

Crystal structure of the AAA⁺ α domain of *E. coli* Lon protease at 1.9 Å resolution

Istvan Botos,^a Edward E. Melnikov,^b Scott Cherry,^a Anna G. Khalatova,^b Fatima S. Rasuloval,^c Joseph E. Tropea,^a Michael R. Maurizi,^c Tatyana V. Rotanova,^b Alla Gustchina,^a and Alexander Wlodawer^{a,*}

^a Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, MCL Bldg. 536, Rm. 5, Frederick, MD 21702-1201, USA

^b Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow 117997, Russia

^c Laboratory of Cell Biology, National Cancer Institute, Bethesda, MD 20892, USA

Received 12 August 2003

Abstract

The crystal structure of the small, mostly helical α domain of the AAA⁺ module of the *Escherichia coli* ATP-dependent protease Lon has been solved by single isomorphous replacement combined with anomalous scattering and refined at 1.9 Å resolution to a crystallographic *R* factor of 17.9%. This domain, comprising residues 491–584, was obtained by chymotrypsin digestion of the recombinant full-length protease. The α domain of Lon contains four α helices and two parallel strands and resembles similar domains found in a variety of ATPases and helicases, including the oligomeric proteases HslVU and ClpAP. The highly conserved “sensor-2” Arg residue is located at the beginning of the third helix. Detailed comparison with the structures of 11 similar domains established the putative location of the nucleotide-binding site in this first fragment of Lon for which a crystal structure has become available.

© 2003 Elsevier Inc. All rights reserved.

Keywords: ATP-dependent proteases; Domain structure; Structure comparisons; Structure conservation; Substrate recognition

1. Introduction

Elimination of mutant and aberrant proteins, necessary to assure quality control in living cells, is often achieved by energy-dependent selective intracellular proteolysis. Analogous processes also play a key role in the rapid turnover of short-living regulatory proteins (Goldberg, 1992; Gottesman and Maurizi, 1992; Wickner et al., 1999). In prokaryotic cells and in the organelles of higher eukaryotes, energy-dependent proteolysis is accomplished by oligomeric ATP-dependent proteases, such as Lon, FtsH, ClpAP, ClpXP, and HslVU (Gottesman, 1996), whereas multicatalytic complexes (26 S proteasomes) perform that function in the cytosol of eukaryotic cells (Hershko and Ciechanover, 1992). The ATP-dependent proteases are composed of a

chaperone component or domain that specifically recognizes protein targets and couples ATP hydrolysis to unfolding and translocation of the polypeptide chains into the interior chamber of an associated protease domain where processive proteolysis takes place. These enzymes are members of the extended AAA⁺ family (ATPases Associated with a variety of cellular Activities), a group of proteins whose diverse activities often require the ATP-modulated assembly of oligomeric (often hexameric) rings. Besides being involved in selective proteolysis, AAA⁺ proteins participate in diverse cellular processes, including cell-cycle regulation, organelle biogenesis, vesicle-mediated protein transport, microtubule severing, membrane fusion, and DNA replication (Maurizi and Li, 2001; Ogura and Wilkinson, 2001; Patel and Latterich, 1998).

ATPase functional domains (also called AAA⁺ modules) (Schulz and Schirmer, 1979) of the AAA⁺ proteins consist of single polypeptide chains containing

* Corresponding author. Fax: +1-301-846-6322.

E-mail address: wlodawer@ncifcrf.gov (A. Wlodawer).

220–250 amino acids. Within the AAA⁺ superfamily, these functional domains are present either once or as repeats (Neuwald et al., 1999). Each AAA⁺ module consists of two structural domains (Schulz and Schirmer, 1979), the larger one at its N terminus (referred to as the α/β domain) and the smaller one at its C terminus (referred to as the α domain). The large domains are highly conserved and typically contain a Rossmann fold, common in many nucleotide-binding enzymes. The small domains typically contain three or four α -helices but vary in sizes and exhibit substantially greater structural variation (Hattendorf and Lindquist, 2002; Lupas and Martin, 2002).

The α/β domains of AAA⁺ proteins contain the Walker A and B nucleotide binding motifs (Walker et al., 1982) shared by other P-loop type ATPases. In addition, the α/β domain contains the “sensor-1” motif, consisting of a polar residue, which forms a hydrogen-bonding network crucial in positioning of a nucleotide-binding water molecule (Lenzen et al., 1998; Yu et al., 1998), as well as a highly conserved arginine within the Box VII motif that projects from one subunit into the active site of the neighboring subunit, acting as an “arginine finger” (Karata et al., 1999; Neuwald et al., 1999; Putnam et al., 2001).

Another motif, “sensor-2,” present in many but not all AAA⁺ proteins, is located near the start of helix 3 of the α domain. The sensor-2 residue, Arg or Lys, participates in binding and hydrolysis of the ATP, which binds in a crevice at the interface of the large and small domains (Hattendorf and Lindquist, 2002; Neuwald et al., 1999). Independently expressed α domains from Lon and several Clp family members are compactly and stably folded in vitro (Smith et al., 1999) and were found to bind to certain disordered regions of either C- or N-terminal peptide sequences of their specific substrates. For the latter reason, α domains have also been called “sensor- and substrate-discrimination (SSD) domains” (Smith et al., 1999). The binding properties of the α domains correlate with studies indicating that the AAA⁺ modules are involved in protein target selection and regulate the activities of the functional components of AAA⁺ proteins (Lupas and Martin, 2002; Maurizi and Li, 2001; Neuwald et al., 1999; Ogura and Wilkinson, 2001; Patel and Latterich, 1998; Wickner et al., 1999). However, additional sites within the entire AAA⁺ module are likely to contribute to target protein binding (Ebel et al., 1999; Leonhard et al., 1999; Ortega et al., 2000). In this paper, we will use the descriptive name, α domain, to refer to the small domain of the functional ATPase domain (AAA⁺ module).

Escherichia coli Lon protease was the first ATP-dependent protease described in detail (Swamy and Goldberg, 1981), with homologs later identified in other prokaryotes and eukaryotes. This enzyme is active as a homooligomer consisting of four to eight copies

(Goldberg et al., 1994) of a single 784 amino acid polypeptide chain (Amerik et al., 1990; Amerik et al., 1988). Three functional domains with different activities can be identified within each subunit of Lon. The N-terminal domain (N domain) is putatively involved in the recognition and binding of target proteins (Lupas and Martin, 2002). The central part of the chain (A domain) is the ATPase discussed above, while the proteolytic domain (P domain) is located at its C terminus (Amerik et al., 1991; Amerik et al., 1990). The crucial residues of the latter domain that are directly involved in the proteolytic activity of Lon were identified by site-directed mutagenesis as Ser679 and Lys722 (Amerik et al., 1991; Rotanova et al., 2003). Based on the pattern of sequence conservation, the existence of two separate subfamilies of Lon, A and B, was postulated (Rotanova et al., 2003).

The members of the Lon A subfamily, identified in bacteria and eukaryotes, have highly homologous α/β and P domains, while their α domains display significant differences in sizes and lower sequence homology. The N domains are the most variable in both their size and sequence (Rotanova, 1999). The catalytic residues Ser679 and Lys722 are located in the consensus fragments PKDGPS*AG and (K/R)(E/D)K*X Φ (A/S) (Φ denotes a hydrophobic residue), respectively (Rotanova et al., 2003). Lon A protease from *E. coli* (Goldberg et al., 1994) is a “classical” representative of the Lon family (EC 3.4.21.53; MEROPS, clan SF, ID: S16.001; www.merops.ac.uk).

The enzymatic properties of Lon protease and its mode of activity have been the subject of extensive studies (Goldberg et al., 1994; Gottesman et al., 1997; Melnikov et al., 2000; Melnikov et al., 2001; Rotanova, 1999). However, despite considerable efforts, the three-dimensional structure of Lon has not yet been reported, while medium-to-high resolution crystal structures are available for a number of heterooligomeric ATP-dependent proteases. Structural information is available for the proteolytic subunits of ClpP (Wang et al., 1997) and HslV (Bochtler et al., 1997), the ATPase domains of ClpA (Guo et al., 2002) and HslU (Bochtler et al., 2000), as well as for the full-length HslU–HslV complex (Bochtler et al., 2000; Sousa et al., 2002; Wang et al., 2001). A crystal structure of the ATPase domain of the homooligomeric FtsH has also been recently reported (Krzywda et al., 2002; Niwa et al., 2002).

Since our efforts to crystallize the full-length *E. coli* Lon A protease have not yet been successful, we have obtained individual domains by using either protease digestion of the full-length protein, or by expression of those recombinant domains that are stable and soluble. A number of such fragments have been obtained in sufficient quantity for crystallization. We report here the high-resolution crystal structure of the α domain of *E. coli* Lon A (called Lon further on) and its comparison

to corresponding domains of other ATP-dependent proteases (ClpAP, HslVU, and FtsH), as well as other AAA⁺ proteins (NSF (Lenzen et al., 1998; Yu et al., 1998), p97 (Zhang et al., 2000), the clamp loader complex (Guenther et al., 1997; Jeruzalmi et al., 2001a; Jeruzalmi et al., 2001b; Oyama et al., 2001), ClpB (Li and Sha, 2002), Cdc6p (Liu et al., 2000), and the branch migration helicase RuvB (Putnam et al., 2001; Yamada et al., 2001)). The variability of these α domain structures may be one factor in the remarkable diversity of AAA⁺ protein function (Hattendorf and Lindquist, 2002).

2. Materials and methods

2.1. Purification of Lon-S679A and its α domain

A proteolytically inactive mutant form of Lon, Lon-S679A (Amerik et al., 1991), was used to generate chymotryptic fragments. Full-length Lon-S679A was expressed in the *lon*-deficient *E. coli* strain, BL21 (Novagen, Madison, WI), using the plasmid construct pBR-lonS679A (Amerik et al., 1991), and was purified as described previously (Rotanova et al., 1994). Limited proteolysis was performed at 37 °C in 50 mM Tris–HCl buffer, pH 8.0. The proteolysis was monitored by SDS-PAGE on 4–12% NuPage gels (Invitrogen, Carlsbad, CA). The final concentrations of Lon-S679A and α -chymotrypsin (Sigma, St. Louis, MO) in a 100 ml reaction volume were 0.5 mg/ml and 20 μ g/ml, respectively. After 3 h of incubation, the reaction was stopped by adding β -mercaptoethanol to 1% (v/v). The mixture was loaded with a flow rate of 1 ml/min onto a 5 ml HiTrap Q–Sepharose column (Amersham Biosciences, Piscataway, NJ) equilibrated with 50 mM Tris–HCl buffer, pH 8.0. The target protein was collected in the flow-through, concentrated on a Centrprep 10 concentrator (Millipore, Bedford, MA), and fractionated by size exclusion chromatography on a 26/60 Superdex 75 pg column (Amersham Biosciences, Piscataway, NJ) equilibrated in 20 mM Tris–HCl buffer, pH 7.5, 0.15 M NaCl. The purity and homogeneity of the desired fragment (eluted as a 10 kDa protein) was verified by N-terminal sequencing and electrospray ionization mass-spectroscopy (Agilent 1100). These procedures established that the fragment consisted of residues 491–584.

2.2. Protein crystallization

The purified α domain of Lon protease was concentrated to 10 mg/ml. Screening of crystallization conditions (Jancarik and Kim, 1991) was carried out by the hanging drop, vapor-diffusion method (McPherson, 1982), using the Hampton (Hampton Research, Laguna Niguel, CA) and Wizard (Emerald Biostructures, Bain-

bridge Island, WA) screening kits. Diffraction-quality crystals were grown in 30% PEG 3000, 0.1 M CHES, pH 9.5, as well as in 20% PEG 8000, 0.1 M CHES, pH 9.5. The largest crystals grew in 14 days at room temperature to the size of 0.4 mm \times 0.1 mm \times 0.05 mm. Before flash freezing, the crystals were transferred into a cryoprotectant solution consisting of 80% mother liquor and 20% ethylene glycol. For heavy atom derivatization, crystals were soaked for 1 week in their mother liquor to which 10 mM uranyl chloride was added.

2.3. Crystallographic procedures

X-ray data were collected on a Mar345 detector mounted on a Rigaku RU-200 rotating anode X-ray generator operated at 50 kV, 100 mA. The Cu K α radiation was focused by an MSC/Osmic mirror system. Native data were collected to 1.9 Å resolution and were processed and scaled with HKL2000 (Otwinowski, 1992) (Table 1). The complex with uranyl chloride provided a derivative useful for phasing with data

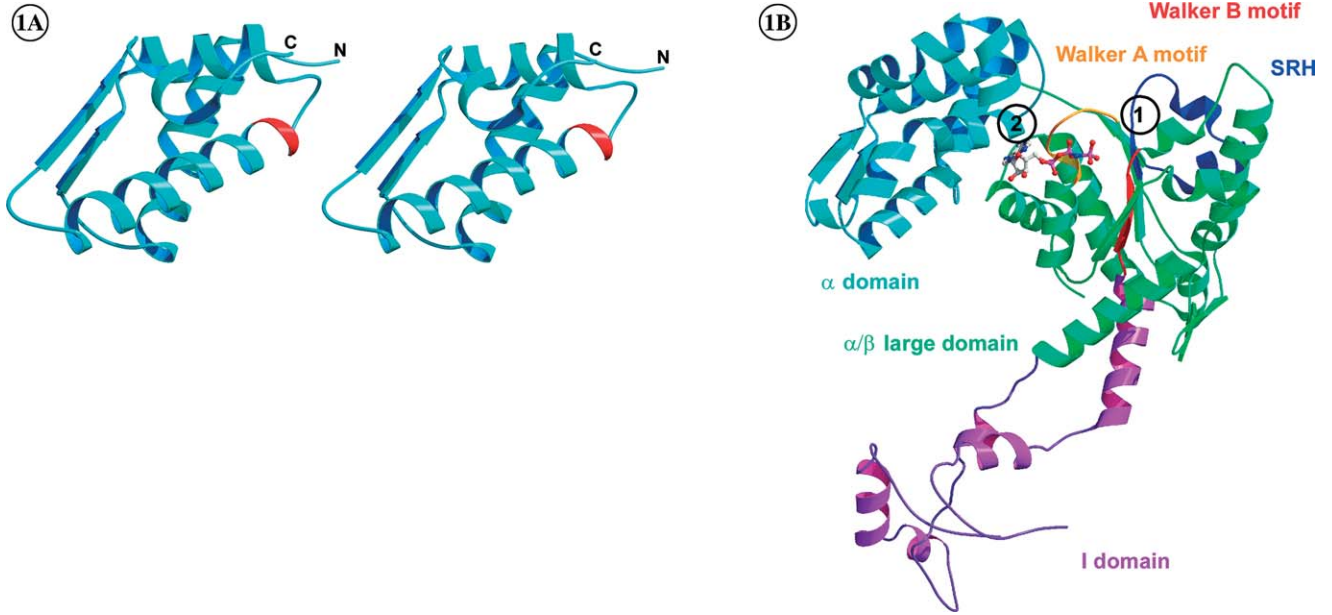
Table 1
Statistics of data collection and structure refinement

	Native	UOCl ₄
<i>Data collection</i>		
Space group		P4 ₃ 2 ₁ 2
Molec/a.u.		1
Unit cell parameters (Å)	$a = b = 46.43,$ $c = 85.2$	$a = b = 46.4,$ $c = 85.1$
Resolution (Å)	20–1.9	20–2.45
Total reflections	46 524	47 748
Unique reflections	7808 (Friedel merged)	6462 (Friedel unmerged)
Completeness (%)	98.8	98.5
Avg. I/σ	12.6	11.0
R_{merge} (%) ^a	8.1	9.0
<i>Phasing statistics (20–2.45 Å)</i>		
Phasing power, acentric	—	1.28
$R_{\text{cullis, iso}}$ acentric	—	0.73
Phasing power, centric	—	1.7
$R_{\text{cullis, iso}}$ centric	—	0.68
Phasing power, anomalous	—	1.06
$R_{\text{cullis, ano}}$	—	0.84
<i>Refinement statistics</i>		
R (%) ^b	17.9	—
R_{free} (%) ^c	28.4	—
r.m.s.d. bond lengths (Å)	0.011	—
Angle distances (Å)	0.038	—
Temp. factor (protein, Å ³)	36.2	—
(Solvent, Å ³)	49.2	—
Number of protein atoms	762	—
Number of solvent molecules	90	—

^a $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections after rejections.

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

^c R_{free} defined in Brünger (1992).



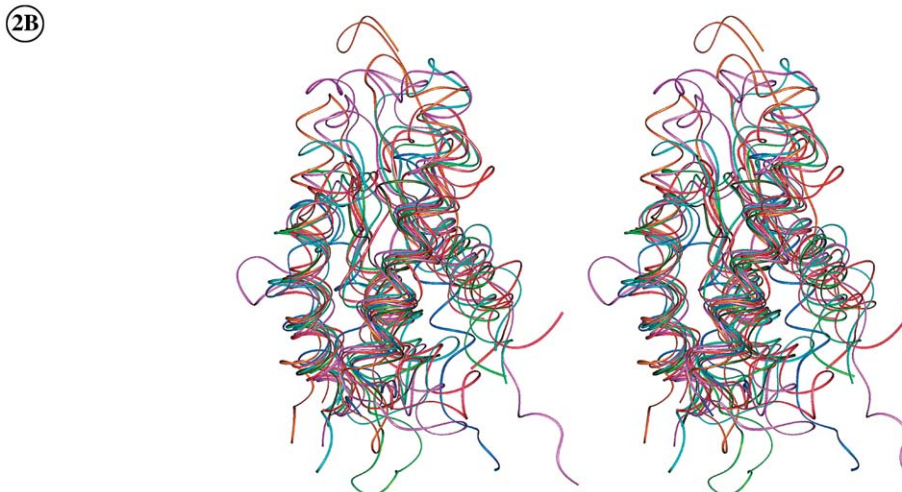
2A

LON	491	SGY-TEDEKLNIAK-R--HLLPKQIERNA-LKKGELTV-DDSAIIGIIRYY-----TREA	539
HSLU-Ec	333	QAL-TTSDFERILTEPNASITVQYKALMA-TEGVNIEF-TDSGIKRIEAAWQVNES-----TENI--	390
HSLU-Hi	334	TAL-SAADFERILTEPHASLTEQYKALMA-TEGVNIAF-TTDAVKKIAEAAFRVNEK-----TENI--	391
CLPA-D1	350	--P-SIEETVQIIN-G---LKPQYEAHHD-----VRY-TAKAVRAAVELAVKYIN-----DPRH	393
CLPA-D2	652	HL-STDVIHQVVD-K---FIVELQVQLD-QKGVSLV-SQEARNWLAEKGY-----DRAM-	699
NSF-D2	677	NIATGEQLLEALE-L-----LGNFKD---KERTTIAQVQ-----KGKKVWI--	713
C-LOADER	169	--L-RDEDIAKRLR-Y---IAENEG-----LEL-TEEGLQAILYI-----AEG-	203
CDC6P	192	--Y-TKDQIFDILL-D---RAKAGL-----AEGSY-SEDILQMIAD-----ITGAQTPLDNRG	237
P97	372	--P-DATGRLEILQ-I---HT-----KNMKLADDV---DLEQVANE-----THGH	406
FTSH-Tt	326	--P-DVKGREQILR-I---HAR-----GKPLAEDVDLALLAKRT-----PGF--	358
FTSH-Ec	326	--P-DVRGREQILK-V---HMR-----RVPLAPDIDAAIARG-----TPGF--	357
RUVB-Tt	166	EYY-TPEELAQGV-M---DARL-----LGVRI-TEEAALIGRR-----SR-G	202

helix 1
helix 2

LON	540	GVRLEREISKLCKRAVKQLLLDK-----SLKHIEI----NGDNLHDYLG---VQRF-----	584
HSLU-Ec	391	GARRLHTVLERLMEEISYDADLS-----GQNITI---DADYVSKHLD---ALVAEDLSRFIL---	443
HSLU-Hi	392	GARRLHTVMERLMDKISFSASDMN-----GQTVNI---DAAVYADALG---EVVENEDLSRFIL---	444
CLPA-D1	394	LDPKATDVIDEAGARARLMP-----VSKRKKTV---NVADIESVARIARI-----	433
CLPA-D2	700	GARPMARVIQDNLKKPLANELLF-----GSLVDGGQVVALDKE-----KNELTYGFSQAQKHAE	755
NSF-D2	714	GIKKLLMLIEMSLQM-----DPEY-----RVRKFLALLREE---GASPLDFK---	751
C-LOADER	204	DMRRAINILQAAAAL-----DKKI---TDENVMVAS-----	232
CDC6P	238	DARLAIDILYRSAYAAQONG-----RKHI---APEDVRKSSKEV--LF-----	276
P97	407	VGADLAALCSEAALQAIRKKM---DLIDLEDETIDAEMNSLAV---TMDDFRWALSQ-----	458
FTSH-Tt	359	VGADLENLLNEAALLAAREG-----RRKI---TMKDLLEAAS-----	392
FTSH-Ec	358	SGADLANLVNEAALFAARGN-----KRVV---SMVEFEKAKDKMMGLG-----	398
RUVB-Tt	203	TMRVAKRLFRRVRDF-----AQVAGEEVI---TRERALEAALGDELG-----	244

helix 3
helix 4



collected under the same conditions as those used as for the native protein. SIRAS phasing was performed using the program SHARP (Global Phasing Ltd., Cambridge) and resulted in a map into which 15 residues could be built automatically with the program ARP/wARP (Perrakis et al., 1999). Several cycles of the program RESOLVE (Terwilliger, 2001) yielded various other fragments of the structure, which were combined into a 54-residue partial model. The rest of the model was built manually into the initial map using the program O (Jones and Kjeldgaard, 1997). 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 parameters (Engh and Huber, 1991). Final rounds of refinement were carried out with SHELXL (Sheldrick, 1998), leading to a model with an R of 17.9% and R_{free} (Brünger, 1992) of 28.2% for all data between 10 and 1.9 Å resolution. A considerable difference between these indicators is not uncommon for structures refined with SHELXL at similar resolution, and the final residual electron density maps were featureless. The Ramachandran plot for the final structure, obtained with the program PROCHECK (Laskowski et al., 1993), showed 91.7% of the residues in the core region and 8.3% in the additional allowed region. Other quality parameters also indicated a well-refined structure (Table 1). The coordinates and the structure factors have been submitted to the Protein Data Bank (PDB) with the accession code 1qzm for immediate release.

3. Results and discussion

3.1. Description of the structure

The small α domain of the *E. coli* Lon AAA⁺ module was previously defined as residues 495–607, including 13 C-terminal residues past the region of Clp homology, reflecting uncertainty in the domain boundary in the absence of structural data (Smith et al., 1999). However, trypsin cleavage of full-length Lon produced a stable fragment containing residues 485–588 (data not shown), and limited digestion with α -chymotrypsin yielded a

stable fragment consisting of residues 491–584, which, as shown below, had a well-defined three-dimensional structure and was found to constitute a complete α domain.

The structure of the α domain of *E. coli* Lon is shown in Fig. 1A. Despite variation in the number of residues within the contributing secondary structural elements, it follows a conserved α domain topology (helix–strand–helix–helix–strand–helix). Helix 1 contains residues 494–514 with a slight bend around His505. A loop leads to β -strand 1, which is followed by helix 2 consisting of residues 524–535. A short turn links helix 2 to a long helix 3, containing residues 540–563. The well-conserved sensor-2 Arg542 residue is located at the beginning of helix 3. The following second β -strand loops back to form a parallel β -sheet with the first strand. The C-terminal helix 4 consisting of residues 571–580 is unwound at the end. However, with the terminal residues having very clearly defined electron density, such unwinding does not appear to be an artifact of the C-terminal proteolytic cleavage. There are no extensive crystal contacts between helix 4 and the neighboring molecules.

The architecture of the Lon molecule and the predicted location of the α domain can be modeled on the basis of the known structure of HslU. It is expected that the α domain of Lon should form an interface with the larger α/β domain and that the nucleotide should be cradled between them. The α domain is also expected to make contacts with the α/β domain of the adjacent subunit and contribute the interface for oligomerization, as is seen in HslU (Fig. 1B), ClpA (Guo et al., 2002), and other AAA⁺ proteins (Lenzen et al., 1998; Yu et al., 1998; Zhang et al., 2000).

3.2. Structural homology

The crystal structure of the α domain of Lon protease was compared to the structures of the corresponding domains for 11 members of the AAA⁺ family, listed in the legend of Fig. 2. Although these enzymes are topologically very similar, their C α atoms do not superimpose exactly and, in most cases, automatic alignment (using the program ALIGN (Cohen,

Fig. 1. Crystal structure of the α domain of Lon and its comparison with the structure of the HslU subunit of protease HslVU. (A) Stereo view of the α domain of *E. coli* Lon. The sensor-2 residue (Arg) is marked in red. (B) Architecture of a typical ATPase functional domain (*E. coli* HslU, PDB entry 1e94). The large α/β domain with the Rossmann fold is shown in green, the α domain in cyan, and the additional I domain in magenta. The locations of the Walker A and B motifs and the second region of homology (SRH) are color-coded, and the positions of the sensor-1 and sensor-2 residues are shown as circled numbers. The nucleotide (AMP-PNP) is cradled between the two domains. The figure was generated with programs Bobscript (Esnouf, 1999) and Raster3D (Meritt and Murphy, 1994).

Fig. 2. Comparison of the α domains. (A) Structure-based sequence alignment of the α domains of *E. coli* Lon, *E. coli* HslU, *H. influenzae* HslU, *E. coli* ClpA domains D1 and D2, *C. griseus* NSF domain D2, *P. furiosus* clamp loader, *P. aerophilum* Cdc6p, *M. musculus* p97, *T. thermophilus* FtsH, *E. coli* FtsH, and *T. thermophilus* RuvB. Residues involved in nucleotide binding are highlighted in yellow and the sensor-2 residue in blue. (B) Stereo view of the superimposed C α traces of the proteins compared in (A). This panel generated with InsightII (Accelrys, Inc.).

1997)) was not successful. A preliminary alignment was done manually and was followed by an optimization procedure using the program PROFIT (Martin, 1996) for the homologous secondary structural elements (Fig. 2A). The range of r.m.s. deviations for those regions was 1.5–2.0 Å. Despite the different lengths of the corresponding helices in the structures, the mode of superposition was unambiguous because of the presence of guide points, such as the sensor-2 arginine, and the residues forming the binding site for the nucleotide. Since those functionally important residues are located within helices 1 and 3, it is not surprising that these two helices are superimposed much more accurately in all 12 structures than helices 2 or 4. Helix 4 is located at the C-terminus of the α domain and is unwound in the majority of the structures. In NSF-D2, helices 2 and 4 occupy distinctly different topological positions when helices 1 and 3 are superimposed, with helix 1 being significantly shorter than in the other structures. The α domain of ClpA-D2 lacks helix 4, which is replaced by a β -loop at the C-terminus which contributes the third antiparallel strand to the β -sheet formed by strands 1 and 2. In HslU, helix 1 has a kink in the middle with the insertion of four residues. Superposition of various α domains reveals that helix 1 is slightly bent or ends at this position in other AAA⁺ proteins, but, aside from HslU, only Lon protease has an extra residue (His505) in the middle of helix 1.

A sequence alignment for the α domains of the AAA⁺ proteins (Fig. 2A) was created on the basis of the superposition of their structures. The sequence similarity is higher for the α domains from the LonA subfamily (not shown). Conservation is highest among the residues that form the hydrophobic cores of the domains. Only very few hydrophilic residues are conserved and there is no clear pattern for the distribution of negatively and positively charged residues, with the exception of the position of sensor-2 residue.

3.3. Dual role of the α domain

Data from different laboratories indicate that the α domains of AAA⁺ proteins contribute to several aspects of their function. They appear to be essential for oligomerization, by mediating interactions between adjacent monomers in the oligomer (Ogura and Wilkinson, 2001). For instance, ClpB mutants lacking the α domain fail to oligomerize (Barnett et al., 2000). In addition, isolated α domains of different AAA⁺ proteins could interact with appropriate substrates (Smith et al., 1999), leading to the proposal that this region acts as a “sensor- and substrate-discrimination” domain.

Structural data also support two important functional roles for α domains. Residues within the α domain, assisting the residues from the α/β domain, participate in forming the binding site for the nucleotide. In addition, interactions between these two structural domains maintain the interface between the monomers upon oligomerization. The nucleotide binding site of the α domain is formed by residues from helices 1 and 3. Four residues from helix 1 interacting with the nucleotide have well conserved topology. They are generally hydrophobic, although they show some variability (Table 2). In most structures, helix 1 residues interact predominantly with the adenine moiety except for residues Ile334 and Ile337 in FtSH (PDB code 1IY1) which interact with the ribose as well. In the structures of NSF and RuvA/B the nucleotide does not interact with the first helix, most likely due to the specific conformation of the N-terminal loop of the α/β domain in NSF and the presence of a Gly residue at position 176 on helix 1 in the α domain of RuvB. A comparison of the interactions of the α domain residues with the nucleotides in several superimposed structures allows us to map the corresponding residues in the α domain of Lon protease with high probability, even in the absence of a bound nucleotide. We propose that Tyr493, Ile501, Arg504, or

Table 2

Interactions between the protein and the bound nucleotides in the members of the AAA⁺ family. Hydrophilic residues within the distance of 3.5 Å from the nucleotide and hydrophobic residues within 4.5 Å distance are bolded

AAA ⁺ protein	Bound nucl.	Residues						
		Helix 1			Helix 3			
Lon Ec	—	Y493	I501	R504–H505	L506	V541	R542	E545
HslU Hi	ATP	L336	I344	P348	I352	A393	R394	H397
HslU Ec	DAT	L335	I343	P347	I351	A392	R393	H396
ClpA-D1	ADP	P349	I357	G360	L361	P395	D396	I399
ClpA-D2	ADP	L653	V661	K664	F665	A701	R702	A705
NSF-D2	ANP	I678	A687	L690	L691	I715	K716	L719
p97	ADP	P371	I380	I383	H384	G408	A409	A412
Cdc6p	ADP	Y192	I200	D203	R204	A239	R240	I243
Clamp loader	ADP	L169	R177	Y180	I181	M205	R206	I209
RuvB Tt	ANP	Y168	G176	R179	D180	M204	R205	K208
FtSH Tt	ADP	P326	I334	I337	H338	G362	A363	E366

Abbreviations: DAT, 2'-deoxyadenosine-5'-diphosphate; ANP, phosphoaminophosphonic acid-adenylate ester (also called 5'-adenylyl-imidodiphosphate).

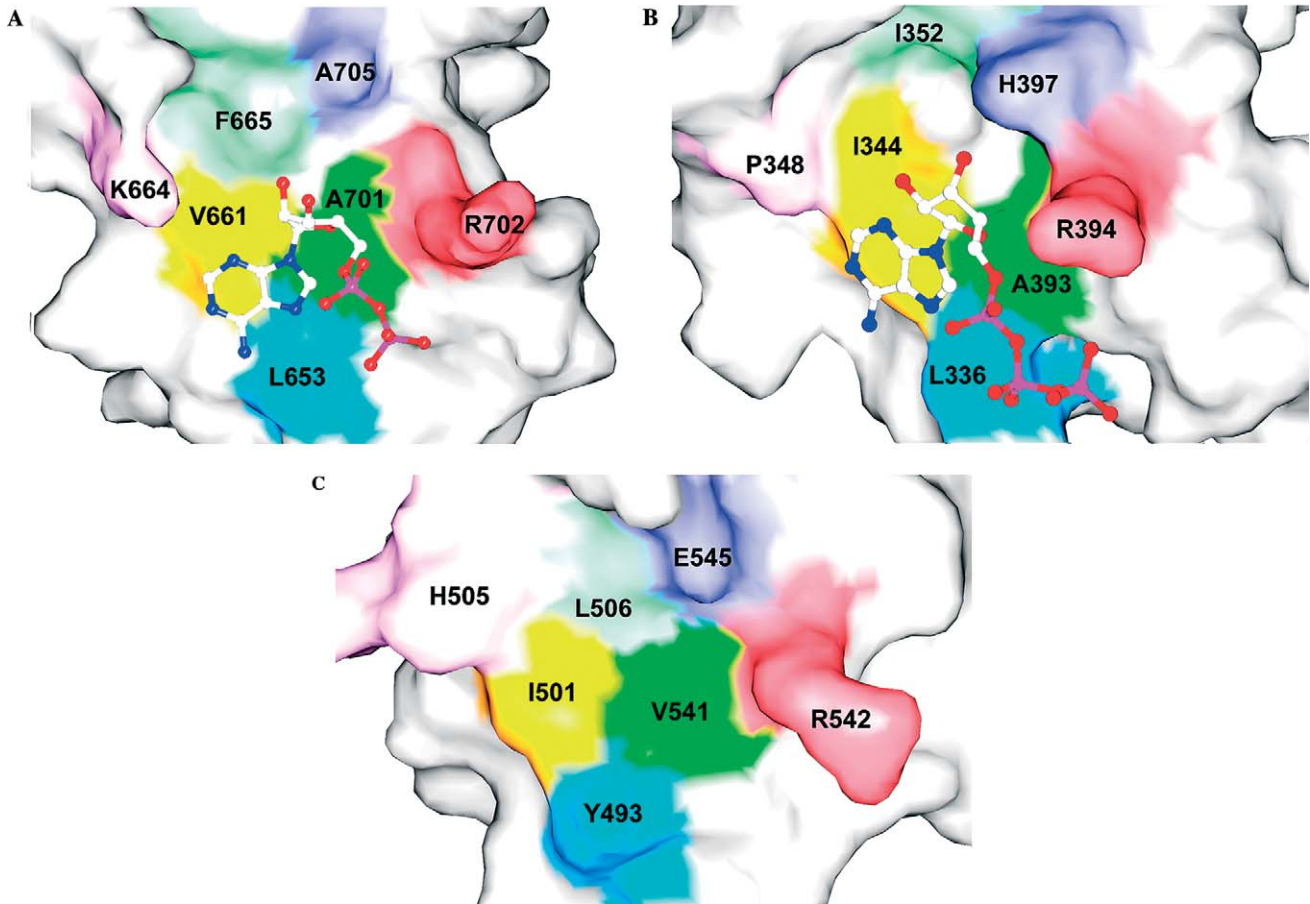


Fig. 3. Residues of the α domains that form nucleotide-binding sites in ATP-dependent proteases. (A) *E. coli* ClpA D2-small domain (PDB entry 1ksf). (B) *H. influenzae* HslU (PDB entry 1kyi). (C) A homologous area of the *E. coli* Lon, predicted to be involved in nucleotide binding. This figure is based on data summarized in Table 2.

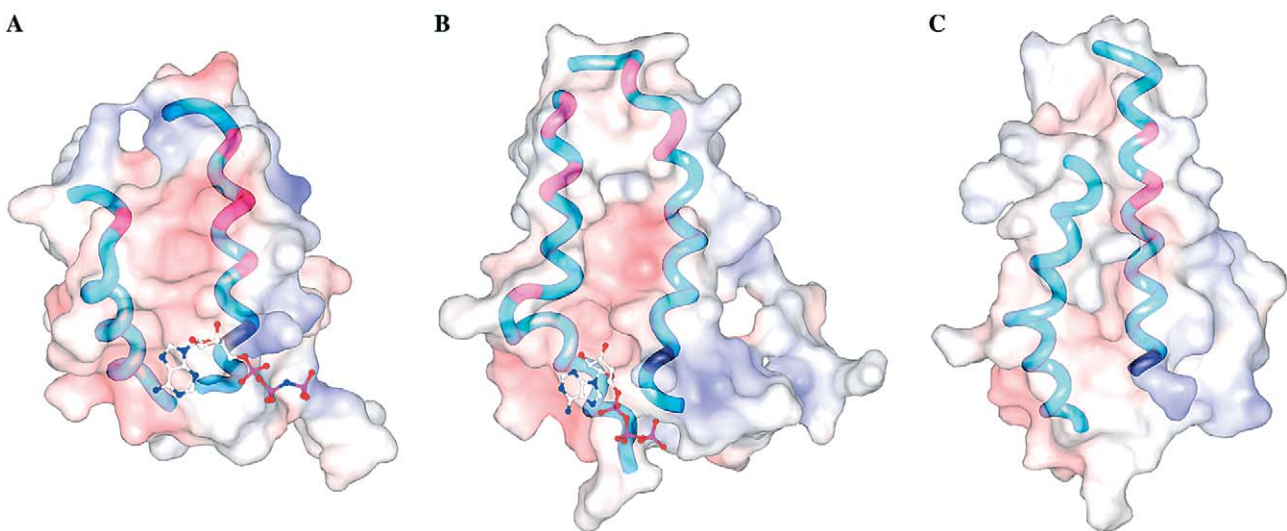


Fig. 4. Oligomerization interface in selected α domains, showing surface charges. (A) NSF (PDB entry 1d2n), (B) *H. influenzae* HslU (PDB entry 1kyi), and (C) *E. coli* Lon. Helices 1 and 3 are shown as blue ribbons with their parts involved in forming the interface with the α/β domain highlighted in pink. Putative sensor-2 residues are highlighted in dark blue. The bound nucleotide is shown in ball-and-stick representation. Figure generated with SPOCK (Christopher, 1998).

His505, and Leu506 (all located on helix 1) are likely to be involved in the binding of nucleotide.

Three residues from helix 3 occupy structurally equivalent positions and interact with the ribose and the phosphate in all structures, with the exception of FtsH (Table 2), although the set of interacting pairs in each structure depends on the type of the bound nucleotide and its conformation. Two out of three residues are well conserved in the majority of the structures, i.e., sensor-2 Arg (Lys in NSF) and the preceding residue, which is always hydrophobic and interacts with the nucleotide. In the structures of the α domains that do not contain a sensor-2 Arg, the latter residue is either Gly or Pro (Table 2). The third residue varies in its nature from hydrophobic to negatively or positively charged. We can postulate that Val541, Arg542, and Glu545 in helix 3 of the α domain of Lon should interact with the ribose and the phosphate of the nucleotide. In the compared structures, residues from the vicinity of the sensor-2 Arg project into the active site. Contacts between sensor-2 and the nucleotide (Fig. 3) probably regulate the activity of this domain via conformational changes associated with nucleotide binding or hydrolysis (Hattendorf and Lindquist, 2002), although the nature of this regulation is still not elucidated.

In the crystals of two AAA⁺ proteins, HslU and the small subunit of the clamp loader complex from *Pyrococcus furiosus*, the functional ATPase domains are present either as a hexamer or as a dimer of trimers, respectively, in the asymmetric unit. In addition, p97-D1 and NSF-D2 are hexameric in solution and their oligomerization state was preserved in the crystals. In these cases, it is most likely that crystal symmetry should not have imposed or changed the mode of oligomerization, allowing the characterization and comparison of the residues that maintain the interactions within the interface between the monomers in each protein. The interface is formed predominantly between the α domain of one ATPase monomer and the α/β domain of the adjacent one. The residues of the α domains contributing to the interface belong to the same area of the domain surface (Fig. 4). A common characteristic of these electrostatic and hydrophobic interactions is that they mainly involve residues located on the upper part of the third helix of the α domain. In HslU, extensive interactions are also found between the upper part of the third helix of the α domain and the N-terminal loop of the larger domain of monomer 2. In the clamp loader subunit interface, a few very specific interactions are found involving residues on the C-terminus of the α domain. The interacting residues are not particularly well conserved even within the similar parts of the interface. This indicates that the interactions within the interface are very specific for each protein, making it difficult to assign all individual residues that might be playing a similar role in the α domain of Lon, although the general area of the

interface surface of the latter can be suggested (Fig. 4). Some predictions, nevertheless, can be made. In particular, in the α domain of HslU, several positions which interact with residues from the adjacent subunit are consistently occupied by longer side-chains (Leu, Tyr, Arg, Lys, or Glu). In ClpX, mutations of residues which are expected to occupy these positions by homology modeling (Leu381, Asp382, or Tyr385) led to severe defects in activity, possibly by hindering nucleotide- and protein substrate-dependent conformational changes (Joshi et al., 2003). In Lon, the equivalent residues are Arg553, Lys554, and Lys557. These basic residues could interact with acidic residues in the first two helices of the α/β domain, (e. g., Glu250, Glu252, Asp259, Glu266, or Glu269), which form the interface with the α domain in other AAA⁺ proteins.

The similarity of the isolated α domain of Lon to the corresponding domains of the AAA⁺ proteins, with or without bound nucleotides, emphasizes the relevance of the structure presented here. Conservation of the structural features of this domain observed in different environments makes it likely that its structure is not significantly modified in the context of the intact subunit of Lon. The prediction of the residues in the α domain of Lon protease which are involved in nucleotide binding and in oligomerization will be tested by mutagenesis. However, only the structure of the holoenzyme and of a nucleotide complex of the AAA⁺ module of Lon can verify these points in an unambiguous manner.

Acknowledgments

We are grateful to Dr. Zbigniew Dauter for his help in obtaining the initial phases used for structure solution. This work was supported in part by a grant from the Russian Foundation for Basic Research (Project No. 02-04-48481) to TVR and by the US Civilian Research and Development grant RB1-2505-MO-03 to TVR and AW.

References

- 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.
- Amerik, A.Y., Chistyakova, L.G., Ostroumova, N.I., Gurevich, A.I., Antonov, V.K., 1988. Cloning, expression and structure of the functionally active shortened *lon* gene in *Escherichia coli*. *Bioorg. Khim.* 14, 408–411.
- Amerik, A.Y., Antonov, V.K., Gorbalenya, A.E., Kotova, S.A., Rotanova, T.V., Shimbarevich, E.V., 1991. Site-directed mutagenesis of La protease. A catalytically active serine residue. *FEBS Lett.* 287, 211–214.

- Barnett, M.E., Zolkiewska, A., Zolkiewski, M., 2000. Structure and activity of ClpB from *Escherichia coli*. Role of the amino- and carboxyl-terminal domains. *J. Biol. Chem.* 275, 37565–37571.
- 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.
- Bochtler, M., Hartmann, C., Song, H.K., Bourenkov, G.P., Bartunik, H.D., Huber, R., 2000. The structures of HslU and the ATP-dependent protease HslU–HslV. *Nature* 403, 800–805.
- 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., Adams, P.D., Clore, G.M., et al., 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54, 905–921.
- Brünger, A.T., Krukowski, A., Erickson, J.W., 1990. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. A* 46, 585–593.
- Christopher, J.A., 1998. SPOCK: The Structural Properties Observation and Calculation Kit. The Center for Macromolecular Design, Texas A&M University, College Station, TX.
- Cohen, G.E., 1997. ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.* 30, 1160–1161.
- Ebel, W., Skinner, M.M., Dierksen, K.P., Scott, J.M., Trempe, J.E., 1999. A conserved domain in *Escherichia coli* Lon protease is involved in substrate discriminator activity. *J. Bacteriol.* 181, 2236–2243.
- Engh, R., Huber, R., 1991. Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr. A* 47, 392–400.
- Esnouf, R.M., 1999. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* 55, 938–940.
- Goldberg, A.L., 1992. The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. *Eur. J. Biochem.* 203, 9–23.
- 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.
- Gottesman, S., 1996. Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* 30, 465–506.
- Gottesman, S., Maurizi, M.R., 1992. Regulation by proteolysis: energy-dependent proteases and their targets. *Microbiol. Rev.* 56, 592–621.
- Gottesman, S., Wickner, S., Maurizi, M.R., 1997. Protein quality control: triage by chaperones and proteases. *Genes Dev.* 11, 815–823.
- Guenther, B., Onrust, R., Sali, A., O'Donnell, M., Kuriyan, J., 1997. Crystal structure of the δ' subunit of the clamp-loader complex of *E. coli* DNA polymerase III. *Cell* 91, 335–345.
- Guo, F., Maurizi, M.R., Esser, L., Xia, D., 2002. Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease. *J. Biol. Chem.* 277, 46743–46752.
- Hattendorf, D.A., Lindquist, S.L., 2002. Analysis of the AAA sensor-2 motif in the C-terminal ATPase domain of Hsp104 with a site-specific fluorescent probe of nucleotide binding. *Proc. Natl. Acad. Sci. USA* 99, 2732–2737.
- Hershko, A., Ciechanover, A., 1992. The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* 61, 761–807.
- Jancarik, J., Kim, S.H., 1991. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* 21, 916–924.
- Jeruzalmi, D., O'Donnell, M., Kuriyan, J., 2001a. Crystal structure of the processivity clamp loader gamma (γ) complex of *E. coli* DNA polymerase III. *Cell* 106, 429–441.
- Jeruzalmi, D., Yurieva, O., Zhao, Y., et al., 2001b. Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. *Cell* 106, 417–428.
- Jones, T.A., Kjeldgaard, M., 1997. Electron-density map interpretation. *Methods Enzymol.* 277, 173–208.
- Joshi, S.A., Baker, T.A., Sauer, R.T., 2003. C-terminal domain mutations in ClpX uncouple substrate binding from an engagement step required for unfolding. *Mol. Microbiol.* 48, 67–76.
- Karata, K., Inagawa, T., Wilkinson, A.J., Tatsuta, T., Ogura, T., 1999. Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. Site-directed mutagenesis of the ATP-dependent protease FtsH. *J. Biol. Chem.* 274, 26225–26232.
- Krzywdka, S., Brzozowski, A.M., Verma, C., Karata, K., Ogura, T., Wilkinson, A.J., 2002. The crystal structure of the AAA domain of the ATP-dependent protease FtsH of *Escherichia coli* at 1.5 Å resolution. *Structure* 10, 1073–1083.
- 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. Crystallogr.* 26, 283–291.
- Lenzen, C.U., Steinmann, D., Whiteheart, S.W., Weis, W.I., 1998. Crystal structure of the hexamerization domain of *N*-ethylmaleimide-sensitive fusion protein. *Cell* 94, 525–536.
- Leonhard, K., Stiegler, A., Neupert, W., Langer, T., 1999. Chaperone-like activity of the AAA domain of the yeast Yme1 AAA protease. *Nature* 398, 348–351.
- Li, J., Sha, B., 2002. Crystal structure of *E. coli* Hsp100 ClpB nucleotide-binding domain 1 (NBD1) and mechanistic studies on ClpB ATPase activity. *J. Mol. Biol.* 318, 1127–1137.
- Liu, J., Smith, C.L., DeRyckere, D., DeAngelis, K., Martin, G.S., Berger, J.M., 2000. Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol. Cell* 6, 637–648.
- Lupas, A.N., Martin, J., 2002. AAA proteins. *Curr. Opin. Struct. Biol.* 12, 746–753.
- Martin, A.C.R., 1996. ProFit. Available from <http://www.biochem.ucl.ac.uk/~martin/#profit>.
- Maurizi, M.R., Li, C.C., 2001. AAA proteins: in search of a common molecular basis. *International Meeting on Cellular Functions of AAA Proteins. EMBO Rep.* 2, 980–985.
- McPherson, A., 1982. *Preparation and Analysis of Protein Crystals*. Wiley, New York.
- Melnikov, E.E., Tsiurulnikov, K.B., 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.
- Melnikov, E.E., Tsiurulnikov, 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. *Bioorg. Khim.* 27, 120–129.
- Meritt, E.A., Murphy, M.E.P., 1994. *Raster3D* Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* 50, 869–873.
- 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.
- Niwa, H., Tsuchiya, D., Makyio, H., Yoshida, M., Morikawa, K., 2002. Hexameric ring structure of the ATPase domain of the membrane-integrated metalloprotease FtsH from *Thermus thermophilus* HB8. *Structure* 10, 1415–1423.
- Ogura, T., Wilkinson, A.J., 2001. AAA⁺ superfamily ATPases: common structure–diverse function. *Genes Cells* 6, 575–597.
- Ortega, J., Singh, S.K., Ishikawa, T., Maurizi, M.R., Steven, A.C., 2000. Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Mol. Cell* 6, 1515–1521.
- Otwinowski, Z., 1992. *An Oscillation Data Processing Suite for Macromolecular Crystallography*. Yale University, New Haven.

- Oyama, T., Ishino, Y., Cann, I.K., Ishino, S., Morikawa, K., 2001. Atomic structure of the clamp loader small subunit from *Pyrococcus furiosus*. *Mol. Cell* 8, 455–463.
- Patel, S., Latterich, M., 1998. The AAA team: related ATPases with diverse functions. *Trends Cell Biol.* 8, 65–71.
- Perrakis, A., Morris, R., Lamzin, V.S., 1999. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6, 458–463.
- Putnam, C.D., Clancy, S.B., Tsuruta, H., Gonzalez, S., Wetmur, J.G., Tainer, J.A., 2001. Structure and mechanism of the RuvB Holliday junction branch migration motor. *J. Mol. Biol.* 311, 297–310.
- Rotanova, T.V., 1999. Structural and functional characteristics of ATP-dependent Lon-proteinase from *Escherichia coli*. *Bioorg. Khim.* 25, 883–891.
- Rotanova, T.V., Melnikov, E.E., Tsurulnikov, K.B., 2003. A catalytic Ser–Lys dyad in the active site of the ATP-dependent Lon protease from *Escherichia coli*. *Bioorg. Khim.* 29, 97–99.
- Rotanova, T.V., Kotova, S.A., Amerik, A.Yu., Lykov, I.P., Ginodman, L.M., Antonov, V.K., 1994. ATP dependent protease La from *Escherichia coli*. *Bioorg. Khim.* 20, 114–125.
- Schulz, G.E., Schirmer, R.H., 1979. *Principles of Protein Structure*. Springer, New York.
- Sheldrick, G.M., 1998. SHELX: Applications to macromolecules. In: Fortier, S. (Ed.), *Direct methods for solving macromolecular structures*. Kluwer Academic Publishers, Dordrecht, pp. 401–411.
- Smith, C.K., Baker, T.A., Sauer, R.T., 1999. Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. *Proc. Natl. Acad. Sci. USA* 96, 6678–6682.
- Sousa, M.C., Kessler, B.M., Overkleef, H.S., McKay, D.B., 2002. Crystal structure of HslUV complexed with a vinyl sulfone inhibitor: corroboration of a proposed mechanism of allosteric activation of HslV by HslU. *J. Mol. Biol.* 318, 779–785.
- Swamy, K.H., Goldberg, A.L., 1981. *E. coli* contains eight soluble proteolytic activities, one being ATP dependent. *Nature* 292, 652–654.
- Terwilliger, T.C., 2001. Map-likelihood phasing. *Acta Crystallogr. D* 57, 1763–1775.
- Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J., 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945–951.
- 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.
- Wang, J., Song, J.J., Franklin, M.C., et al., 2001. Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure* 9, 177–184.
- Wickner, S., Maurizi, M.R., Gottesman, S., 1999. Posttranslational quality control: folding, refolding, and degrading proteins. *Science* 286, 1888–1893.
- Yamada, K., Kunishima, N., Mayanagi, K., et al., 2001. Crystal structure of the Holliday junction migration motor protein RuvB from *Thermus thermophilus* HB8. *Proc. Natl. Acad. Sci. USA* 98, 1442–1447.
- Yu, R.C., Hanson, P.I., Jahn, R., Brunger, A.T., 1998. Structure of the ATP-dependent oligomerization domain of *N*-ethylmaleimide sensitive factor complexed with ATP. *Nat. Struct. Biol.* 5, 803–811.
- Zhang, X., Shaw, A., Bates, P.A., et al., 2000. Structure of the AAA ATPase p97. *Mol. Cell* 6, 1473–1484.