Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of *Mycobacterium tuberculosis* ClpP1

**Henrik Ingvarsson,[a] María J. Maté,[b]‡ Martin Högbom,[a] Denis Portnoï,[c] Nadia Benaroudj,[d] Pedro M. Alzari,[b] Miguel Ortiz-Lombardía[b]*§ and Torsten Unge[a]***

[a]Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden, [b]Unité de Biochimie Structurale, CNRS URA 2185, Institut Pasteur, 75724 Paris CEDEX 15, France, [c]Unité de Génétique Mycobactérienne, Institut Pasteur, 75724 Paris CEDEX 15, France, and [d]Unité de Génétique Bactérienne et Différenciation, Institut Pasteur, 75724 Paris CEDEX 15, France

‡ Present address: Centro de Investigaciones Biológicas (CSIC), C. Ramiro de Maeztu 9, 28040 Madrid, Spain.
§ Present address: Centro Nacional de Investigaciones Oncológicas, C. Melchor Fndz Almagro 3, 28029 Madrid, Spain.

Correspondence e-mail: miguel.ortiz@cnio.es, torsten@xray.bmc.uu.se

Received 5 September 2006
Accepted 23 November 2006

**PDB References:** *Mt* ClpP1, 2c8t, r2c8tsf; 2ce3, r2ce3sf; 2cby, r2cbysf.

*Mycobacterium tuberculosis* caseinolytic protease ClpP1 (*Mt* ClpP1) is a self-compartmentalized protease consisting of two heptameric rings stacked on top of each other, thus enclosing a catalytic chamber. Within the chamber, which can be reached through two axial pores, each of the 14 identical monomers possesses a serine protease active site. The unfolding and translocation of substrates into the chamber are mediated by associated hexameric ATPases covering the axial pores. Three crystal structures of *Mt* ClpP1, determined by molecular replacement, are presented in this study. Two of the models were refined to a resolution of 2.6 Å and the third to 3.0 Å. It was found that disorder in the handle domain affects the formation and configuration of the tetradecamer and results in condensed structures with larger equatorial pores when compared with ClpPs from other species. Additionally, this disorder accompanies conformational changes of the residues in the catalytic triad. The models also reveal structural differences within the N-terminal hairpin-loop domain, which possibly reflect the significant differences in amino-acid sequence between *Mt* ClpP1 and other ClpP homologues in this region.

## 1. Introduction

Caseinolytic proteases (Clp; also called Ti) are involved in a number of cellular activities, such as the degradation of proteins misfolded as a result of various types of stress (Kruger *et al.*, 2000; Frees & Ingmer, 1999; Gerth *et al.*, 1998), the regulation of short-lived proteins and the housekeeping removal of dysfunctional proteins, which include denatured and aggregated polypeptides (Maurizi, 1992; Aizenman *et al.*, 1996; Tomoyasu *et al.*, 2001). Clp is also implicated in the control of cell growth, targeting the DNA-binding protein from starved cells (Dps; Stephani *et al.*, 2003). From a medical perspective, it is noteworthy that Clp has been linked to the tight regulation of virulence genes in the pathogens *Listeria monocytogenes* (Gaillot *et al.*, 2000) and *Salmonella typhimurium* (Webb *et al.*, 1999).

In order to function as an effective protease, an assembly of chaperone ATPases including ClpA and ClpX has to be associated with the self-compartmentalized protease ClpP. ClpA and ClpX are members of the Clp/Hsp100 family of molecular chaperones that belongs to the AAA+ superfamily of ATPases associated with various cellular activities. Instead of ClpA, *Mt* possesses its orthologue ClpC (Shanklin *et al.*,

1995). Other Gram-positive bacteria similarly have the orthologues ClpC or ClpE in place of ClpA. The Clp ATPases oligomerize into hexamers and, together with ClpP, form a large hetero-oligomeric complex (Kessel *et al.*, 1995; Grimaud *et al.*, 1998).

The ClpA and ClpX ATPases provide substrate specificity and also activate ClpP for proteolytic activity (Gottesman *et al.*, 1997) by unfolding their globular protein substrates. The proteolytic capacity of ClpP is dramatically reduced in the absence of the Clp ATPases and only short peptides of up to six amino acids can be degraded (Thompson & Maurizi, 1994). The association with the ATPase is required even for the degradation of certain peptides, although in such cases ATP binding but not hydrolysis is necessary (Thompson *et al.*, 1994).

Homologous counterparts of both ClpP and the Clp ATPases have been found in a wide range of prokaryotic and eukaryotic species with similar structures and functions (Schirmer *et al.*, 1996; Maurizi *et al.*, 1990; Clarke *et al.*, 1994; Shanklin *et al.*, 1995). Homologues of ClpX and ClpP are found within the mitochondria and chloroplasts of higher eukaryotes, but their functions in these organelles are not fully known (Bross *et al.*, 1995; Corydon *et al.*, 1998; Adam & Clarke, 2002).

Presently, the structures of ClpPs from three species have been published and made available in the Protein Data Bank (PDB; Berman *et al.*, 2000), namely those from *Escherichia coli* (*Ec*), PDB codes 1tyf (Wang *et al.*, 1997), 1yg6, 1yg8 (V6A mutant; Bewley *et al.*, 2006) and 2fzs (bound to a dipeptide substrate; Szyk & Maurizi, 2006), from *Homo sapiens* (*Hs*), PDB code 1tg6 (Kang *et al.*, 2004), and from *Streptococcus pneumoniae* (*Sp*), PDB code 1y7o (A153P mutant; Gribun *et al.*, 2005). Moreover, a structure of *Plasmodium falciparum* (*Pf*) ClpP has been deposited with PDB code 2f6i by A. Mulichak and coworkers from the Structural Genomics Consortium (SGC). These structures show that the enzyme forms tetradecameric complexes consisting of two heptameric rings stacked head-to-head on top of each other, thus confining a cavity with two axial pores. The narrowness of these pores, which range from 10 to 20 Å in width, suggests that the passage of polypeptides requires the unfolding of their native structure. This is achieved with the aid of the associated ClpP ATPase, which is coaxially stacked onto the pores of ClpP. Substrates are then unfolded and translocated through the continuous Clp ATPase–ClpP channel down into the proteolytic cavity of ClpP (Ortega *et al.*, 2000, 2002; Schirmer *et al.*, 1996). ATP plays a complex role in this process. On one hand, the nucleotide state of the Clp ATPase influences the stability of the Clp ATPase–ClpP complex (Singh *et al.*, 1999). On the other hand, ATP hydrolysis is required for the unfolding and translocation of the substrate.

Structurally, each ClpP monomer can be divided into a globular domain, referred to as the 'head domain', which forms the apical parts of the double ring, and an elongated domain, referred to as the 'handle', consisting of the $\beta$9 strand and the $\alpha E$ helix, which constitutes the equatorial wall of the tetradecamer (for the sake of consistency in comparison, we will use the numbering of secondary-structural elements as defined for the first structure of *Ec* ClpP; Wang *et al.*, 1997). The handle domains protrude from the head domains in each heptameric ring and interact with their counterparts from the opposing ring, giving rise to the tetradecameric structure. Inside the proteolytic chamber, each monomer contributes an active site, which is formed by a Ser-His-Asp catalytic triad typical of serine proteases. The triad is situated at the junction between the head and the handle domains.

The N-termini of the monomers of the *Ec* ClpP, *Sp* ClpP, *Hs* ClpP and *Pf* ClpP structures adopt a looped-out conformation that has been found to display a high plasticity (Kang *et al.*, 2004; Gribun *et al.*, 2005; Bewley *et al.*, 2006; Szyk & Maurizi, 2006). Mutations in this region can dramatically affect the ability of ClpP to interact stably with the Clp ATPases (Kang *et al.*, 2004; Gribun *et al.*, 2005; Bewley *et al.*, 2006).

In this paper, we report the cloning, expression, purification, crystallization and X-ray structure of the tetradecameric *Mt* ClpP1, annotated at the Pasteur Institute *TubercuList* server (http://genolist.pasteur.fr/TubercuList/) as Rv2461c. The structure of *Mt* ClpP1 reveals a partly disordered handle domain, a slightly rotated arrangement of the monomers and an extended $\alpha A$ helix at the N-terminus, traits that distinguish this structure from those of its orthologues. Whereas disorder in the handle domain has been described for other ClpP structures (Gribun *et al.*, 2005; PDB entry 2f6i), the N-terminal configuration of *Mt* ClpP1 is unique, possibly reflecting its sequence divergence in this region. Possible functional implications of these structural variations are discussed.

## 2. Materials and methods

### 2.1. Subcloning and protein purification

The *clpP1* gene from *Mt* (H37Rv) was cloned independently for the three structures presented here, albeit following a similar strategy. In a typical procedure, the Rv2461c gene encoding ClpP1 was PCR-amplified from *Mt* (H37Rv) genomic DNA using Pfu polymerase (Stratagene) and the primers CACCATGGCTAGCCAAGTGACTGACATGCG (forward) and CTGTGCTTCTCCATTGACGTG (reverse). A subsequent PCR reaction introduced a C-terminal $His_4$ tag into the PCR product. In this reaction, the reverse primer was TCAATGATGATGATGGCCCTGTGCTTCTCCATTGACGTG. The PCR product was ligated into the vector pET 101/D-TOPO (Invitrogen) and transformed into *E. coli* Top10 cells (Invitrogen) for cloning. Plasmids were prepared according to the Qiaprep Spin Miniprep kit (Qiagen) protocol.

*E. coli* BL21 Star (DE3) cells (Invitrogen) were used for protein production at 310 K. Expression of the *Mt* ClpP1 gene was induced with 100 µg ml$^{-1}$ isopropyl 1-thio-$\beta$-D-galactopyranoside (Sigma) and growth was continued for 2.5 h. Cells were collected by centrifugation.

The cell pellet was suspended in lysis buffer [50 m*M* $NaH_2PO_4$ pH 8.0, 300 m*M* NaCl, 10 m*M* imidazole, 20%(*v*/*v*) glycerol] with 0.01 mg ml$^{-1}$ RNase and 0.02 mg ml$^{-1}$ DNase and lysed with a Constant Cell Disruptor (Constant Systems

**Table 1**
Data-collection and refinement statistics.

Values in parentheses refer to the highest resolution shell.

| Structure | 2c8t | 2ce3 | 2cby |
|---|---|---|---|
| Detector | MAR Research CCD | ADSC CCD | MAR Research CCD |
| Source | ESRF ID14-3 | ALS beamline 5.0.2 | ESRF ID14-3 |
| Wavelength (Å) | 0.931 | 1.000 | 0.934 |
| Space group | $P2_1$ | $P2_1$ | $P6_122$ |
| Unit-cell parameters (Å, °) | $a$ = 96.7, $b$ = 168.2, $c$ = 103.8, $\alpha$ = 90.0, $\beta$ = 114.6, $\gamma$ = 90.0 | $a$ = 97.7, $b$ = 169.0, $c$ = 104.4, $\alpha$ = 90.0, $\beta$ = 114.8, $\gamma$ = 90.0 | $a$ = 178.3, $b$ = 178.3, $c$ = 264.9, $\alpha$ = 90.0, $\beta$ = 90.0, $\gamma$ = 120.0 |
| Resolution range (Å) | 59.4–3.0 (3.16–3.00) | 49.5–2.6 (2.69–2.60) | 76.7–2.6 (2.74–2.6) |
| No. of measured reflections | 228076 | 538671 | 1706805 |
| No. of unique reflections | 60193 | 90593 | 76728 |
| Average multiplicity | 3.8 (3.8) | 6.0 (3.8) | 22.2 (22.5) |
| Completeness (%) | 100.0 (100.0) | 100.0 (95.7) | 100 (100) |
| $R_{sym}(I)$ | 0.102 (0.405) | 0.08 (0.45) | 0.16 (0.69) |
| $I/\sigma(I)$ | 5.9 (1.9) | 10.0 (1.7) | 24.3 (6.6) |
| $V$ (Å$^3$) | 1535009 | 1563676 | 7289852.5 |
| No. of molecules in ASU | 14 | 14 | 7 |
| Resolution range (Å) | 19.9–3.0 | 169.0–2.6 | 154.0–2.6 |
| No. of reflections | 57087 | 90592 | 76647 |
| Test set (% of data) | 5 | 5 | 5 |
| $R$† (%) | 21.3 | 20.5 | 19.4 |
| $R_{free}$ (%) | 23.8 | 26.1 | 22.9 |
| Non-H atoms | 18508 [2394 residues] | 18114 [2338 residues] | 9149 [1183 residues] |
| Solvent atoms | n/a | 117 | 301 |
| Mean $B$ factor‡ (Å$^2$) | 51.5 | 29.3 | 19.6 |
| Ramachandran plot outliers§ (%) | 2.3 | 2.2 | 1.3 |
| R.m.s.d. from ideal values¶ | | | |
| Bond lengths (Å) | 0.014 | 0.012 | 0.014 |
| Bond angles (°) | 1.447 | 1.387 | 1.480 |

† $R = \sum_{hkl}(F_{obs} - kF_{calc})/\sum_{hkl} F_{obs}$. ‡ Average of the residual $B$ factors, *i.e.* not including the TLS component. § Percentage of non-Gly, non-Pro residues in disallowed regions calculated using a strict-boundary Ramachandran plot (Kleywegt & Jones, 1996). ¶ Using the parameters of Engh & Huber (1991).

Ltd). After centrifugation, the supernatant of the lysate was incubated with Ni–NTA agarose (Qiagen; 1 ml per 10 mg His-tagged protein) for 40 min at 277 K in a batch procedure. The matrix was washed with lysis buffer containing 20 m*M* imidazole and 10%(*v*/*v*) glycerol and the protein was eluted with the same buffer containing 250 m*M* imidazole and 10%(*v*/*v*) glycerol. The eluate was further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex-200 column (Amersham Biosciences) and developed with a buffer containing 20 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl and 10% glycerol. The buffer was changed on a PD-10 Desalting Column (Amersham Biosciences) to 100 m*M* trisodium citrate pH 5.2 and concentrated to 7 mg ml$^{-1}$ in a Vivaspin concentrator (Vivascience, 10 kDa molecular-weight cutoff).

### 2.2. Crystallization, data collection and data processing

Initial crystal screening was performed with either JBScreen HTS I (Jena Bioscience) or a grid of ammonium sulfate/pH conditions by the sitting-drop vapour-diffusion technique (Ducruix & Giegé, 1999) at room temperature. For the 2c8t structure, the crystallization drop contained 5 mg ml$^{-1}$ protein, 100 m*M* trisodium citrate pH 5.2, 3% PEG 4000. Crystals appeared within 48 h and measured approximately 0.2 × 0.2 × 0.3 mm. Prior to flash-cooling in liquid nitrogen, crystals were soaked in a cryoprotectant consisting of 20% glycerol, 100 m*M* trisodium citrate pH 5.2 and 12% PEG 4000.

For the 2cby structure, crystals were grown from a 1:1 mixture of 3 mg ml$^{-1}$ protein in 50 m*M* Tris buffer pH 8.0 and a reservoir solution consisting of 1.0 *M* ammonium sulfate and 100 m*M* HEPES pH 8.0. Crystals were flash-cooled in a gas beam at 120 K after being soaked in a solution including 1.2 *M* ammonium sulfate, 100 m*M* HEPES pH 8.0 and 20%(*v*/*v*) glycerol. Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF) beamline ID-14-3 at a fixed wavelength of 0.931 Å. Diffraction data were indexed and integrated using *MOSFLM* (Leslie, 1999) and further scaled and averaged using *SCALA* (Evans, 1993) as implemented in the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994).

The 2ce3 structure was obtained from crystals grown by the sitting-drop vapour-diffusion method with drops containing 400 nl protein solution and 400 nl mother liquor equilibrated against 100 µl reservoir consisting of 9.5%(*w*/*v*) PEG 2000, 200 m*M* lithium sulfate and 100 m*M* MOPS (4-morpholinopropanesulfonic acid) pH 6.5. Crystals appeared after 2 d.

Data-collection and processing statistics are summarized in Table 1.

### 2.3. Structure solution, model building and refinement

In all three cases, the structure of *Mt* ClpP1 was solved by molecular replacement (MR). The 2c8t structure was solved using *MOLREP* (Vagin & Teplyakov, 1997) from the *CCP*4 suite. A polyalanine model of the tetradecameric

structure of *Ec* ClpP (PDB code 1tyf; Wang *et al.*, 1997) was used as the search model. The 2cby structure was solved using the stand-alone version of *AMoRe* (Navaza, 2001) from a polyalanine model derived from one heptamer of the same structure 1tyf. Finally, the 2ce3 structure was solved using the 2cby heptamer as the search model.

The MR solution for the 2c8t structure first underwent rigid-body refinement in the program *REFMAC*5 (Murshudov *et al.*, 1997), defining each monomer as a rigid unit. Refinement was then continued with *NCSref* (Evans, 1993; Potterton *et al.*, 2003; Ducruix & Giegé, 1999). Based on the resulting model, a density map and symmetry operators were calculated which were used as input to a 14-fold NCS averaging with *RAVE* (Kleywegt & Jones, 1999). Rebuilding and refinement of the tetradecamer were performed in an iterative manner using NCS averaging in *O* (Jones *et al.*, 1991) and electron-density maps generated from *REFMAC*5 refinement. The final rounds of refinement were performed with TLS parameterization in *REFMAC*5 (Winn *et al.*, 2001). The tight NCS restraints option was used throughout the refinement. Statistics relating to the final refined structure are summarized in Table 1.

The MR solution for the 2cby structure was fed into the *ARP*/*wARP* program (Perrakis *et al.*, 1999) to follow an automated model-building procedure combined with iterative structure refinement. *ARP*/*wARP* built more than 1000 residues in the density, of which ∼900 were assigned the correct amino-acid side chain. Further rebuilding and refinement of the model were performed iteratively using the programs *Coot* (Emsley & Cowtan, 2004) and *REFMAC*5, respectively. No NCS constraints were used throughout the refinement, but a TLS model was applied in the last refinement runs in *REFMAC*5, with a TLS domain for each monomer. The 2ce3 structure was rebuilt and refined in a similar way, although the *ARP*/*wARP* program was not needed owing to the completeness of the MR model. Refinement statistics for the 2cby and 2ce3 structures are summarized in Table 1.

Figures were prepared with *O*, *MOLRAY* (Harris & Jones, 2001), *INDONESIA* (D. Madsen, P. Johansson and G. J. Kleywegt, unpublished program) and *Chimera* (Emsley & Cowtan, 2004). Evolutionary sequence-conservation scores were calculated using the *ConSurf* server (Landau *et al.*, 2005), *via* its implemented empirical Bayesian method, from a multiple alignment consisting of 243 ClpP sequences with less than 90% identity.

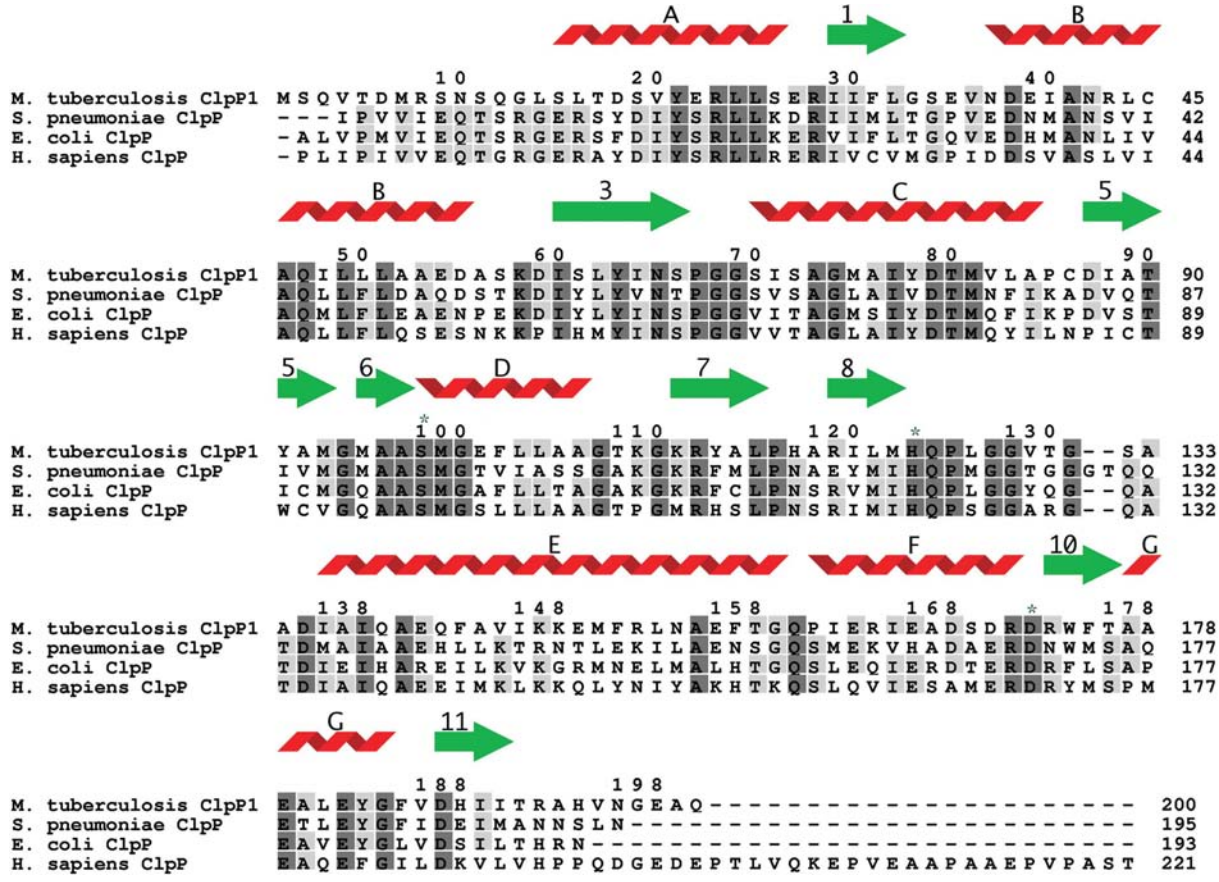


**Figure 1**
Sequence alignment of ClpP homologues corresponding to the NCBI entries for *M. tuberculosis* ClpP1 (gi:13882259), *S. pneumoniae* ClpP (gi:66877867), *E. coli* ClpP (gi:3318866) and *H. sapiens* ClpP (gi:963048) (Benson *et al.*, 2003). The numbering of amino acids is referred to *Mt* ClpP1. Conserved residues are boxed in dark grey and similar residues in light grey. Residues marked with green asterisks correspond to the residues in the catalytic triad. Secondary-structure elements correspond to the structure of *Mt* ClpP1. The secondary-structure elements are labelled according to the structure of *Ec* ClpP1 (Wang *et al.*, 1997).

## 3. Results

### 3.1. The ClpP1 monomer

The *Mt clpP1* gene encodes a polypeptide consisting of 200 amino acids. An alignment of the *Mt* ClpP1, *Hs* ClpP, *Ec* ClpP and *Sp* ClpP peptide sequences shows a high degree of similarity (Fig. 1). The percentage of identical amino acids between *Mt* ClpP1 and the other four homologues are 51.3% for *Ec* ClpP, 45.4% for *Sp* ClpP and 42.2% for *Hs* ClpP. *Pf* ClpP (not shown in the alignment), which has an ∼175-amino-acid sequence N-terminal to the ClpP domain, is the most divergent of the ClpP proteins of known structure, with only 38.2% identity to *Mt* ClpP1. The *Mt clpP1* gene belongs to a genomic region that includes genes for a second ClpP subunit named ClpP2 (Rv2460c), for the ATPase ClpX (Rv2457c) and for the chaperone trigger factor protein Tig (Rv2462c). The second amino acid in the coding sequence of *Mt* ClpP1 (Ser2) aligns with the first amino acid (Ala1) of the mature form of *Ec* ClpP (throughout this paper, the sequence numbering refers to *Mt* ClpP1). However, this does not imply the absence of a prosequence in *Mt* ClpP1. In fact, the most probable start codon of the ClpP1 gene from the closely related *Streptomyces coelicolor* (from the *Actinomycetales* order) aligns with that of *Mt* ClpP1 and yet it is known that *Sc* ClpP1 is processed upstream of its eighth amino acid (Viala & Mazodier, 2002). Therefore, *Mt* ClpP1 is also expected to require a processing step to become an active peptidase. However, the recombinant *Mt* ClpP1 in this study was purified as nonprocessed protein.

In the three models presented here, we could build the stretches Ser15–Pro125/Lys126 and Ile136–Thr191/Val195 in the electron density. The overall fold of *Mt* ClpP1 is similar to the previously solved structures of *Ec* ClpP, *Sp* ClpP, *Hs* ClpP and *Pf* ClpP (Wang *et al.*, 1997; Gribun *et al.*, 2005; Kang *et al.*, 2004; Fig. 2*a*).

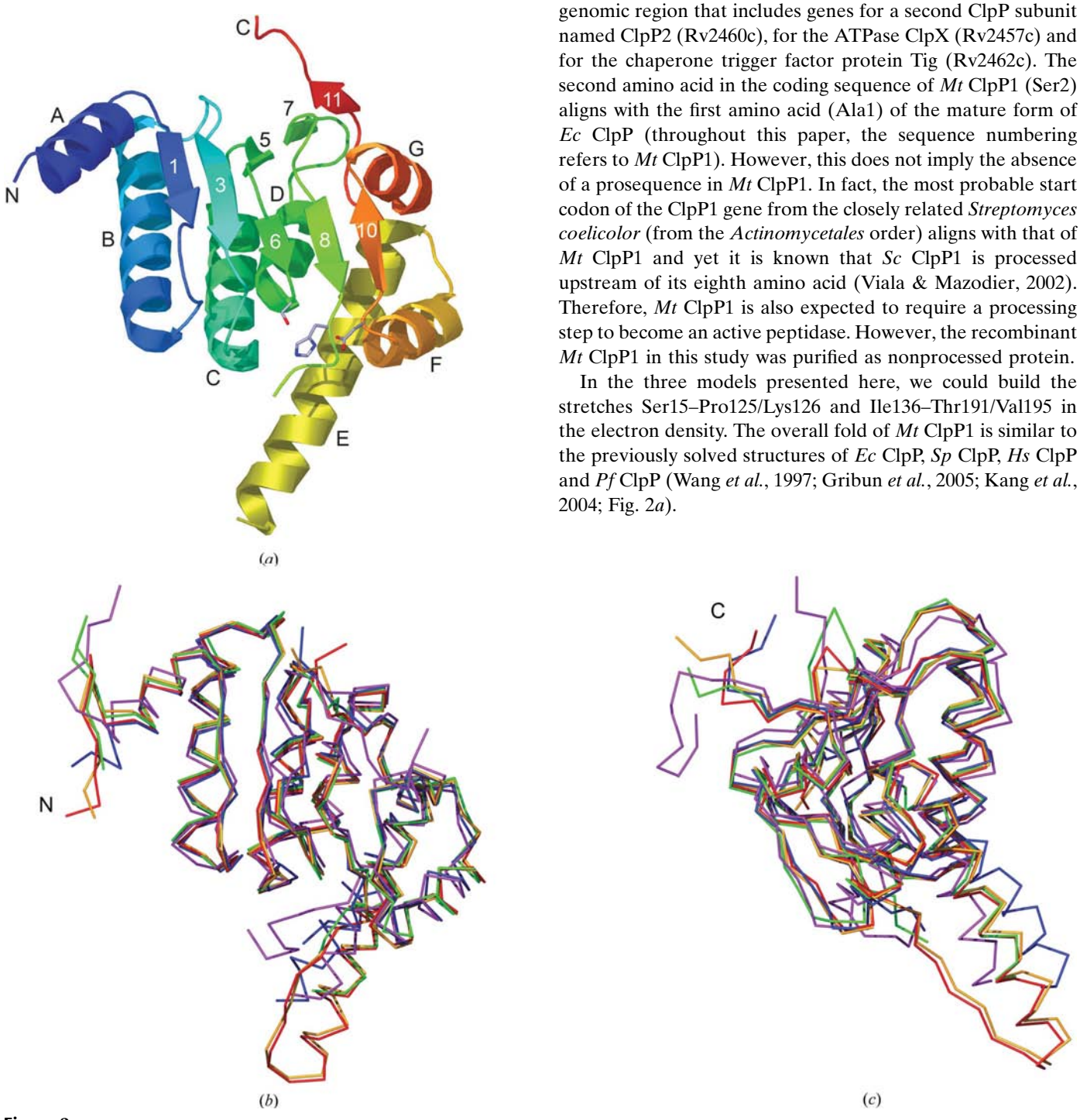


**Figure 2**
Structure of the *Mt* ClpP1 monomer. (*a*) Ribbon-diagram representation, with labelling of the secondary-structure elements according to the structure of *Ec* ClpP1 (Wang *et al.*, 1997). The chain is coloured from blue to red from the N-terminus to the C-terminus. The catalytic triad (Ser98, His123 and Asp172) is shown in stick representation. (*b*) A comparison of the $C^{\alpha}$ backbone of ClpP monomers corresponding to *Mt* (blue), *Sp* (green), *Ec* (gold), *Pf* (violet) and *Hs* (red). Amino acids from the head domains (residues 28–124 and 160–190 in *Mt* ClpP1) were selected for superimposition. (*c*) A 90°-rotated view from (*b*) showing a significant shift of the $\alpha E$ helix in the handle domain of the *Mt* ClpP1 monomer.

In *Mt* ClpP1, the amino acids Arg28–Gln124 and Gln160–Ile190 define the head domain, which is characterized by two perpendicular $\beta$-sheets with five $\alpha$-helices (*B*, *C*, *D*, *F* and *G*) positioned between them (Fig. 2*a*). The first sheet is formed by the parallel strands $\beta$1 and $\beta$2, which are actually a single strand in the present structures, $\beta$3, $\beta$5, $\beta$7 and $\beta$11. The second sheet is built by the parallel strands $\beta$4, $\beta$6 and $\beta$8 together with the antiparallel strand $\beta$10.

In turn, the amino acids Pro125–Gly159 make up the handle domain in the form of a $\beta$-turn–helix motif ($\beta$9 and $\alpha E$, respectively). Helix $\alpha E$ protrudes from the head domain and is anchored to the head domain through interactions with helices *C*, *D* and *F*. In the structures reported here, the first ten residues of this domain, comprising strand $\beta$9, are disordered.

The N-terminus of the *Mt* ClpP1 structure is visible from residue Ser15. Here, residues Ser15–Glu27 constitute a helix

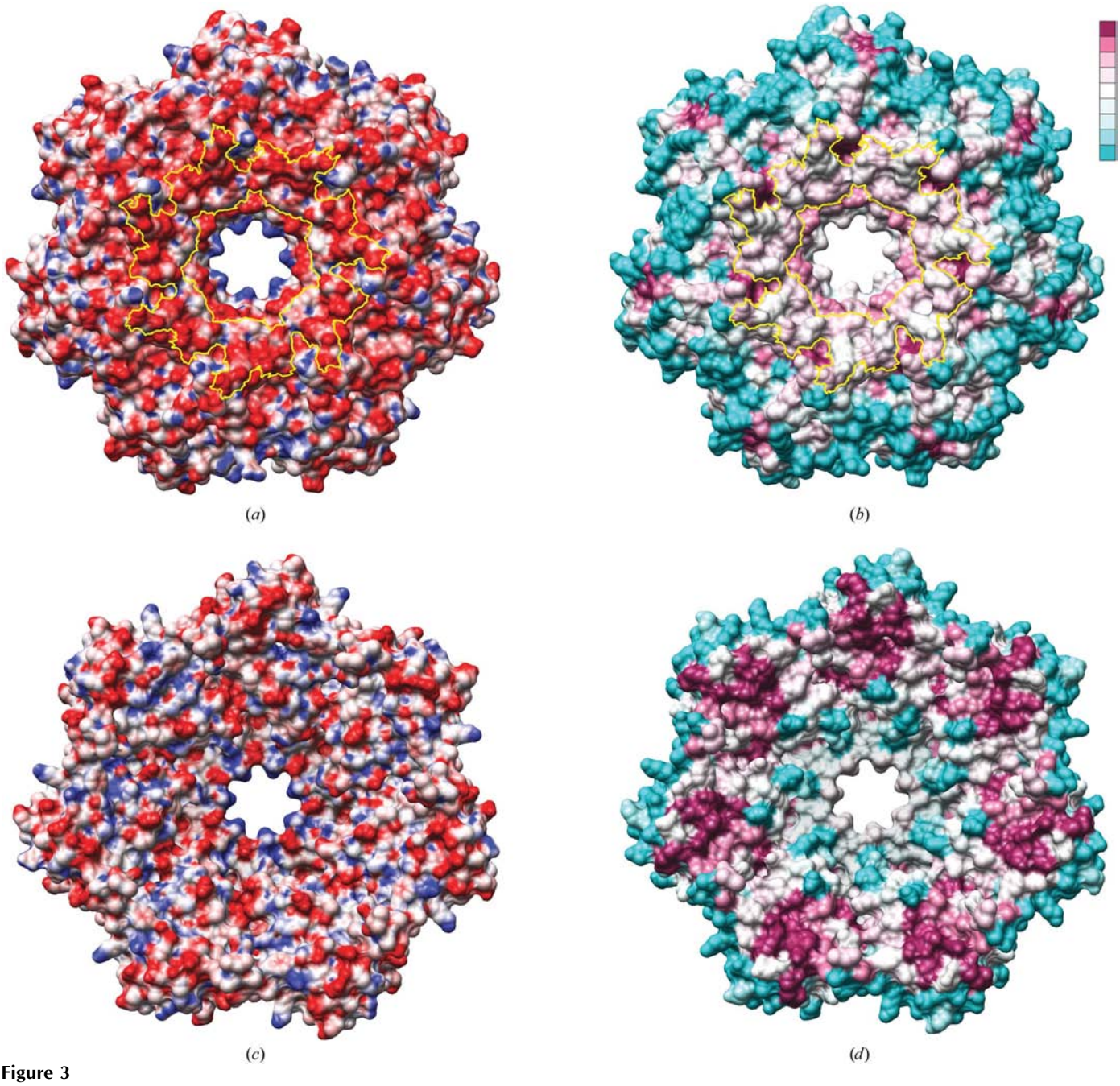


**Figure 3**
Entrance and cavity surfaces of *Mt* ClpP1. (*a*) Electrostatic potential mapped on the external surface of the *Mt* ClpP1 heptamer. Colours range from positive (+20 *kT*/e in blue) to negative (−20 *kT*/e, in red) electrostatic potential. (*b*) Sequence conservation mapped on the same surface as (*a*). In (*a*) and (*b*) two yellow lines trace the outer rim and the narrow part of the axial pore as described in the text. (*c*) Electrostatic potential mapped on the surface of the cavity vault defined by the *Mt* ClpP1 heptamer. (*d*) Sequence conservation mapped on the same surface as (*c*). The high conservation in the zone of the catalytic residues is clearly visible. The vertical bar shows the colour range from low (bottom) to high (top) conservation, determined *via* the *ConSurf* server, as mentioned in §2.

$\alpha A$ that is longer than those seen in previous ClpP structures. This helix protrudes away from the head domain (Fig. 2*a*) into the axial pore of the oligomer.

The C-terminal residues Thr191–Gln200 also project from the head domain to extend almost parallel to the upper rim of the heptameric ring. In our three *Mt* ClpP1 models, the last visible C-terminal residue varies from Thr191 to Val195.

The active site containing the catalytic triad Ser98, His123 and Asp172 is situated at the base of the head domain at the interface with the handle domain. The configuration of the catalytic triad differs from one monomer to another, with the greatest difference being the rotamer adopted by His123. In all cases, the catalytic Ser98 points to the 'oxyanion hole' defined by the amino groups of the conserved amino acids Gly69 and Met99.

We compared the overall structure of the *Mt* ClpP1 monomer with those of *Ec* ClpP, *Sp* ClpP, *Pf* ClpP and *Hs* ClpP by superimposing the head-domain amino acids 28–124 and 161–190 on their equivalent residues in the homologue structures (Figs. 2*b* and 2*c*). We observed root-mean-square (r.m.s.) deviations of 0.79, 0.88, 0.71 and 0.79 Å over the positions of the $C^{\alpha}$ atoms, respectively. Such a striking similarity across such a wide representation of phyla underlines the two most particular traits of the *Mt* ClpP1 monomer, namely the longer $\alpha A$ helix, which penetrates deeper in the axial pore, and a different orientation of the $\alpha E$ helix with respect to the handle domain. The loop comprising residues Gly159–Pro161, which connects the $\alpha E$ helix to the head domain, acts as a hinge in the reorientation of the helices $\alpha E$ and $\alpha F$. This reorientation is also accompanied by the closest packing of the $\alpha F$ helix against the main body of the handle domain.

### 3.2. The heptamer

The first order of oligomerization of ClpP proteins consists of a heptameric ring (Fig. 3). The diameter of the *Mt* ClpP1 heptamer is ~105 Å. In the structures described here, the pore in the centre of this ring has a diameter of ~15 Å when the main-chain amine from Ser15 is considered as the innermost atom. However, since the first 14 amino acids are disordered in these structures, there is uncertainty about the real size of the pore. The thickness of the ring, excluding the handle domain, is ~39 Å. The heptameric ring is stabilized by two sets of interactions: firstly through contacts between helices $\alpha A$, $\alpha B$ and $\alpha C$ of one monomer and helix $\alpha A$, the first $\beta$-sheet and the C-terminus of the next monomer and secondly *via* interactions between the $\alpha E$ helix of the former monomer and the second $\beta$-sheet of the adjacent monomer. The interactions are mainly hydrophobic, but also include hydrogen bonds between the pairs Glu22/Arg23, Asp38/Asn65, Gln42/Gly33, Asn42/Asn65, Arg43/Thr17, Leu83/His194, Glu141/Trp174 and Asn142/Arg119 and the possible ion pairs Glu54/Arg23, Asp79/His117 and Glu149/His117. According to the *PISA* server at the EBI (Krissinel & Henrick, 2005), the average interaction area between two adjacent monomers is around 1365 $Å^2$. Therefore, for a particular monomer, approximately 2730 $Å^2$ is buried upon formation of the heptamer, corresponding to 30.2% of the total surface area of 9047 $Å^2$.

### 3.3. The tetradecamer and the proteolytic chamber

The tetradecameric structure is formed by a twofold symmetrical arrangement of two heptameric rings (Figs. 4*a* and 4*d*). This twofold symmetry is crystallographic for the 2cby structure and noncrystallographic for the other two structures. The tetradecamer has a molecular weight of ~303 kDa and measures ~85 Å in height. The stacked heptamers confine the proteolytic cavity, which can be reached through the axial pores of each heptameric ring. The cavity is an ellipsoid that measures ~68 Å in the plane perpendicular to the pore axis and ~47 Å along the pore axis. These measurements were made on the $C^{\alpha}$ backbone.

The three *Mt* ClpP1 structures described in this paper showed minor differences in their tetradecameric arrangements. Whether these differences are related to the varying crystallization conditions or to crystal-packing forces is not clear. The pH in the three experiments varied from 5.2 to 8.0. 2c8t and 2ce3 crystallized in space group $P2_1$ and 2cby in $P6_122$ (Table 1). The parameters for 2ce3 are almost identical to those of 2c8t, with the largest difference of ~1 Å being in the length of the unit-cell *a* axis. An overall comparison of these three structures shows that flexibility is high in the equatorial region and that shifts of up to 1 Å in the monomer positions are allowed within the heptamers.

In the structures of *Ec* ClpP and *Hs* ClpP, the interface between the heptamers is formed through extensive intercalation of the handle domains from opposite rings. Conversely, the handle domains in the *Mt* ClpP structures are flexible, with no defined electron density for amino acids Gly127–Asp135, thus displaying only a limited number of contact points. This weak interaction, which is driven by the mutual burial of the amino acids Ile136, Ala140 and Phe143, all of which are located in the $\alpha E$ helices of the interacting monomers, is reminiscent of the 'flat' interaction surface described for *Sp* ClpP (Gribun *et al.*, 2005). Another significant characteristic of *Mt* ClpP1 is that there are no inter-heptameric interactions through the head domains. Notably, the $C^{\alpha}$ atoms of the conserved amino acids Asp170 and Arg171 are more than 9 Å apart in *Mt* ClpP1. These residues play a key role in an ion-pair/hydrogen-bond network in *Ec* ClpP, *Hs* ClpP and, although less strongly, in *Sp* ClpP. Indeed, mutation of the residue homologous to Arg171 in *Mt* ClpP1 to a glycine impeded the formation of the double ring of *Sp* ClpP under physiological conditions (Gribun *et al.*, 2005). The central role of this arginine in the structure of the oligomers of *Ec* ClpP and *Hs* ClpP is emphasized by its contribution to the positioning of the $\beta 9$ strand of a neighbouring monomer in the same heptamer through a hydrogen bond to Gln131 (Ser132 in *Mt* ClpP1). The side chain of Arg171 is visible in the *Mt* ClpP structure, but the main chain of the loop where it resides follows a totally different path which precludes all these interactions.

The opening of the axial pore is a funnel-shaped surface (Figs. 3*a* and 3*b*). In *Mt* ClpP1, the outermost rim of this funnel is essentially outlined by the charged and polar amino acids Arg23, Ser26, Glu27, Glu54 and Asp55. Lining the pore, from the outside in we find residues Glu22, Ser19, Asp18, Leu16 and Ser15, with the latter residue defining the edge of the cavity. The proteolytic chamber has a rather heterogeneous composition including hydrophobic, polar and charged residues (Fig. 3*c*). Within the chamber, residues Thr17, Arg43, Glu39 and Glu38 are concentrically distributed in a generally apolar context. Sequence conservation is markedly lower among residues lining the surface of the vault than among those around the catalytic sites (Fig. 3*d*).

The catalytic triads are aligned in an approximately radial distribution next to the equatorial walls of the chamber. The catalytic Ser98, at ∼42 Å, is the closest of the three residues to

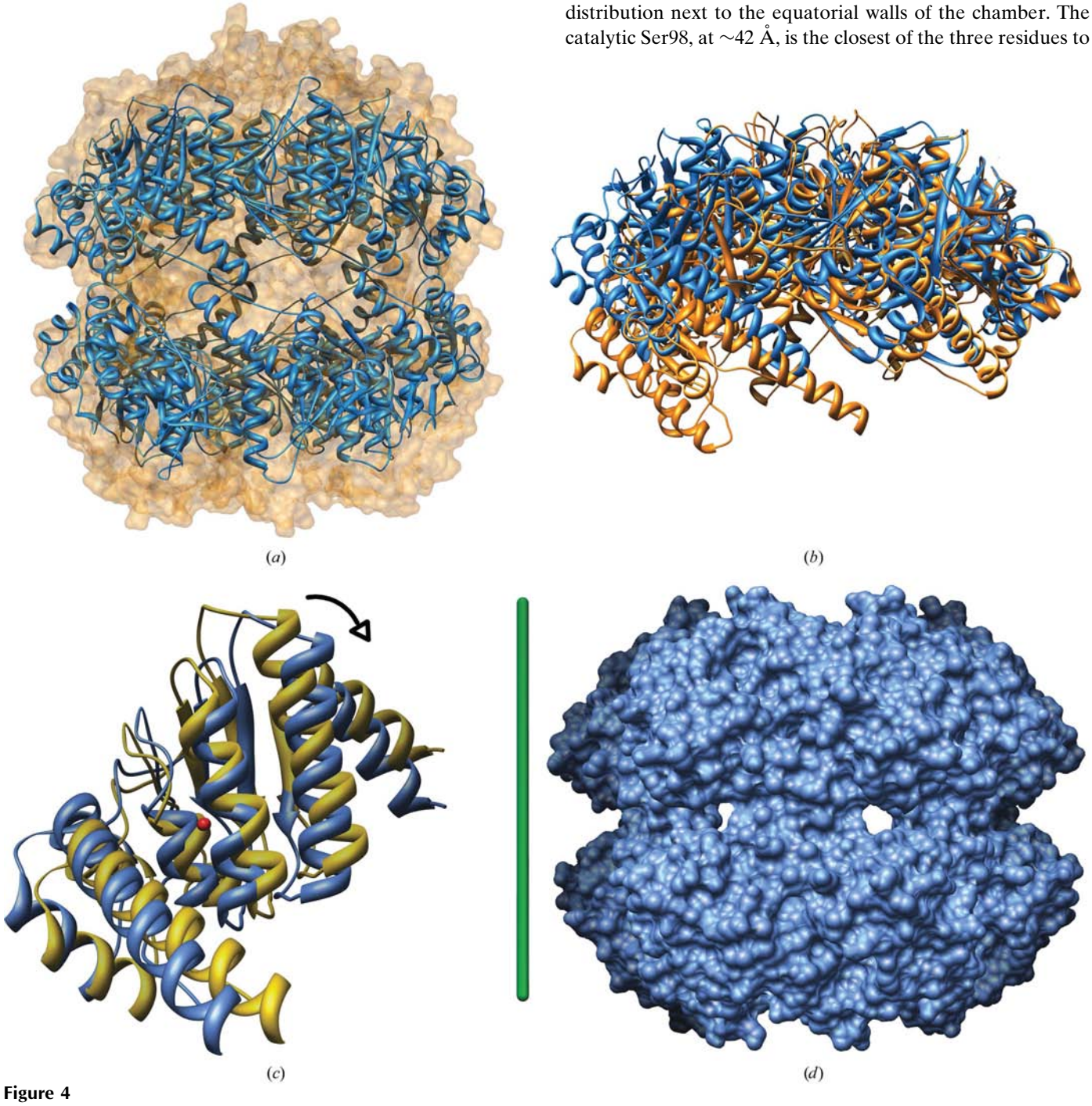


**Figure 4**
Properties of the *Mt* ClpP1 tetradecamer. (*a*) Superimposition of one heptamer from the *Ec* ClpP structure 1yg6 (molecular surface, gold) onto the heptamer of *Mt* ClpP1 structure 2cby (ribbons, blue) shows the greater compactness of the latter. (*b*) Superimposition of monomer *A* of the *Ec* ClpP structure 1yg6 onto monomer *A* of the *Mt* ClpP1 structure 2cby. The superimposed monomers are on the right of the represented heptamers. The colour scheme is as in (*a*). (*c*) Inwards rotation of the monomers of *Mt* ClpP1 shown by overlaying monomer *A* from (*a*) and (*b*). The axis of this rotation (in red) is approximately perpendicular to the pore axis (in green). (*d*) Side view of the front half surface of a complete model of *Mt* ClpP1, *i.e.* including the region Leu126–Asp135, to illustrate the equatorial pores. See text for details.

the central axis and lies in the evolutionarily most conserved region of the central cavity (Fig. 3*d*).

## 4. Discussion

The *Mt* ClpP1 structures described in this study contribute additional information to the notion that the plasticity of inter-ring interactions can be important for the function of ClpP proteins (Gribun *et al.*, 2005; Sprangers *et al.*, 2005; Kang *et al.*, 2005). The most remarkable characteristics of *Mt* ClpP1, compared with previously determined ClpP structures, are the different configuration of the N-termini and the compactness of the tetradecamer.

### 4.1. Flexibility in the handle domains allows a concerted rotation of the monomers and promotes a more compact tetradecamer

Similar to *Hs* ClpP (Kang *et al.*, 2005), *Mt* ClpP1 can exist as a stable heptamer under physiological conditions. Indeed, it was isolated as such throughout the purification steps, as confirmed by analytical ultracentrifugation and size-exclusion chromatography (data not shown). The tetradecamer was only observed in the high ionic strength conditions found in the crystals, suggesting that the interactions keeping these heptamers together are weak and probably mainly driven by the hydrophobic effect. Alternatively, the formation of the tetradecamer may be regulated *in vivo* by the interaction of *Mt* ClpP1 with one of its cognate chaperone ATPases, similarly to what has been shown for *Hs* ClpP (Kang *et al.*, 2005). The resulting tetradecamer is more compact than those found in the structures of *Ec* and *Hs* ClpP, as becomes apparent upon the superimposition of their tetradecamers using one heptameric ring as reference (Fig. 4*a*). In this respect, the structure of *Mt* ClpP1 is similar to that of *Sp* ClpP (A153P) (Gribun *et al.*, 2005). Gribun and coworkers have suggested that such a compact arrangement might represent an early stage of the oligomerization process (Gribun *et al.*, 2005). The resemblance between these two homologous structures extends to the fact that neither the *Sp* mutant ClpP nor the *Mt* ClpP1 presented here showed peptidase activity. Yet, whether there is a causal relationship between these two common features remains unclear.

Whereas the plasticity of the handle–handle interactions ultimately determines the degree of compactness of the tetradecamer, the underlying structural causes are twofold. One reason lies in the different angle subtended by the handle and the head domains at their hinge point, as has been described in the previous section (Fig. 2*b*). The second cause of the higher compactness of the *Mt* ClpP1 tetradecamer is disclosed when the heptameric rings of *Mt* ClpP1 and of *Ec* (or *Hs*) ClpP are observed upon their superimposition using only one of their monomers as reference (Fig. 4*b*). After such an operation, the planes of the heptamers form an angle that actually reflects a concerted rotation of the monomers. For each monomer in *Mt* ClpP1, this rotation can be calculated using as extremes its position in the *Mt* heptamer *versus Ec* heptamer superimposition (Fig. 4*a*) on one hand and that in the *Mt* monomer *versus Ec* monomer superimposition (Fig. 4*b*) on the other. This procedure shows that there is only a minor translation ($0.8 \pm 0.3$ Å) of the monomers, but a rotation of $13.2 \pm 0.5^{\circ}$ (Fig. 4*c*). The axis of such rotation is almost perpendicular ($89.9 \pm 5.9^{\circ}$) to the central axis of the heptamer (Fig. 4*c*). We propose that this concerted rotation of the monomers could play a crucial role in the function of the ClpP proteins. Firstly, it may facilitate the entrance of the substrate by changing the angular tilt of the N-termini of the monomers with respect to the axial pore. Secondly, since it contributes to the opening of equatorial pores next to the active sites, it could assist the exit of the products if the hypothesis of Gribun and coworkers is correct (Gribun *et al.*, 2005; Sprangers *et al.*, 2005). To approximately appreciate the size of such pores, we built a series of models based on the 2ce3 structure, but including the disordered Leu126–Asp135 region. These models were subject to a simulated-annealing refinement protocol. The best among them, in terms of $R$ and $R_{\rm free}$ factors, was further refined in *REFMAC*5 using the same protocols as for the 2cby and 2ce3 structures. Although this model consistently shows no clear supporting density for the disordered region, it contains no clashes and hence could be considered as a qualitative snapshot in the dynamics of this region. This snapshot suggests that equatorial pores (Fig. 4*d*) can appear owing to the plasticity of the inter-ring interactions and are not artefacts arising from the incompleteness of Gribun's or our model.

### 4.2. The conformation of the active sites is consistent with an inactive enzyme

The active sites are located in the clefts between the head and handle domains. Owing to the rotational shift of the monomers and the different orientation of the $\alpha E$ helix, the catalytic triads in *Mt* ClpP1 are more exposed to the environment outside of the tetradecamer. Superimposition of the catalytic triads of *Mt* ClpP1 on the equivalent triads in *Hs* ClpP, *Ec* ClpP and *Sp* ClpP revealed that the position of $C^{\alpha}$ from His123 in *Mt* ClpP1 is shifted away by $\sim$1.7 Å with respect to the position of the histidine in the structures of the catalytically active *Hs* ClpP and *Ec* ClpP proteins (Fig. 5). This dislocation, which abolishes the Ser98–His123 interaction, is probably related to the conformational changes and the concomitant disorganization of the handle domain in *Mt* ClpP1. In this respect, we note that His123 is located two amino acids from the N-terminus of the missing strand $\beta 9$ in the handle domain. Although the hydrogen bond between His123 $N^{\delta 1}$ and Ala172 $O^{\delta 1}$ is maintained, this is clearly an inactive conformation of the active site. This is consistent with our failure to detect any peptidasic activity or to covalently bind the inhibitor PMSF to Ser98.

### 4.3. The N-termini of the monomers build a longer $\alpha A$ helix

The highly conserved N-terminal residues of mature ClpP proteins (Fig. 3*b*) are meant to be important both for the stability of these proteins and for their binding to the cognate

ATPases (Kang *et al.*, 2005). Interestingly, the coded sequence of *Mt* ClpP1 and that of *Sp* ClpP start at the level of the mature sequence of other ClpP proteins, such as the *Ec* and *Hs* homologues. However, whereas the N-terminus of *Sp* ClpP complies with the consensus sequence ΦΦPΦΦΦE (where Φ represents any hydrophobic amino acid; Kang *et al.*, 2005), the N-terminus of *Mt* ClpP1 significantly deviates from the consensus sequence. Therefore, although the full-length *Sp* ClpP may be considered to be a bona fide mature ClpP, this statement cannot be extrapolated to *Mt* ClpP1. Furthermore, the protein sequence of ClpP1 from the taxonomically related actinomycetal bacteria *S. coelicolor* (*Sc*) starts at the same level, but nonetheless seems to require a proteolytic maturation step.

Perhaps as a consequence of this sequence dissimilarity, in the *Mt* ClpP1 structures presented here we do not observe the N-terminal protruding loops found in *Ec* ClpP, *Hs* ClpP and *Sp* ClpP (Kang *et al.*, 2004; Gribun *et al.*, 2005; Bewley *et al.*, 2006; Szyk & Maurizi, 2006). Instead, these residues fit in the $\alpha A$ helix. Nevertheless, it must be noted that the 14 most terminal residues of *Mt* ClpP1 are present, as confirmed by mass-spectrometry analysis, although they are not visible in the electron-density maps. Thus, we cannot rule out any possible orientation of the N-termini, including the formation of disordered protruding loops. Another intriguing possibility is that the disordered nonprocessed N-termini may preclude the entry of even small peptides, thus explaining the lack of peptidasic activity in our preparations. Moreover, it is worth noting that in *S. coelicolor* the presence of ClpP2 is needed to obtain a processed ClpP1 and *vice versa* (Viala & Mazodier, 2002). This fact has been taken as evidence for an obligate cross-processing step involving independently folded ClpP1 and ClpP2 homo-oligomers. However, an alternative hypothesis was also suggested according to which the two proteins would form a hetero-tetradecamer in order to become active (Viala & Mazodier, 2002). The *Mt clpP1* gene forms part of a genomic region that is indeed similar to that containing *Sc clpP1*, notably including the genes *clpP2* and *clpX*. In this context, it would not be surprising that *Mt* ClpP2 is required for the maturation of *Mt* ClpP1.

In conclusion, the structure of *Mt* ClpP1 shows an alternative arrangement of the tetradecamer that may correspond to a different intermediate in the mechanism of action of the ClpP caseinolytic proteases. Specifically, we suggest that the structures presented here may be used to shed light on the processes of substrate uptake and product release. However, our hypothesis must be tempered by the fact that disordered handle domains have so far been observed only for structures corresponding to inactive ClpP proteins, at least until more is known about *Pf* ClpP. Finding conditions under which *Mt* ClpP1 will be active would clearly help in elucidating this problem.

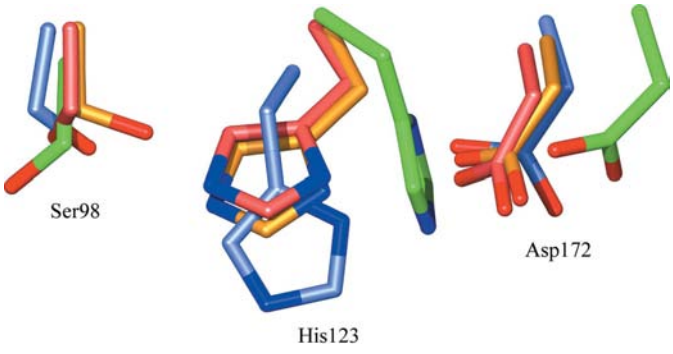


**Figure 5**
A superimposition of the catalytic triads in ClpP proteins corresponding to the active sites in *Mt* (Ser98-His123-Asp172, blue), *Hs* (Ser97-His122-Asp171, red), *Ec* (Ser97-His122-Asp171, gold) and *Sp* (Ser96-His121-Asp172, green). Labels in the figure refer to *Mt* ClpP1. Only one of the two rotamers adopted by His123 in *Mt* ClpP1 is shown.

The authors thank Patrik Johansson, Lena Henriksson, Annette Roos, Wojciech Krajewski and Eva-Lena Andersson for interesting discussions and help with various things during this project and the staff of the ID14-3 beamline from the European Synchrotron Radiation Facility (ESRF, Grenoble, France) for their support. We are indebted to Timothy Lekin, Brent Segelke and Chang Yub Kim from the TB Structural Genomics Consortium (TBSGC) pipeline for providing us with the data from which we built the 2ce3 structure. Financial support was received from the Swedish Research Council (VR), the Foundation for Strategic Research (SSF) and X-TB (QLRT-2000-02018)

## References

Adam, Z. & Clarke, A. K. (2002). *Trends Plant Sci.* **7**, 451–456.
Aizenman, E., Engelberg-Kulka, H. & Glaser, G. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 6059–6063.
Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. (2003). *Nucleic Acids Res.* **31**, 23–27.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bewley, M. C., Graziano, V., Griffin, K. & Flanagan, J. M. (2006). *J. Struct. Biol.* **153**, 113–128.
Bross, P., Andresen, B. S., Knudsen, I., Kruse, T. A. & Gregersen, N. (1995). *FEBS Lett.* **377**, 249–252.
Clarke, A. K., Gustafsson, P. & Lidholm, J. A. (1994). *Plant Mol. Biol.* **26**, 851–862.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Corydon, T. J., Bross, P., Holst, H. U., Neve, S., Kristiansen, K., Gregersen, N. & Bolund, L. (1998). *Biochem. J.* **331**, 309–316.
Ducruix, A. & Giegé, R (1999). *Crystallization of Nucleic Acids and Proteins: A Practical Approach.* Oxford University Press.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Engh, R. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.
Frees, D. & Ingmer, H. (1999). *Mol. Microbiol.* **31**, 79–87.
Gaillot, O., Pellegrini, E., Bregenholt, S., Nair, S. & Berche, P. (2000). *Mol. Microbiol.* **35**, 1286–1294.
Gerth, U., Kruger, E., Derre, I., Msadek, T. & Hecker, M. (1998). *Mol. Microbiol.* **28**, 787–802.
Gottesman, S., Wickner, S. & Maurizi, M. R. (1997). *Genes Dev.* **11**, 815–823.

Gribun, A., Kimber, M. S., Ching, R., Sprangers, R., Fiebig, K. M. & Houry, W. A. (2005). *J. Biol. Chem.* **280**, 16185–16196.
Grimaud, R., Kessel, M., Beuron, F., Steven, A. C. & Maurizi, M. R. (1998). *J. Biol. Chem.* **273**, 12476–12481.
Harris, M. & Jones, T. A. (2001). *Acta Cryst.* D**57**, 1201–1203.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kang, S. G., Dimitrova, M. N., Ortega, J., Ginsburg, A. & Maurizi, M. R. (2005). *J. Biol. Chem.* **280**, 35424–35432.
Kang, S. G., Maurizi, M. R., Thompson, M., Mueser, T. & Ahvazi, B. (2004). *J. Struct. Biol.* **148**, 338–352.
Kessel, M., Maurizi, M. R., Kim, B., Kocsis, E., Trus, B. L., Singh, S. K. & Steven, A. C. (1995). *J. Mol. Biol.* **250**, 587–594.
Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1395–1400.
Kleywegt, G. J. & Jones, T. A. (1999). *Acta Cryst.* D**55**, 941–944.
Krissinel, E. & Henrick, K. (2005). *CompLife 2005*, edited by M. R. Berthold, pp. 163–174. Berlin: Springer-Verlag.
Kruger, E., Witt, E., Ohlmeier, S., Hanschke, R. & Hecker, M. (2000). *J. Bacteriol.* **182**, 3259–3265.
Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005). *Nucleic Acids Res.* **33**, W299–W302.
Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.
Maurizi, M. R. (1992). *Experientia*, **48**, 178–201.
Maurizi, M. R., Clark, W. P., Kim, S.-H. & Gottesman, S. (1990). *J. Biol. Chem.* **265**, 12546–12552.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Ortega, J., Lee, H. S., Maurizi, M. R. & Steven, A. C. (2002). *EMBO J.* **21**, 4938–4949.
Ortega, J., Singh, S. K., Ishikawa, T., Maurizi, M. R. & Steven, A. C. (2000). *Mol. Cell*, **6**, 1515–1521.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.
Schirmer, E. C., Glover, J. R., Singer, M. A. & Lindquist, S. (1996). *Trends Biochem. Sci.* **21**, 289–296.
Shanklin, J., DeWitt, N. D. & Flanagan, J. M. (1995). *Plant Cell*, **7**, 1713–1722.
Singh, S. K., Guo, F. & Maurizi, M. R. (1999). *Biochemistry*, **38**, 14906–14915.
Sprangers, R., Gribun, A., Hwang, P. M., Houry, W. A. & Kay, L. E. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 16678–16683.
Stephani, K., Weichart, D. & Hengge, R. (2003). *Mol. Microbiol.* **49**, 1605–1614.
Szyk, A. & Maurizi, M. R. (2006). *J. Struct. Biol.* **156**, 165–174.
Thompson, M. W. & Maurizi, M. R. (1994). *J. Biol. Chem.* **269**, 18201–18208.
Thompson, M. W., Singh, S. K. & Maurizi, M. R. (1994). *J. Biol. Chem.* **269**, 18209–18215.
Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P. & Bukau, B. (2001). *Mol. Microbiol.* **40**, 397–413.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Viala, J. & Mazodier, P. (2002). *Mol. Microbiol.* **44**, 633–643.
Wang, J., Hartling, J. A. & Flanagan, J. M. (1997). *Cell*, **91**, 447–456.
Webb, C., Moreno, M., Wilmes-Riesenberg, M., Curtiss, R. III & Foster, J. W. (1999). *Mol. Microbiol.* **34**, 112–123.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.