J. H. *et al.* (2004). The active site of a *lon* protease from *Methanococcus jannaschii* distinctly differs from the canonical catalytic dyad of *lon* proteases. *J. Biol. Chem.* **279**, 53451–53457.
 30. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
 31. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
 32. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucl. Acids Res.* **32**, W615–W619.
 33. Dauter, Z., Botos, I., LaRonde-LeBlanc, N. & Wlodawer, A. (2005). Pathological crystallography—case studies of several unusual macromolecular crystals. *Acta Crystallog. sect. D*, in the press.
 34. Melnikov, E. E., Tsurulnikov, K. B. & Rotanova, T. V. (2001). Coupling of proteolysis to ATP hydrolysis upon *Escherichia coli lon* protease functioning II. Hydrolysis of ATP and activity of the enzyme peptide hydrolase sites. *Bioorgan. Khim.* **27**, 120–129.
 35. Patterson, J., Vineyard, D., Thomas-Wohlever, J., Behshad, R., Burke, M. & Lee, I. (2004). Correlation of an adenine-specific conformational change with the ATP-dependent peptidase activity of *Escherichia coli lon*. *Biochemistry*, **43**, 7432–7442.
 36. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.
 37. Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R. & Davies, D. R. (1998). Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl Acad. Sci. USA*, **95**, 9150–9154.

38. Kimber, M. S. & Pai, E. F. (2000). The active site architecture of *Pisum sativum* beta-carbonic anhydrase is a mirror image of that of alpha-carbonic anhydrases. *EMBO J.* **19**, 1407–1418.
39. Strop, P., Smith, K. S., Iverson, T. M., Ferry, J. G. & Rees, D. C. (2001). Crystal structure of the “cab”-type beta class carbonic anhydrase from the archaeon *Methanobacterium thermoautotrophicum*. *J. Biol. Chem.* **276**, 10299–10305.
40. Mitsushashi, S., Mizushima, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Moriyama, H. *et al.* (2000). X-ray structure of beta-carbonic anhydrase from the red alga, *Porphyridium purpureum*, reveals a novel catalytic site for CO₂ hydration. *J. Biol. Chem.* **275**, 5521–5526.
41. Cronk, J. D., Endrizzi, J. A., Cronk, M. R., O’neill, J. W. & Zhang, K. Y. (2001). Crystal structure of *E. coli* beta-carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein Sci.* **10**, 911–922.
42. Bochtler, M., Ditzel, L., Groll, M. & Huber, R. (1997). Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **94**, 6070–6074.
43. Wang, J., Hartling, J. A. & Flanagan, J. M. (1997). The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell*, **91**, 447–456.
44. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M. & Clausen, T. (2002). Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature*, **416**, 455–459.
45. Doublet, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530.
46. Jancarik, J. & Kim, S. H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallog.* **21**, 916–924.
47. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
48. Dauter, Z., Dauter, M. & Dodson, E. (2002). Jolly SAD. *Acta Crystallog. sect. D*, **58**, 494–506.
49. de la Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494.
50. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
51. Jones, T. A. & Kjeldgaard, M. (1997). Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208.
52. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
53. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
54. Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319–343.
55. Brünger, A. T. (1992). The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
56. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
57. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Crystallog. sect. D*, **55**, 484–491.
58. Schomaker, V. & Trueblood, K. N. (1968). On the rigid-body motion of molecules in crystals. *Acta Crystallog. sect. B*, **24**, 63–76.
59. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134.
60. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.

Edited by R. Huber

(Received 18 April 2005; received in revised form 31 May 2005; accepted 2 June 2005)

Available online 22 June 2005