Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the N-terminal fragment of *Escherichia coli* Lon protease

Mi Li,[a,b] Alla Gustchina,[a] Fatima S. Rasulova,[c] Edward E. Melnikov,[d] Michael R. Maurizi,[c] Tatyana V. Rotanova,[d] Zbigniew Dauter[e] and Alexander Wlodawer[a]*

[a]Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, MD 21702-1201, USA, [b]Basic Research Program, SAIC-Frederick, Frederick, MD 21702, USA, [c]Laboratory of Cell Biology, National Cancer Institute, Bethesda, MD 20892, USA, [d]Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow 117997, Russia, and [e]Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, NCI, Argonne National Laboratory, Argonne, IL 60439, USA

Correspondence e-mail: wlodawer@nih.gov

Received 29 April 2010
Accepted 25 May 2010

PDB Reference: N-terminal fragment of Lon protease, 3ljc.

The structure of a recombinant construct consisting of residues 1–245 of *Escherichia coli* Lon protease, the prototypical member of the A-type Lon family, is reported. This construct encompasses all or most of the N-terminal domain of the enzyme. The structure was solved by SeMet SAD to 2.6 Å resolution utilizing trigonal crystals that contained one molecule in the asymmetric unit. The molecule consists of two compact subdomains and a very long C-terminal α-helix. The structure of the first subdomain (residues 1–117), which consists mostly of β-strands, is similar to that of the shorter fragment previously expressed and crystallized, whereas the second subdomain is almost entirely helical. The fold and spatial relationship of the two subdomains, with the exception of the C-terminal helix, closely resemble the structure of BPP1347, a 203-amino-acid protein of unknown function from *Bordetella parapertussis*, and more distantly several other proteins. It was not possible to refine the structure to satisfactory convergence; however, since almost all of the Se atoms could be located on the basis of their anomalous scattering the correctness of the overall structure is not in question. The structure reported here was also compared with the structures of the putative substrate-binding domains of several proteins, showing topological similarities that should help in defining the binding sites used by Lon substrates.

1. Introduction

Lon proteases were the first well characterized enzymes that contained both protease and ATPase domains. Studies of the enzymatic properties of *Escherichia coli* Lon protease (*Ec*Lon) established the general properties of the whole family of such proteins (Goldberg *et al.*, 1994; Gottesman *et al.*, 1997; Melnikov *et al.*, 2000). The enzymatic function of Lon proteases is to couple ATP hydrolysis to structural disruption and processive degradation of proteins into peptides consisting of 5–12 amino acids. As determined by electron microscopy (Park *et al.*, 2006), *Ec*Lon forms hexamers of identical 784-amino-acid polypeptide chains (Amerik *et al.*, 1990; Goldberg *et al.*, 1994). Although a crystal structure of the full-length enzyme is not yet available, a number of structures of fragments of *Ec*Lon and its orthologs from other organisms have been reported (Botos, Melnikov, Cherry, Khalatova *et al.*, 2004; Botos, Melnikov, Cherry, Tropea *et al.*, 2004; Botos *et al.*, 2005; Garcia-Nafria *et al.*, 2010; Im *et al.*, 2004; Li *et al.*, 2005). Each subunit of *Ec*Lon contains three functional domains (Amerik *et al.*, 1990; Gottesman *et al.*, 1995; Rotanova *et al.*, 2006); however, the number of separately folded structural subdomains is most likely to be larger. *Ec*Lon was classified as a member of the LonA family, all of which have been shown

to consist of tandem N-terminal, ATPase and proteolytic domains (Rotanova *et al.*, 2004). The central region of LonA proteases, named the A domain, is an ATPase belonging to the AAA^+ superfamily (Neuwald *et al.*, 1999). The proteolytically active C-terminal region, or P domain, defines a unique serine protease family and has a serine–lysine catalytic dyad (Botos, Melnikov, Cherry, Tropea *et al.*, 2004; Rotanova *et al.*, 2003, 2004).

The defining feature of the LonA family in bacteria is its N-terminal domain, which is around 300–350 amino acids in length and is differentiated by sequence conservation and presumed functional properties. The N domain is divided into two or more subdomains (Li *et al.*, 2005) and is also predicted to include an extended coiled coil (CC; amino acids ∼175–280 in *Ec*Lon; Ebel *et al.*, 1999; Lee *et al.*, 2004; Melnikov *et al.*, 2008). The CC segment joins the two parts of the N domain comprising residues 1 to ∼200 to a small helical subdomain (∼280–300) that is the final link to the A domain. The CC segment contains a region (amino acids 233–240) that is sensitive to proteolytic digestion (Melnikov *et al.*, 2008). By analogy with other ATP-dependent proteases, the N-terminal domain of LonA is predicted to participate in recognition and binding of target proteins or their adaptors (Ebel *et al.*, 1999; Iyer *et al.*, 2004; Rotanova *et al.*, 2004). Previously, only the first ∼120 residues of the N-terminal domain have been characterized in structural terms (Li *et al.*, 2005). We have now crystallized and solved the structure of a longer construct consisting of residues 1–245 of *Ec*Lon (Lon-N245), which extends just beyond the proteolytically sensitive region of the CC segment. Because it is not certain that the protease-sensitive region precisely defines a subdomain border, we will refer to the part of the N domain purified and crystallized in this study as the Lon-N245 fragment.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the last shell.

	Data set 1	Data set 2
Data collection		
Wavelength (Å)	0.9756	1.000
Space group	$P3_221$	$P3_221$
Unit-cell parameters (Å)	$a = b = 91.69$, $c = 81.97$	$a = b = 91.58$, $c = 81.86$
Resolution (Å)	30–2.8	30–2.6
No. of reflections (unique/total)	10106 (81603)	12507 (107497)
Completeness (%)	99.9 (99.7)	99.6 (96.3)
$I/\sigma(I)$	35.6 (2.3)	33.5 (2.5)
$R_{\rm merge}$ (%)	7.2 (73.3)	5.3 (53.3)
Refinement		
No. of molecules in asymmetric unit		1
No. of protein atoms		1897
No. of solvent molecules		0
Average B factor (Å^2)		96.3
$R_{\rm cryst}$ (%)		23.5
$R_{\rm free}$ (3% of reflections) (%)		28.2
R.m.s. deviations from ideality		
Bond lengths (Å)		0.011
Bond angles (°)		1.31

Table 2
Twinning analysis of the data sets.

Twinning analysis	Data set 1	Data set 2	Theoretical nontwinned	Theoretical twinned
$\langle I^2\rangle/\langle I\rangle^2$ acentric	2.087	2.055	2.0	1.5
$\langle I^2\rangle/\langle I\rangle^2$ centric ($h0l$)	3.310	3.284	3.0	2.0
$\langle E^2 - 1\rangle$ acentric	0.747	0.735	0.736	0.541
$\langle E^2 - 1\rangle$ centric ($h0l$)	1.120	1.066	0.968	0.736
$N(z)$, $z < 0.1$ acentric	0.091	0.095	0.095	0.018
$N(z)$, $z < 0.1$ centric	0.228	0.253	0.248	0.095
$\langle\|L\|\rangle$	0.492	0.493	0.5	0.375
$\langle L^2\rangle$	0.325	0.327	0.333	0.2
$\langle\|H\|\rangle$	0.466	0.468	0.5	0.0
$\langle H^2\rangle$	0.302	0.308	0.333	0.0

2. Materials and methods

2.1. Protein expression, purification and crystallization

To obtain a construct expressing Lon-N245, a clone of intact *E. coli* Lon from plasmid pLon500 (Maurizi *et al.*, 1985) was moved to the pET30a vector using a two-step procedure. A fragment of *lon* with an internal *Bst*BI deletion was amplified by PCR and inserted into the vector. The *Bst*BI-deleted portion of Lon was then restored by cutting and ligation. Prior to restoration of the fragment, a silent mutation was introduced into the *Bst*BI fragment using the QuikChange procedure to eliminate an internal *Nde*I site. The Lon-N245 reading frame was amplified by PCR using the pET30a *lon* clone as a template, placing a stop codon after codon 245 and providing *Nde*I and *Xho*I restriction sites for cleavage and ligation into pBAD33 under the control of the arabinose regulon. A further modification was made by cleavage of the *Nde*I site and ligation of a short synthetic duplex encoding a tandem array of six histidines in frame immediately following the initiator methionine. For expression of His_6-Lon-N245, the plasmid was transformed into *E. coli* MG1655 carrying a *lon* deletion and cells were grown in LB medium. To incorporate selenomethionine (SeMet) into the expressed protein, the plasmid was transformed into *E. coli* B834 cells (Novagen). Transformed cells were grown in the defined medium recommended by the supplier initially with limiting methionine (10 µg ml^{-1}) to an A_{600} of 0.8. SeMet (30 µg ml^{-1}) was added to the medium and Lon-N245 expression was induced by the addition of 0.2% arabinose for 4 h, after which the cells were harvested and stored frozen.

Cells containing SeMet-modified His_6-Lon-N245 were suspended in 4 ml buffer [50 m*M* Tris–HCl pH 8.0, 0.1 *M* KCl, 10%(*v*/*v*) glycerol] per gram of cells and the suspension was passed once through a French pressure cell at 138 MPa. After clarifying the extract by centrifugation at 30 000*g* for 45 min, the supernatant solution was passed over a Talon metal-chelate affinity column and His_6-Lon-N245 was eluted with a gradient from 0.02 to 1.0 *M* imidazole in the same buffer. Unbound His_6-Lon-N245 was loaded back onto a fresh Talon column after adding KCl to 0.5 *M* and the protein was eluted as before in buffer containing 0.5 *M* KCl. Fractions containing His_6-Lon-N245 were combined and run over a Superdex 200 column equilibrated with 50 m*M* Tris–HCl pH 7.5, 0.4 *M* KCl,

10%(*v*/*v*) glycerol. The fractions of the highest purity were combined, concentrated using a Centricon-10 membrane and dialyzed before setting up for crystallization.

Crystals of SeMet-His$_6$-Lon-N245 were obtained using the hanging-drop vapor-diffusion method. The sample was concentrated to 10 mg ml^{-1} in 20 m*M* Tris–HCl buffer pH 7.5 also containing 1 m*M* EDTA, 0.4 *M* NaCl and 10% glycerol. The well solution contained 0.6 *M* magnesium formate in 0.1 *M* bis-tris buffer pH 5.5. The hanging drop was prepared by mixing 4 µl sample with 4 µl well solution. Crystals could be observed overnight and grew to a length of more than 1 mm in about 10 d.

2.2. Data collection and processing

Two data sets were collected on the SER-CAT beamline 22ID at APS, Argonne National Laboratory from the trigonal crystals of the SeMet derivative of Lon-N245 at 100 K using a MAR 300 CCD detector. The first set was collected at a wavelength of 0.9756 Å, close to the Se absorption edge, and included data extending to a resolution of 2.8 Å. The second, 2.6 Å data set was collected at a wavelength of 1.000 Å. Both data sets were processed with *HKL*-2000 (Otwinowski & Minor, 1997) and the resulting statistics of data processing are shown in Table 1.

Owing to problems encountered subsequently in structure refinement (see below), these data were analyzed in detail in order to verify that the crystals were not subject to twinning or other crystallographic artifacts. Diffraction data from both crystals merged well in point groups 321 and 3, but not in point group 622. Regular absences among the 00*l* reflections clearly demonstrated the presence of a threefold screw axis. The symmetry of these crystals was therefore consistent with either the true or partially merohedrally twinned space groups $P3_121$ or $P3_221$ or perfectly twinned $P3_1$ or $P3_2$. Several tests for twinning were performed and Table 2 shows the results obtained from *phenix.xtriage* (Zwart *et al.*, 2005). The results for both data sets agree well with the statistics expected for nontwinned crystals. Moreover, the clear difference between the statistics obtained for acentric and potentially centric (*h*0*l*) reflections for the 321 point group confirm the presence of twofold axes and therefore exclude the possibility that these crystals display perfectly twinned symmetry in space groups $P3_1$ or $P3_2$. It is known that the simultaneous occurrence of merohedral twinning and translational pseudosymmetry may neutralize the effect of either of these phenomena on the intensity statistics. However, the highest native Patterson peaks obtained from the data sets have heights of 6.8 and 6.0% of the origin maximum, precluding the presence of such effects.

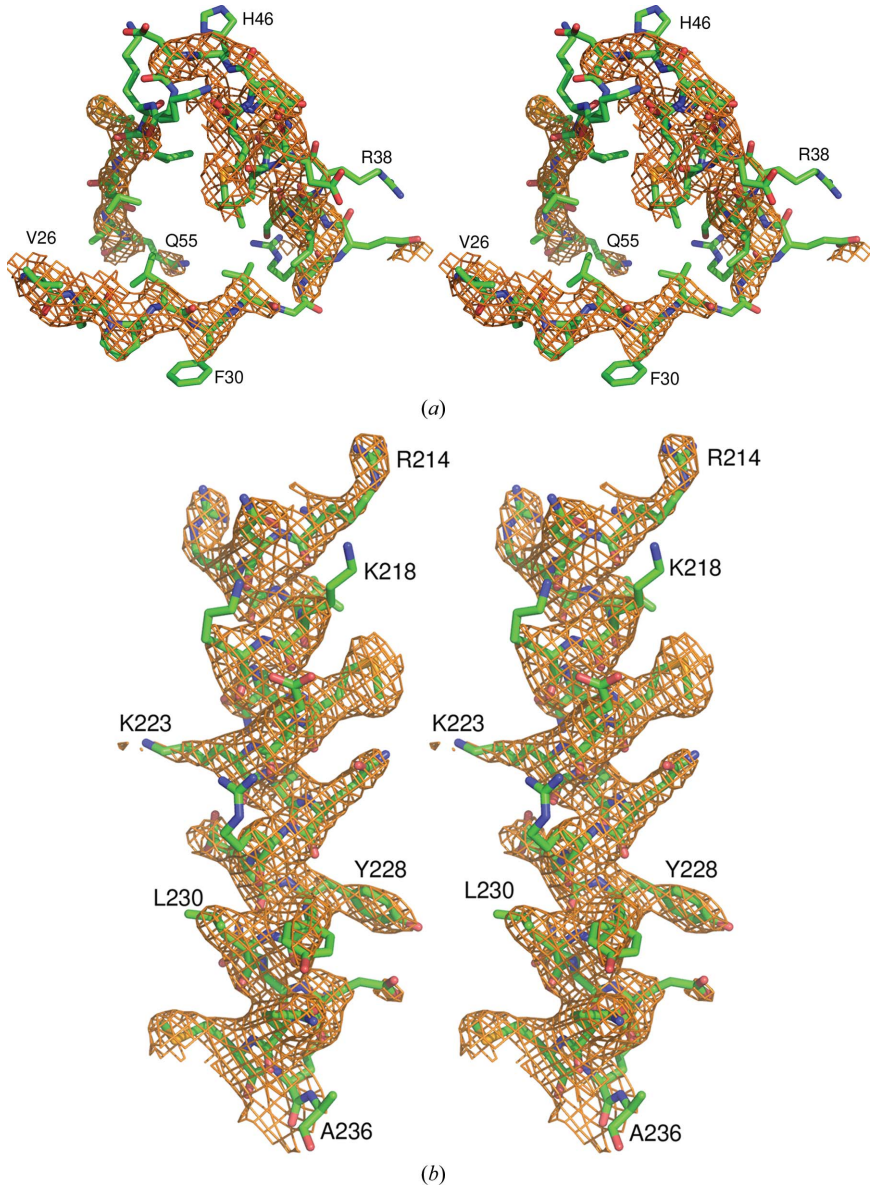

Figure 1
The final $2F_o - F_c$ electron-density maps of Lon-N245. (*a*) Representative density in the N-terminal subdomain. (*b*) Representative density in the C-terminal subdomain. The maps were contoured at the 1.5σ level and selected residues are labeled.

2.3. Structure solution and refinement

The structure of Lon-N245 was solved by single-wavelength anomalous diffraction of crystals of selenomethionine-containing protein. Data from the 2.8 Å set 1 were used to search for Se sites with the programs *SHELXD*/*E* (Sheldrick, 2008). 11 of the 12 expected Se sites (not counting the N-terminal methionine, which was expected to be disordered) were located and served as input into the program *auto-*

SHARP (Global Phasing Ltd, Cambridge) for refinement, phase calculation and extension and density modification using both data sets 1 and 2. The final phases had an FOM of 0.53 for the acentric reflections and 0.38 for the centric reflections when space group $P3_221$ was selected. The scores from density modification were 0.94 for space group $P3_221$ and 0.32 for $P3_121$, unambiguously proving the former to be correct. The resulting electron-density map exhibited good density for the C-terminal subdomain of Lon-N245 (residues 123–245), showing a very clear long helix (residues 189–245) as well as four shorter helical segments. However, the electron density for the N-terminal subdomain was quite poor. The C-terminal subdomain was built both manually with the program *Coot* (Emsley & Cowtan, 2004) as well as automatically with the program *Buccaneer* (Cowtan, 2006). Since the N-terminal subdomain could not be traced based on the experimental map alone, the previously solved structure of the *Ec*Lon N-terminal subdomain (residues 1–119; Lon-N119; Li *et al.*, 2005) was used to assist in model building. Two different approaches were attempted. First, the structure of Lon-N119 was manually rotated and translated into the position where the S atoms of Met25, Met44, Met51 and Met80 overlapped with the Se sites located by *SHELXD* and *SHARP*. The resulting model was inspected in *Coot*, showing good crystal packing with no collisions between the symmetry-related molecules. The second approach involved molecular-replacement analysis with the program *Phaser* (McCoy, 2007) using Lon-N119 as a search model. Such runs were not successful against the original data with no partial model, but succeeded when the coordinates of the C-terminal subdomain were used. For the C-terminal subdomain itself a molecular-replacement run resulted in a *Z* score of 27.1 and an LLG of 598, whereas the *Z* score was lowered to 19.7 and the LLG increased to 790 after the N-terminal subdomain was added. The solution from the *Phaser* run compared well with the manually inserted N-terminal subdomain resulting from the first approach above, verifying the correctness of locating this part of the structure. A post-mortem check of the anomalous difference Fourier map confirmed the positions of all 12 Se atoms, but the peak heights on this map differed considerably, which was in general agreement with the varying refined *B* factors of the Se atoms and the occupancies of the peaks in the *SHELXD* substructure solution (Table 3).

Table 3
Peak heights (in map σ units) in the anomalous difference Fourier map calculated with refined phases, refined *B* factors of the Se atoms and relative occupancies of peaks in the *SHELXD* substructure solution.

SeMet residue	Peak height (σ)	*B* factor (Å^2)	Occupancy
25	9.0	92	0.959
44	5.8	95	—
51	6.7	112	0.559
80	6.0	77	0.500
173	12.3	56	0.893
186	10.3	58	0.907
197	11.7	49	0.842
199	12.8	58	0.743
200	15.6	54	1.000
221	7.7	67	0.501
234	7.7†	88	0.568†
244	9.2	59	0.965

† This peak is shifted to the special position on the twofold axis between two symmetry-equivalent Se234 atoms.

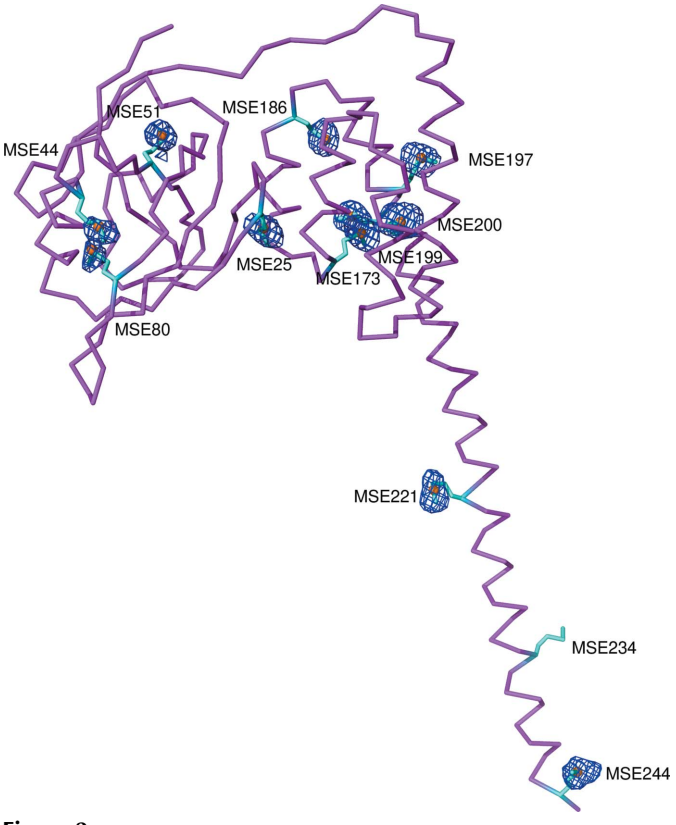

Figure 2
An anomalous difference map showing the experimentally determined location of the Se atoms in the crystals of Lon-N245 superimposed on the backbone trace of the protein, with SeMet residues shown as sticks. The map was calculated with the phases calculated from the final model after deletion of Se atoms and was contoured at the 3.0σ level.

The structure was rebuilt and refined with *Coot* and *REFMAC*5 (Murshudov *et al.*, 1997) using the higher resolution data set 2. Throughout the refinement the electron density for the N-terminal subdomain remained poor and the conventional approach for refinement that utilized rigid-body and positional refinement, annealing and TLS could not lower $R_{\rm free}$ below 30%. Since at this stage we started to suspect that the reasons for such behavior arose from the mobility of the protein molecule, especially its N-terminal subdomain, we used the normal-mode-based refinement algorithm for the final refinement cycles, as implemented in the programs *ENM-CALC* (Lu *et al.*, 2006) and *REFMAC-NM*. Such an approach involves rigid-body refinement, rebuilding with *Coot* and subsequent normal-mode refinement. After the latter refinement, positional refinement that included TLS, but with no refinement of *B* factors, was conducted with *REFMAC*. The procedures were repeated in several rounds of rebuilding and refinement, significantly improving the statistics (Table 1). The final model of Lon-N245 consists of residues 7–245; the His tag and the first six residues at the N-terminus are disordered. In view of the limited resolution of the diffraction data

and the less-than-optimal quality of the electron-density maps (Fig. 1) we did not attempt to model any solvent. The coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession code 3ljc.

3. Results and discussion

3.1. Analysis of the correctness of the structure

Since the electron-density maps were less than fully satisfactory and some nonstandard procedures were utilized during refinement, it was necessary to analyze the resulting structure very carefully in order to verify its correctness. The geometry of the refined model is satisfactory considering the limited resolution of the diffraction data. The Ramachandran plot for the final structure obtained with the program *PROCHECK* (Laskowski *et al.*, 1993) showed 84.4% of the residues in the core region and 15.6% in the additionally allowed region, with none in the generously allowed or disallowed regions. The positions of the anomalous scatterers determined from the *SHELXD* analysis are in good agreement with the final model (Fig. 2, Table 3), further supporting the general correctness of the structure as currently interpreted, although the values for the *R* factors are higher than expected and the values of the *B* factors are also quite high.

3.2. Description of the molecule

Analysis of multiple sequence alignments suggested that LonA N domains might consist of two or more independently folded subdomains with a boundary at around residue 119 (*Ec*Lon numbering). Indeed, crystals of Lon-N119 have been obtained and its structure has been solved in two different crystal forms, yielding 14 crystallographically independent views of this fragment of the molecule (Li *et al.*, 2005). With the exception of some parts of the loops that exhibited limited conformational variability, all of the independent structures agreed well with each other. However, in Lon-N245 this region was quite difficult to trace and the electron density for parts of the molecule was still poor in the current structure even at the conclusion of the refinement process.

The structure of the N-terminal subdomain of Lon-N245 predominantly consists of β-strands (Fig. 3). The first two, $\beta1$ (9–16) and $\beta2$ (26–31), are followed by the sole prominent helix present in this subdomain, $\alpha1$ (34–45). Strand $\beta3$ (49–54) leads to a very irregular wide loop, which is followed by the helical turn $\alpha2$ made by residues 65–67. The remainder of the subdomain is made of a long strand $\beta4$ (71–82), the even longer $\beta5$ (88–105) and the terminal strand $\beta6$ that starts at residue 109 and ends at 123, forming a linker to the second, helical subdomain. Strands $\beta1$, $\beta3$, $\beta4$ and part of strand $\beta5$ form a mixed β-sheet, while strands $\beta2$, the rest of strand $\beta5$ and strand $\beta4$ form an antiparallel sheet that is almost exactly perpendicular to the former. Finally, a third antiparallel sheet is formed by strands 1, 6 and 5.

The C-terminal subdomain is all-helical (Fig. 3), with helix $\alpha3$ (124–145) followed by helices $\alpha4$ (149–159), $\alpha5$ (162–172), $\alpha6$ (181–185) and finally the very long helix $\alpha7$ (189–245). The length of the latter helix is $\sim$85 Å and helices $\alpha6$ and $\alpha7$ belong to the predicted CC region of *Ec*Lon. These five helices, up to and including residues 189–207 of $\alpha7$, form a compact bundle with up–down–up–down–up topology. Owing to crystal packing, helix $\alpha7$ makes extensive contacts with its symmetry mate from another molecule.

3.3. Comparison with other structures

We have previously noted that the structure of Lon-N119 exhibits significant similarity to the structure deposited in the PDB with accession code 1zbo but not described in further detail (F. Forouhar, W. Yong, K. Conover, T. B. Acton, G. T. Montelione, L. Tong & J. F. Hunt, unpublished work). These coordinates represent a hypothetical protein from *Bordetella parapertussis*, BPP1347, that is annotated as having unknown function. This structure was solved at a resolution of 2.6 Å and the asymmetric unit contains two identical molecules with 197 visible residues each. Similarly to Lon-N245, a molecule of BPP1347 consists of two subdomains that are connected by a single extended linker. Superposition of Lon-N245 on molecule *A* of BPP1347 with the program *SSM* (Krissinel & Henrick, 2004) resulted in an r.m.s. deviation of 2.35 Å for 154 C^{α} pairs belonging to both subdomains. A similar superposition with the program *ALIGN* (Cohen, 1997) resulted in an r.m.s. deviation of 2.15 Å for 161 C^{α} pairs, whereas superposition with *DALI* (Holm & Sander, 1993) resulted in an r.m.s. deviation of 3.3 Å for 175 pairs (*Z* score 17.2). Thus, not only are the individual subdomains of Lon-N245 and BPP1347 similar to each other but their mutual disposition is also practically the same (Fig. 4), despite their very low sequence homology (only 34 residues are identical in the aligned segments). This result could not have happened by chance and offers independent support for the correctness of the structure of Lon-N245.

BPP1347 is a representative of a large diverse family of proteins or domains of larger proteins, previously named LAN (Iyer *et al.*, 2004). The conserved-domains database on the NCBI site (http://www.ncbi.nlm.nih.gov/) identified these proteins as members of COG2802, a conserved domain related to the N-terminal region of LonA. The regions of homology within the COG correspond to regions of similarity between *Ec*Lon-N207 and BBP1347.

The search for related structures with *DALI* did not identify any other proteins with a similar overall fold, although the helical part of the structure appeared to exhibit some similarity to a fragment of phenylalanine-ammonia lyase 1 (PDB code 1w27; Ritter & Schulz, 2004). Although the *Z* score of 7.2 is comparatively high, the alignment of 111 residues (of the 690 present in each molecule of phenylalanine-ammonia lyase 1) resulted in an r.m.s. deviation of 4.7 Å and a sequence identity of 10%, indicating that the apparent similarity is not significant. A number of structures of tyrosine aminomutase could be aligned with *Z* scores of 5.4–5.7, again exhibiting no significant similarity in the amino-acid sequence.

However, the overall architecture of Lon-N245 shows some similarity to the architecture of the response regulator RssB

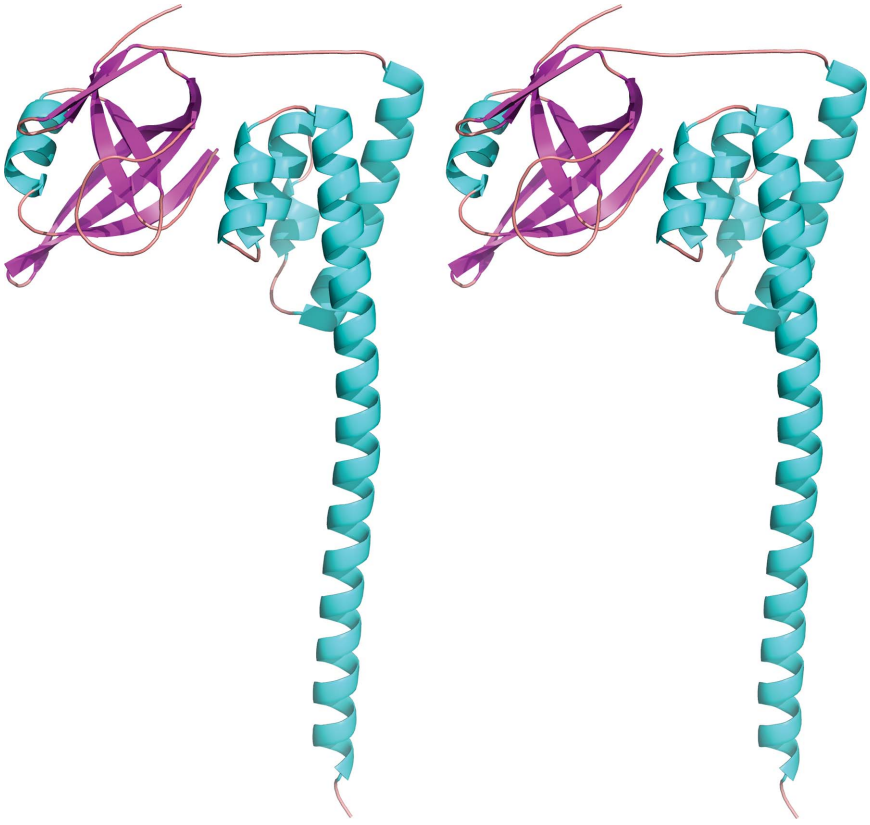

Figure 3
Stereoview of the crystal structure of Lon-N245. The elements of the secondary structure are colored magenta for β-strands, blue for α-helices and brown for coils.

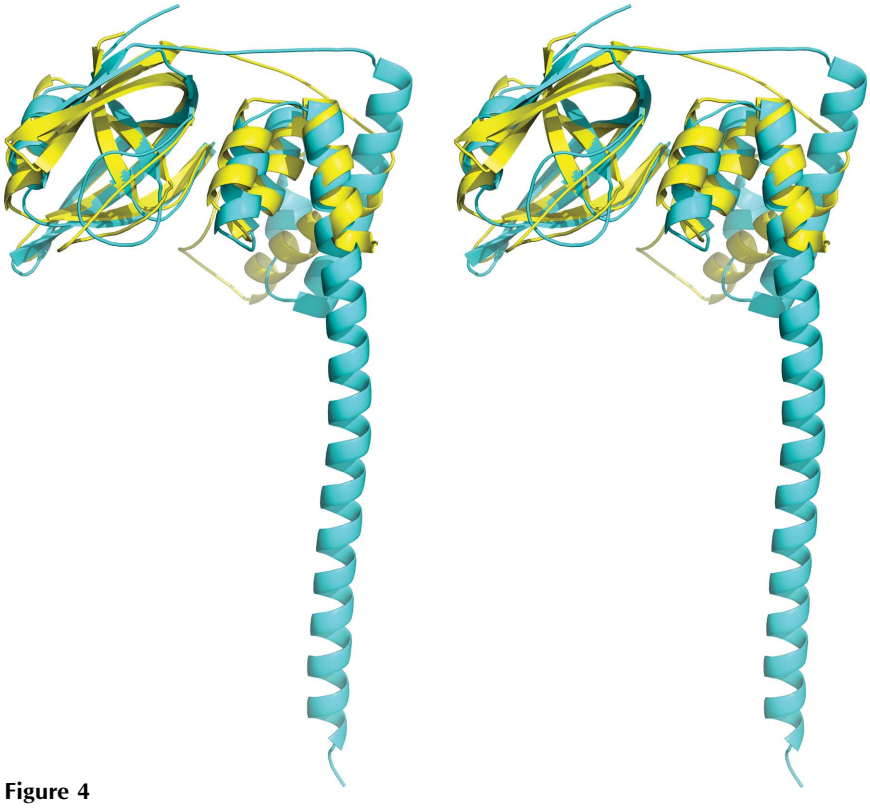

Figure 4
Comparison of the backbone trace of Lon-N245 (blue) with the hypothetical *B. parapertussis* protein BB1347 (yellow).

(Studemann *et al.*, 2003) from *E. coli* and *Pseudomonas aeruginosa*. Structures of this protein have been deposited in the PDB with accession codes 3eq2 and 3f7a, respectively, but have not otherwise been described (I. Levchenko, R. A. Grant, R. T. Sauer & T. A. Baker, unpublished work). RssB is a targeting factor for the ATP-dependent protease ClpXP and, reminiscent of Lon-N245, its structure contains a very long α-helix. However, in the case of RssB the helix connects two widely separated compact domains, neither of which resembles the two compact subdomains of Lon-N245. Considering that the latter molecule contains only a fragment of a larger protease and that the fragment studied here is followed by the C-terminal part of the N-terminal fragment, which is predicted to form a small mostly helical globular subdomain that then joins the ATPase domain, the topological similarity does not seem to be completely far-fetched. Incidentally, similar to the case described here, the refinement of the structures of RssB also presented significant crystallographic difficulties, with much lower resolution of diffraction data and much higher values of the final *R* factors. This fact may indicate that the presence of a long exposed helix might make the protein more flexible and thus make the crystals less able to diffract to high resolution.

3.4. Function of the N-terminal domain of LonA proteases

Direct evidence that the N domain of Lon plays a role in substrate discrimination comes from several directions. Deletion of up to 250 residues from the N-terminus of Lon from *Mycobacterium smegmatis* (Roudiak & Shrader, 1998), *E. coli* (Melnikov *et al.*, 2008) or *Brevibacils thermoruber* (Chir *et al.*, 2009) led to >90% loss of ATP-dependent protein-degrading activity without affecting the peptidase or ATPase activities to a significant degree. These results suggest that the loss of the N-terminus causes a defect in protein–substrate interaction which might arise from impaired oligomerization of the truncated enzyme (Chir *et al.*, 2009; Melnikov *et al.*, 2008). In a separate study (FSR, G. G. Leffers, S. Gottesman and MRM, unpublished work) it has been found that deletion of a portion of the N-terminus abrogates the

ability of *Ec*Lon to degrade RcsA and overexpression of N-terminal fragments of various lengths in Lon$^+$ *E. coli* cells can block the degradation of RcsA. These results also support the model postulating that the N-terminus of *Ec*Lon plays a role in recognizing RcsA. It has also been reported (Ebel *et al.*, 1999) that a point mutation Glu240Lys of *Ec*Lon led to a specific defect in the ability to degrade RcsA *in vivo* but had no effect on another physiological substrate, SulA, suggesting that this region of the N-terminus might be involved in substrate recognition or in a specific interaction needed to process a subset of substrates. Partial proteolysis of Lon by trypsin, chymotrypsin or glutamyl endopeptidase V8 resulted in cleavage near Glu240 in the 233–240 segment of the enzyme (Melnikov *et al.*, 2008; Patterson *et al.*, 2004; Vasilyeva *et al.*, 2002), indicating that this portion of the N-terminus is exposed and perhaps somewhat flexible in the native structure. One might suppose that the Glu240Lys mutation could disrupt the long helix observed in the Lon-N245 crystal and/or prevent formation of a predicted coiled coil to which this helix contributes.

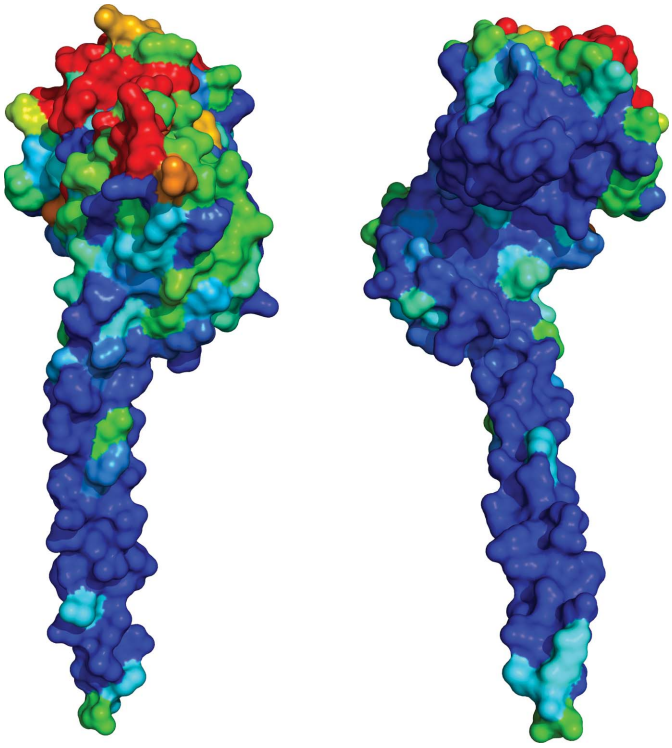

Figure 5
Conservation of the amino-acid sequence among the N domains of LonA proteins. The two views of the cudgel-shaped Lon-N245 are related by ∼180° rotation around the vertical axis. A surface rendering shows highly conserved residues around a putative substrate-binding cleft in the pommel and along the long helical shaft extending toward the C-terminus. The color gradient shows a descending order of conservation: from darkest blue (identical) through light blue, green, yellow and orange to darkest red (nonconserved). Conservation scores were obtained with the program *ConSurf* 2005 (Glaser *et al.*, 2003; Landau *et al.*, 2005) using a multiple sequence alignment of 250 Lon sequences as input. The alignment was generated with the *ClustalW* program at the EMBL–EBI website (http://www.ebi.ac.uk/Tools/clustalw/). This figure was prepared with *PyMOL* (DeLano, 2002).

A potential protein target–LonA interaction region can be identified by examining the nonrandom distribution of highly conserved residues within the fragments of LonA proteases which correspond to Lon-N245 (Fig. 5). A deep hydrophobic cleft found on the edge of the N-terminal subdomain where it joins the C-terminal subdomain is made up of highly conserved residues. Highly conserved mostly hydrophobic residues continue from the cleft across the inside surface of the junction between the pommel and the shaft joint and then extend all along the shaft. Given the known affinity of Lon for unfolded regions of proteins rich in aromatic residues (Gur & Sauer, 2008), this conserved surface covering the deep cleft appears to be an ideal site at which such target regions in Lon substrates could interact.

Whereas helices as long as the C-terminal helix of Lon-N245 (residues 189–245) and almost completely superimposable with it are not uncommon among known protein structures, they are usually found in coiled coils (for example, in the GTPase-activating protein Git1; PDB code 2w6a; Schlenker & Rittinger, 2009). However, despite the very low r.m.s.d. of only 0.8 Å for 56 residues superimposed in this particular case, such a resemblance is most likely just to be a result of the helix being quite regular. With the sequence identity as low as 4%, there is no reason to believe that such apparent structural similarity would be biologically significant. A helix of a comparable length was also seen in the middle of ClpB and Hsp104 subunits (Lee *et al.*, 2003), where it forms a part of a CC region that spans compact protein domains. The CC region may act as a 'molecular crowbar' (Glover & Lindquist, 1998) to disrupt local structural elements in protein aggregates and thus help untangle and solubilize aggregated proteins. It was recently postulated that coiled coils and their surrounding regions in *Ec*Lon and bacterial ClpB exhibit a topological similarity (Rotanova & Melnikov, 2010). However, this postulate needs to be further verified in structural terms.

ATP-dependent proteases interact with substrates in several modes. Recognition and capture are thought to occur in two primary steps: an initial binding event followed by a second step that results in overall tighter binding that requires ATP hydrolysis to disrupt. In some cases, either the first or the second step (or both) can be specific (Gottesman & Maurizi, 1992). SspB-mediated substrate binding to ClpXP involves the adaptor–substrate complex interacting with the N domains of ClpX followed by binding of the C-terminal two amino acids to a site within the ATPase domain of ClpX (Bolon *et al.*, 2004). A similar model holds for ClpS-mediated degradation with ClpAP, except that the specific binding of the N-terminus of the substrate to ClpS is followed by a nonspecific interaction of an unstructured region of the protein with a still-unknown site in ClpA (Erbse *et al.*, 2006; Wang *et al.*, 2008). In LonA, the substrate-recognition sites have not been directly identified, but a similar model of bipartite interaction of substrates has been proposed and one site is likely to be located within the axial channel of the hexamer (Gur & Sauer, 2009). The second site is likely to be within the N-terminal fragment or at least to include a part of it. Such a bipartite mode of interaction implies that the N-terminal fragment of

LonA is dynamic and can help to position bound substrates so that the degradation determinants or degrons have access to the axial channel. The elongated structure of the nearly complete N-terminal fragment presented here indicates that even slight movement of its C-terminal subdomain about the point of attachment to the A domain will result in dramatic movement of the N-terminal subdomain (residues 1–119) and the first helical region (residues 120–220). Such a wide range of movement suggests that those portions undergo a sweeping motion that could survey the area for unfolded proteins or that could be used to exert disruptive forces on bound proteins that it encounters. Higher resolution structures that include both the N-terminal fragment and the A domain are needed to put such a model of the modes of action on firmer footing.

We thank Gerald G. Leffers (NCI) for constructing the pET30a clone of *E. coli* Lon with the synonymous mutation eliminating the endogenous *Nde*I restriction enzyme site. We acknowledge the use of beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT), located at the Advanced Photon Source, Argonne National Laboratory. Use of the APS was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38. This work was supported in part by a grant from the Russian Foundation for Basic Research (Project No. 08-04-00977) to TVR, by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and by Federal funds from the National Cancer Institute, National Institutes of Health under Contract HHSN261200800001E. 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 US Government.

References

Amerik, A. Y., Antonov, V. K., Ostroumova, N. I., Rotanova, T. V. & Chistyakova, L. G. (1990). *Bioorgan. Khim.* **16**, 869–880.
Bolon, D. N., Wah, D. A., Hersch, G. L., Baker, T. A. & Sauer, R. T. (2004). *Mol. Cell*, **13**, 443–449.
Botos, I., Melnikov, E. E., Cherry, S., Khalatova, A. G., Rasulova, F. S., Tropea, J. E., Maurizi, M. R., Rotanova, T. V., Gustchina, A. & Wlodawer, A. (2004). *J. Struct. Biol.* **146**, 113–122.
Botos, I., Melnikov, E. E., Cherry, S., Kozlov, S., Makhovskaya, O. V., Tropea, J. E., Gustchina, A., Rotanova, T. V. & Wlodawer, A. (2005). *J. Mol. Biol.* **351**, 144–157.
Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F., Dauter, Z., Maurizi, M. R., Rotanova, T. V., Wlodawer, A. & Gustchina, A. (2004). *J. Biol. Chem.* **279**, 8140–8148.
Chir, J.-L., Liao, J.-H., Lin, Y.-C. & Wu, S.-H. (2009). *Biochem. Biophys. Res. Commun.* **382**, 762–765.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
DeLano, W. L. (2002). *The PyMOL Molecular Viewer*. San Carlos, California, USA: DeLano Scientific.
Ebel, W., Skinner, M. M., Dierksen, K. P., Scott, J. M. & Trempy, J. E. (1999). *J. Bacteriol.* **181**, 2236–2243.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D. A. & Bukau, B. (2006). *Nature (London)*, **439**, 753–756.
Garcia-Nafria, J., Ondrovicova, G., Blagova, E., Levdikov, V. M., Bauer, J. A., Suzuki, C. K., Kutejova, E., Wilkinson, A. J. & Wilson, K. S. (2010). *Protein Sci.* **19**, 987–999.
Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben Tal, N. (2003). *Bioinformatics*, **19**, 163–164.
Glover, J. R. & Lindquist, S. (1998). *Cell*, **94**, 73–82.
Goldberg, A. L., Moerschell, R. P., Chung, C. H. & Maurizi, M. R. (1994). *Methods Enzymol.* **244**, 350–375.
Gottesman, S. & Maurizi, M. R. (1992). *Microbiol. Rev.* **56**, 592–621.
Gottesman, S., Wickner, S., Jubete, Y., Singh, S. K., Kessel, M. & Maurizi, M. (1995). *Cold Spring Harb. Symp. Quant. Biol.* **60**, 533–548.
Gottesman, S., Wickner, S. & Maurizi, M. R. (1997). *Genes Dev.* **11**, 815–823.
Gur, E. & Sauer, R. T. (2008). *Genes Dev.* **22**, 2267–2277.
Gur, E. & Sauer, R. T. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 18503–18508.
Holm, L. & Sander, C. (1993). *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. & Eom, S. H. (2004). *J. Biol. Chem.* **279**, 53451–53457.
Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. (2004). *J. Struct. Biol.* **146**, 11–31.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben Tal, N. (2005). *Nucleic Acids Res.* **33**, W299–W302.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, A. Y.-L., Hsu, C.-H. & Wu, S.-H. (2004). *J. Biol. Chem.* **279**, 34903–34912.
Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M. & Tsai, F. T. (2003). *Cell*, **115**, 229–240.
Li, M., Rasulova, F., Melnikov, E. E., Rotanova, T. V., Gustchina, A., Maurizi, M. R. & Wlodawer, A. (2005). *Protein Sci.* **14**, 2895–2900.
Lu, M., Poon, B. & Ma, J. (2006). *J. Chem. Theory Comput.* **2**, 464–471.
Maurizi, M. R., Trisler, P. & Gottesman, S. (1985). *J. Bacteriol.* **164**, 1124–1135.
McCoy, A. J. (2007). *Acta Cryst.* D**63**, 32–41.
Melnikov, E. E., Andrianova, A. G., Morozkin, A. D., Stepnov, A. A., Makhovskaya, O. V., Botos, I., Gustchina, A., Wlodawer, A. & Rotanova, T. V. (2008). *Acta Biochim. Pol.* **55**, 281–296.
Melnikov, E. E., Tsirulnikov, K. B. & Rotanova, T. V. (2000). *Bioorgan. Khim.* **26**, 530–538.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. (1999). *Genome Res.* **9**, 27–43.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Park, S.-C., Jia, B., Yang, J.-K., Van, D. L., Shao, Y. G., Han, S. W., Jeon, Y.-J., Chung, C. H. & Cheong, G.-W. (2006). *Mol. Cells*, **21**, 129–134.
Patterson, J., Vineyard, D., Thomas-Wohlever, J., Behshad, R., Burke, M. & Lee, I. (2004). *Biochemistry*, **43**, 7432–7442.
Ritter, H. & Schulz, G. E. (2004). *Plant Cell*, **16**, 3426–3436.
Rotanova, T. V., Botos, I., Melnikov, E. E., Rasulova, F., Gustchina, A., Maurizi, M. R. & Wlodawer, A. (2006). *Protein Sci.* **15**, 1815–1828.
Rotanova, T. V. & Melnikov, E. E. (2010). In the press.
Rotanova, T. V., Melnikov, E. E., Khalatova, A. G., Makhovskaya, O. V., Botos, I., Wlodawer, A. & Gustchina, A. (2004). *Eur. J. Biochem.* **271**, 4865–4871.
Rotanova, T. V., Melnikov, E. E. & Tsirulnikov, K. B. (2003). *Bioorg. Khim.* **29**, 97–99.

Roudiak, S. G. & Shrader, T. E. (1998). *Biochemistry*, **37**, 11255–11263.
Schlenker, O. & Rittinger, K. (2009). *J. Mol. Biol.* **386**, 280–289.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Studemann, A., Noirclerc-Savoye, M., Klauck, E., Becker, G., Schneider, D. & Hengge, R. (2003). *EMBO J.* **22**, 4111–4120.
Vasilyeva, O. V., Kolygo, K. B., Leonova, Y. F., Potapenko, N. A. & Ovchinnikova, T. V. (2002). *FEBS Lett.* **526**, 66–70.
Wang, K. H., Oakes, E. S., Sauer, R. T. & Baker, T. A. (2008). *J. Biol. Chem.* **283**, 24600–24607.
Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl.* **44**, contribution 10.