PROTEIN STRUCTURE REPORT

Crystal structure of the N-terminal domain of *E. coli* Lon protease

MI LI,[1,2,5] FATIMA RASULOVA,[3,5,6] EDWARD E. MELNIKOV,[4] TATYANA V. ROTANOVA,[4] ALLA GUSTCHINA,[1] MICHAEL R. MAURIZI,[3] AND ALEXANDER WLODAWER[1]

[1]Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, Maryland 21702-1201, USA
[2]Basic Research Program, SAIC-Frederick, Frederick, Maryland 21702, USA
[3]Laboratory of Cell Biology, National Cancer Institute, Bethesda, Maryland 20892, USA
[4]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow 117997, Russia

(RECEIVED July 28, 2005; FINAL REVISION July 28, 2005; ACCEPTED August 10, 2005)

Abstract

We report here the first crystal structure of the N-terminal domain of an A-type Lon protease. Lon proteases are ubiquitous, multidomain, ATP-dependent enzymes with both highly specific and non-specific protein binding, unfolding, and degrading activities. We expressed and purified a stable, monomeric 119-amino acid N-terminal subdomain of the *Escherichia coli* A-type Lon protease and determined its crystal structure at 2.03 Å (Protein Data Bank [PDB] code 2ANE). The structure was solved in two crystal forms, yielding 14 independent views. The domain exhibits a unique fold consisting primarily of three twisted β-sheets and a single long α-helix. Analysis of recent PDB depositions identified a similar fold in BPP1347 (PDB code 1ZBO), a 203-amino acid protein of unknown function from *Bordetella parapertussis*, crystallized as part of a structural genomics effort. BPP1347 shares sequence homology with Lon N-domains and with a family of other independently expressed proteins of unknown functions. We postulate that, as is the case in Lon proteases, this structural domain represents a general protein and polypeptide interaction domain.

Keywords: ATP-dependent proteases; protein domain; structural genomics

ATP-dependent Lon proteases are conserved in all living organisms and catalyze rapid turnover of short-lived regulatory proteins and many damaged or denatured proteins (Gottesman and Maurizi 1992). Bacterial Lons are required for survival from DNA damage, developmental changes induced by stress, and quorum sensing, whereas eukaryotic Lons are needed for mitochondrial maintenance and respiratory functions. Mammalian mitochondrial Lon, which is critically required for removal of oxidatively damaged proteins, has been implicated in age related loss of cellular functions. Lon degrades both specific proteins and unfolded polypeptides, although the basis of substrate recognition is poorly understood. In vivo, Lon's ability to interact with both single-stranded DNA and with polyphosphate may increase its access to a selective range of cellular proteins.

The enzymatic properties of *Escherichia coli* Lon protease (*Ec*Lon) have been extensively studied (Goldberg et al. 1994; Gottesman et al. 1997; Melnikov et al. 2000). Lon couples ATP hydrolysis to structural disruption and processive degradation of proteins into peptides of 5–12

[5]These authors contributed equally to this work.
[6]Present address: Laboratory of Microbial Pathogenesis, Navy Medical Research Institute, Silver Spring, MD 20910, USA.

Reprint requests to: Alexander Wlodawer, National Cancer Institute, MCL, Building 536, Room 5, Frederick, MD 21702-1201, USA; e-mail: wlodawer@ncifcrf.gov; fax: (301) 846-6128.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.051736805.

amino acids. *Ec*Lon is active as an oligomer of identical 784-amino-acid polypeptide chains (Botos et al. 2004b). Three functional domains have been identified within each subunit: the central region (A domain), an ATPase belonging to the AAA^+ superfamily (Neuwald et al. 1999); the C-terminal region (P domain), a unique serine protease with a serine-lysine catalytic dyad (Rotanova et al. 2004); and the ~300-amino-acid N-terminal region (N domain), which is divided into two or more subdomains and is expected by analogy with other ATP-dependent proteases to participate in recognition and binding of target proteins or their adaptors (Iyer et al. 2004; Rotanova et al. 2004). However, detailed analysis of the organization and structural interactions between the domains of Lon proteases is lacking. To date, crystal structures have been solved for a small α-helical portion of the *Ec*Lon A domain (residues 491–584) (Botos et al. 2004b) and for the P domains of *E. coli* (Botos et al. 2004a), *Methanococcus jannaschii* (Im et al. 2004), and *Archaeoglobus fulgidus* (Botos et al. 2005) Lons. By contrast, no structural data for the N domain of Lon have been available until now, and its lack of sequence similarity to any proteins of known structure has prevented homology modeling. This report provides initial data on the structure of the N domain of *Ec*Lon.

Results and Discussion

Possible boundaries within the N domain of *Ec*Lon were found by limited proteolysis, which gave fragments 1–209, 1–245, and 1–309. Analysis of multiple sequence alignments suggested that Lon N domains might consist of two or more independently folded subdomains with another boundary around residue 119. Several constructs consisting of Lon N domains of varying length were expressed in *E. coli* and the resulting proteins subjected to crystallization. Crystals of constructs 1–245 and 1–267 diffracted poorly, and structures could not be solved. However, better crystals of the construct 1–119 (Lon-N119) could be obtained, and its structure was solved by single-wavelength anomalous dispersion of the selenomethionine-containing protein. Two different crystal forms of the Lon-N119 construct, an orthorhombic one initially used to solve the structure and the monoclinic one used for refinement, were analyzed in this study. The orthorhombic structure did not refine well (Table 1), and in the discussion below, we refer to the structure obtained with the monoclinic crystals only, unless specifically mentioned otherwise.

Residues 8–117 of Lon-N119 are ordered in the monoclinic crystal structure refined at 2.03 Å resolution

Table 1. *Statistics of data collection and structure refinement*

	Native	SeMet
Data collection	Home	X9B BNL
Space group	$P2_1$	$P2_12_12_1$
Molecules/a.u.	8	6
Unit cell parameters (Å)	a = 90.61, b = 53.45, c = 111.97 β = 107.5°	a = 60.28, b = 87.35, c = 129.52
Resolution (Å)	20–2.03	20–2.45
Unique reflections	63,721 (Friedel merged)	48,706 (Friedel unmerged)
Redundance	4.1 (3.4)	3.3 (3.2)
Completeness (%)	95.8 (78.3)	99.8 (99.3)
Avg. I/σ	37.7 (3.0)	24.6 (2.4)
R_{merge} (%)[a]	4.0 (35.2)	4.2 (54.8)
Phasing statistics (20–2.45 Å)		
Phasing power, acentric	—	1.53
$R_{cullis\ ano}$ acentric	—	0.68
Refinement statistics		
R (%)[b]	21.7	28.0
R_{free} (%)[c]	27.2	35.8
RMSD bond lengths (Å)	0.016	0.09
angle (°)	1.8	1.51
Temp. factor (protein, Å^3)	36.2	50.97
(solvent, Å^3)	49.2	—
No. of protein atoms	6749	5056
No. of solvent molecules	672	—

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the observed intensity, and $\langle I\rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections after rejections.
[b] $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[c] R_{free} defined in Brünger (1992).

(Fig. 1A). Eight molecules in the asymmetric unit (labeled A–H) form two similar tetramers, each exhibiting ~222 symmetry. In each chain, two β-strands, β1 (9–16) and β2 (26–31), are followed by the sole prominent helix, α1 (34–45). Strand β3 (49–54) leads to a very irregular turn that varies considerably between different crystallographically independent molecules (Fig. 2). A helical turn α2 is made by residues 65–67. The remainder of the fragment is made of a long strand β4 (71–82), even longer β5 (88–105), which is slightly distorted in the middle in molecules D and E, and the terminal strand β6 (110–116). Strands β1, β3, β4, and part of strand β5 form a mixed β-sheet, while strands β2, the rest of strand β5, and strand β4 form an anti-parallel sheet almost exactly perpendicular to the former. Finally, a third anti-parallel sheet is formed by strands 1, 6, and 5. A turn involving residues 83–85 is also quite variable between different molecules (Fig. 2).

A search of the Protein Data Bank (PDB) entries by using the program DALI (Holm and Sander 1993) resulted in a single significant hit (Z = 13.6) for coordinates with PDB code 1ZBO. These coordinates were recently deposited by the Northeast Structural Genomics Consortium and represent a hypothetical protein from *Bordetella parapertussis*, BPP1347, annotated as having unknown function. That structure was solved at a resolution of 2.6 Å, and the asymmetric unit contains two identical molecules with 197 visible residues each. Superposition of the N domain of *Ec*Lon on BPP1347 with the program ALIGN (Cohen 1997) results in the root mean square deviation of 2.1 Å for 104 Cα pairs. Structure-based sequence alignment shows 22 identical residues in the aligned structures (19% identity). A similar degree of sequence homology can also be observed between *Ec*Lon and the C-terminal half of BPP1347 (Fig. 3), suggesting that this protein could be a good model for the N-terminal domain of *Ec*Lon from residues 1–209.

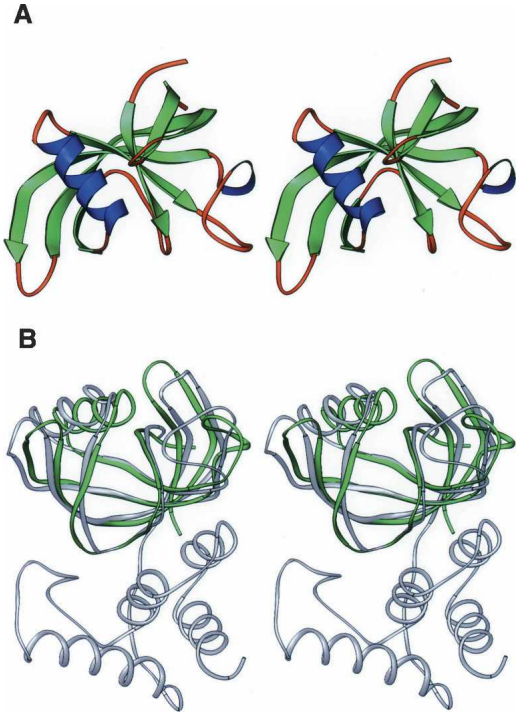

Figure 1. Crystal structure of the N119 domain of *E. coli* Lon and its comparison with the structure of the hypothetical protein BPP1347. (*A*) Stereo view of LonN119 showing the elements of the secondary structure colored green for β strands, blue for α helices, and brown for coils. (*B*) Superposition of LonN119 (green) and *B. parapertussis* BPP1347 (gray).

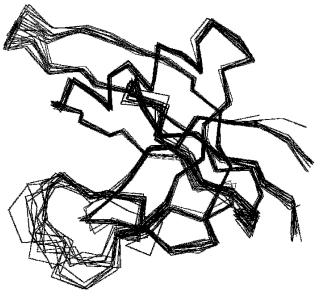

Figure 2. Superposition of the Cα traces for 14 crystallographically independent molecules of the N119 domain of *Ec*Lon. Eight molecules are derived from the asymmetric unit of the monoclinic crystals and six from the orthorhombic crystals. The region of largest deviation surrounds residues 83–85.

Each molecule of BPP1347 consists of two subdomains connected by a single extended linker. The N-terminal half has an overall fold practically identical to the Lon-N119 (Fig. 1B), although the detailed differences between the two structures prevented solving either of them by molecular replacement, using the other as a starting model (data not shown). The C-terminal fragment of BPP1347 contains four prominent α-helices with a topology that appears to be unique, since a DALI search found no structural homologs of this domain. Secondary structure prediction based on the sequences of >100 Lon proteins predicted a mostly helical structure for residues 119–209 of Lon, with turns or loops between helices at nearly the precise positions observed in the structure of BBP1347. Thus, we postulate that the N-terminal domain of *Ec*Lon is likely to have the same overall structure as BBP1347.

The genome sequence of *B. parapertussis* encodes two full-length homologs of Lon, BPP1777 and BPP1033, and there is little reason to question that BPP1347 is a standalone homolog of the Lon N domain. The coding region of BPP1347 is followed by another open reading frame not related to Lon proteases and has numerous

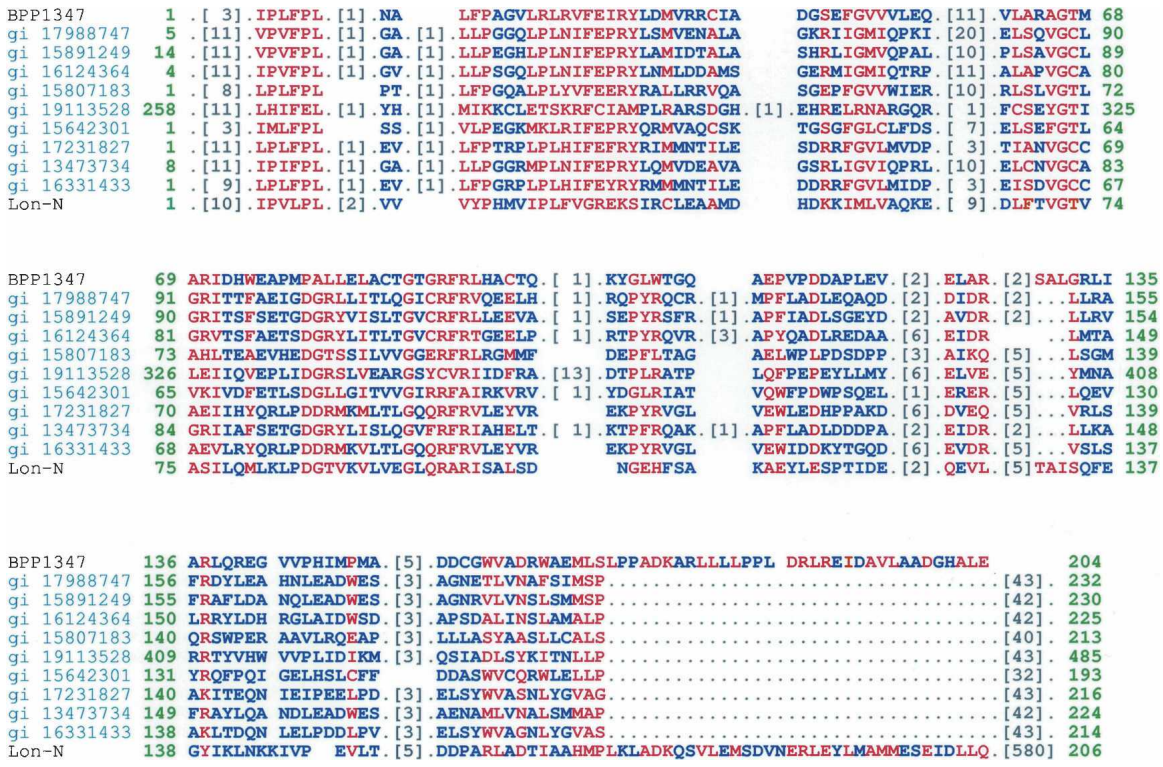

Figure 3. Sequence alignment for proteins belonging to COG2802 (Marchler-Bauer and Bryant 2004) compared with *E. coli* LonN209. Alignment was obtained by using the CD Search program at NCBI with the *B. parapertussis* BB1347 sequence as the query. A GAP sequence alignment of the N-terminal domain of *E. coli* line is included. Identical and highly similar residues between Lon and BP1347 are shown in red. This conserved domain is found at variable positions within the primary structure of different proteins in the family.

stop codons in the other reading frames. BPP1347 is representative of a large, diverse family of proteins or domains of larger proteins. A BLAST search using either *Ec*Lon-N209 or BPP1347 as a query revealed >100 hypothetical proteins that had significant homology to Lon N domain ($e = 10^{-60} - 10^{-6}$), but lacked AAA^+ or Lon P domains and thus could not be ATP-dependent proteases. The conserved domains database on the National Center for Biotechnology Information (NCBI) Web site (http://www.ncbi.nlm.nih.gov/) identified these proteins as members of COG2802, a conserved domain related to Lon N domain (Fig. 3). The regions of homology within the COG correspond to regions of similarity between *Ec*Lon-N209 and BBP1347.

It is not known if, and under what conditions, BBP1347 is expressed in vivo, as is the case for most proteins possessing this conserved domain. The N domain of *Ec*Lon apparently has protein binding ability, as shown by the ability of isolated N domains to inhibit Lon protease activity in vivo (F.S. Rasulova and M.R. Maurizi, unpubl.) and by decreased affinity of N domain-deleted forms of Lon for unfolded substrates such as casein (Rasulova et al. 1998; Roudiak and Shrader 1998). We speculate that the function of the Lon N–like regions in the more distantly related proteins and domains also involves similar binding activity directed at unfolded proteins or polypeptides. The combination of structural genomics and targeted structural studies (Wlodawer 2005) presented here has provided significant new information that should help in establishing the role of the N-terminal domain of Lon proteases and perhaps provide a means of investigating the function of a novel protein family.

Materials and methods

The N domain of *Ec*Lon (residues 1–119, LonN119) with a six-histidine N-terminal extension was cloned in pBAD33 in two steps. The PCR product obtained by amplifying pBAD33-lon DNA with the primers 5′-GCTCGGTACCCGGG GATCCTCTAGA-3′ and 5′-ACAGTCCG**CTCGAG**TCAC TACGACTCCAGATACTCCGCCTTCG-3′ was cut with XbaI and XhoI and inserted between XbaI and SalI restriction sites in pBAD33 to generate pBAD33-lon119. An NdeI linker encoding six histidines was then inserted at the NdeI site of this plasmid to generate pBAD33-hislon119, which was confirmed by DNA sequencing and Western blot analysis. Transformants of a lon^- derivative of MG1655 carrying pBAD33-hislon119

were induced with arabinose, and cells were harvested after 3 h. French pressure cell extracts were clarified by centrifugation and passed over a cobalt-chelate metal ion affinity column, and LonN119 was eluted with 0.2 M imidazole. The protein was concentrated by ammonium sulfate precipitation and run on a Superdex75 gel filtration column in 50 mM Tris (pH 7.5), 0.1 M KCl, and 10% (w/v) glycerol.

After concentration to 10 mg/mL by ultrafiltration using a Centriprep apparatus with a YM3 membrane, LonN119 was screened for crystallization conditions (Jancarik and Kim 1991) by the hanging drop, vapor-diffusion method (McPherson 1982), using the Hampton (Hampton Research) and Wizard (Emerald Biostructures) screening kits. Diffraction-quality crystals were grown in 30% PEG 3000, 0.1M CHES (pH 9.5), as well as in 20% PEG 8000, 0.1M CHES (pH 9.5). The largest crystals grew in 14 d at room temperature to the size of $0.4\times0.1\times0.05$ mm. Before flash freezing, the crystals were transferred into a cryoprotectant solution consisting of 80% mother liquor and 20% ethylene glycol.

The 2.4 Å data collected from an orthorhombic crystal of a SeMet derivative (Table 1) were measured on beamline X9B in BNL at the wavelength of 0.9788 Å and processed with HKL2000 (Otwinowski and Minor 1997). SHELXD/E was used to solve the substructure and the AUTO-SHARP (Global Phasing Ltd.) for phasing and refinement, followed by SOLOMON solvent flipping. The model was initially built automatically by the program RESOLVE (Terwilliger 2001), which located 441 residues—164 with side chains and 277 without. The rest of the model was built manually into the initial map by using the program O (Jones and Kjeldgaard 1997). Only preliminary refinement of this model was attempted (Table 1). The native X-ray data were collected from a monoclinic crystal on a Mar345 detector mounted on a Rigaku RU-200 rotating anode X-ray generator, operated at 50 kV and 100 mA. The Cu Kα radiation was focused by an MSC/Osmic mirror system. Native data were collected to 2 Å resolution and were processed and scaled with HKL2000 (Otwinowski and Minor 1997; Table 1). The native structure was solved by molecular replacement with program EPMR (Kissinger et al. 1999), with a search model consisting of a dimer created from monomers A and B of the orthorhombic structure. Initial rigid body refinement with CNS (Brünger et al. 1998), using maximum-likelihood targets, was followed by simulated annealing (Brünger et al. 1990) with Engh and Huber (1991) parameters. Final rounds of refinement were carried out with CNS, leading to a model with an R of 21.5% and R_{free} (Brünger 1992) of 27.2% for all data between 30 and 2.03 Å resolution. The Ramachandran plot for the final structure, obtained with the program PROCHECK (Laskowski et al. 1993), showed 88% of the residues in the core region, 9.5% in the additionally allowed region, and 2.5% in generously allowed region. The coordinates and the structure factors have been submitted to the PDB with accession code 2ANE for immediate release.

Acknowledgments

This work was supported in part by a grant from the Russian Foundation for Basic Research (project no. 05-04-48383) to T.V.R.; by the U.S. Civilian Research and Development Foundation grant to T.V.R. and A.W.; by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research; and with Federal funds from the National Cancer Institute, National Institutes of Health, under contract no. NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

References

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

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., et al. 2004b. 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.

Botos, I., Melnikov, E.E., Cherry, S., Kozlov, S., Makhovskaya, O.V., Tropea, J.E., Gustchina, A., Rotanova, T.V., and Wlodawer, A. 2005. Atomic-resolution crystal structure of the proteolytic domain of *Archaeoglobus fulgidus* Lon reveals the conformational variability in the active sites of Lon proteases. *J. Mol. Biol.* **351:** 144–157.

Brünger, A.T. 1992. The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355:** 472–474.

Brünger, A.T., Krukowski, A., and Erickson, J.W. 1990. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. A* **46:** 585–593.

Brünger, 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., et al. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Cohen, G.E. 1997. ALIGN: A program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.* **30:** 1160–1161.

Engh, R. and Huber, R. 1991. Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr. A* **47:** 392–400.

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

Gottesman, S. and Maurizi, M.R. 1992. Regulation by proteolysis: Energy-dependent proteases and their targets. *Microbiol. Rev.* **56:** 592–621.

Gottesman, S., Wickner, S., and Maurizi, M.R. 1997. Protein quality control: Triage by chaperones and proteases. *Genes & Dev.* **11:** 815–823.

Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233:** 123–138.

Im, Y.J., Na, Y., Kang, G.B., Rho, S.H., Kim, M.K., Lee, J.H., Chung, C.H., and Eom, S.H. 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.

Iyer, L.M., Leipe, D.D., Koonin, E.V., and Aravind, L. 2004. Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146:** 11–31.

Jancarik, J. and Kim, S.H. 1991. Sparse matrix sampling: A screening method for crystallization of proteins. *J. Appl. Crystallogr.* **21:** 916–924.

Jones, T.A. and Kjeldgaard, M. 1997. Electron-density map interpretation. *Methods Enzymol.* **277:** 173–208.

Kissinger, C.R., Gehlhaar, D.K., and Fogel, D.B. 1999. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D Biol. Crystallogr.* **55:** 484–491.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: Program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26:** 283–291.

Marchler-Bauer, A. and Bryant S.H. 2004. CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* **32:** W327–W331.

McPherson, A. 1982. *Preparation and analysis of protein crystals*. John Wiley and Sons, Inc., New York.

Melnikov, E.E., Tsirulnikov, K.B., and Rotanova, T.V. 2000. Coupling of proteolysis to ATP hydrolysis upon *Escherichia coli* Lon protease functioning, I: Kinetic aspects of ATP hydrolysis. *Bioorg. Khim.* **26:** 530–538.

Neuwald, A.F., Aravind, L., Spouge, J.L., and 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.

Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. **276:** 307–326.

Rasulova, F.S., Dergousova, N.I., Melnikov, E.E., Ginodman, L.M., and Rotanova, T.V. 1998. Synthesis and characterisation of ATP-dependent forms of Lon-proteinase with modified N-terminal domain from *Escherichia coli*. *Bioorg. Khim*. **24:** 370–375.

Rotanova, T.V., Melnikov, E.E., Khalatova, A.G., Makhovskaya, O.V., Botos, I., Wlodawer, A., and 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.

Roudiak, S.G. and Shrader, T.E. 1998. Functional role of the N-terminal region of the Lon protease from *Mycobacterium smegmatis*. *Biochemistry* **37:** 11255–11263.

Terwilliger, T.C. 2001. Map-likelihood phasing. *Acta Crystallogr. D Biol. Crystallogr*. **57:** 1763–1775.

Wlodawer, A. 2005. Giving credit where credit is due. *Nat. Struct. Mol. Biol*. **12:** 634.