

# Structure of the Archaeal Pab87 Peptidase Reveals a Novel Self-Compartmentalizing Protease Family

Vanessa Delfosse<sup>1</sup>, Eric Girard<sup>2</sup>, Catherine Birck<sup>3</sup>, Michaël Delmarcelle<sup>4</sup>, Marc Delarue<sup>5</sup>, Olivier Poch<sup>3</sup>, Patrick Schultz<sup>3</sup>, Claudine Mayer<sup>1,5\*</sup>

**1** Centre de Recherche des Cordeliers, LRMA, INSERM UMR-S 872, Université Pierre et Marie Curie, Paris, France, **2** Institut de Biologie Structurale UMR5075 CEA-CNRS-UJF, Grenoble, France, **3** Département de Biologie et Génomique Structurales, IGBMC, Illkirch, France, **4** Centre d'Ingénierie des Protéines, Université de Liège, Institut de Chimie B6, Sart-Tilman, Belgium, **5** Département de Biologie Structurale et Chimie, Institut Pasteur, CNRS URA 2185, Paris, France

## Abstract

Self-compartmentalizing proteases orchestrate protein turnover through an original architecture characterized by a central catalytic chamber. Here we report the first structure of an archaeal member of a new self-compartmentalizing protease family forming a cubic-shaped octamer with  $D_4$  symmetry and referred to as CubicO. We solved the structure of the *Pyrococcus abyssi* Pab87 protein at 2.2 Å resolution using the anomalous signal of the high-phasing-power lanthanide derivative Lu-HPDO3A. A 20 Å wide channel runs through this supramolecular assembly of 0.4 MDa, giving access to a 60 Å wide central chamber holding the eight active sites. Surprisingly, activity assays revealed that Pab87 degrades specifically D-amino acid containing peptides, which have never been observed in archaea. Genomic context of the Pab87 gene showed that it is surrounded by genes involved in the amino acid/peptide transport or metabolism. We propose that CubicO proteases are involved in the processing of D-peptides from environmental origins.

**Citation:** Delfosse V, Girard E, Birck C, Delmarcelle M, Delarue M, et al. (2009) Structure of the Archaeal Pab87 Peptidase Reveals a Novel Self-Compartmentalizing Protease Family. PLoS ONE 4(3): e4712. doi:10.1371/journal.pone.0004712

**Editor:** Bostjan Kobe, University of Queensland, Australia

**Received:** January 3, 2009; **Accepted:** January 28, 2009; **Published:** March 5, 2009

**Copyright:** © 2009 Delfosse et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Vanessa Delfosse was supported by a scholarship from the French Ministry of Research and higher education. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: mayer@pasteur.fr

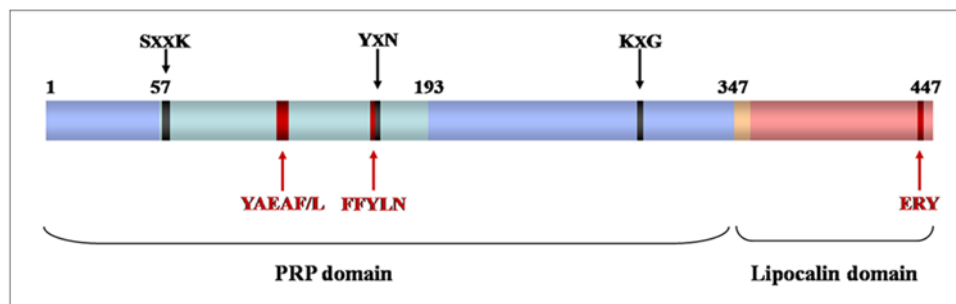
## Introduction

Intracellular protein breakdown is a universal process that implicates protein degradation carried out by high-molecular-weight self-compartmentalizing proteases [1,2]. They belong to different MEROPS families [3] but have converged towards the same barrel-shaped architecture. Compartmentalization confines the peptidase activity to inner cavities only accessible to unfolded polypeptides. Initial proteolysis is fulfilled by energy-dependent proteasomes in eukaryotes and archaea [4] or by the equivalent bacterial counterparts, HslVU [5], ClpAP [6], or Lon [7], which generate peptides of about 10 amino acids long. In addition to the energy-dependent proteasome systems, energy-independent protease complexes have been described to further process and degrade the peptides produced by proteasomes. Organisms diverged in terms of peptide hydrolysis processes and have their own specific pool of different energy-independent proteases that functionally complements the proteasome activity [8]. Structural studies of such complexes, that putatively take part in the degradation of small oligopeptides, reveal several types of organization. The 720 kDa Tricorn protease from *Thermoplasma acidophilum* contains six subunits associated into rings with  $D_3$  symmetry and possesses trypsin- and chymotrypsin-like activities [9]. A larger assembly has been observed *in vivo* for Tricorn where 20 hexamers form a 14.6 MDa icosahedral capsid, whose physiological role is unknown [10]. The TET aminopeptidases are present in organisms where Tricorn is not found [11]. The crystal structures of archaeal *Pyrococcus horikoshii* PhTET1 and PhTET2 have revealed that these proteases adopt a

tetrahedral shape containing twelve subunits [12,13]. Like for Tricorn, TET can form a supramolecular assembly of 800 kDa composed of 24 subunits and referred to as octahedron, only observed *in vivo* yet [14]. TppII is the only energy-independent self-compartmentalizing protease found exclusively in eukaryotes and consists of a particle of more than 1 MDa displaying both exo- and endo-proteolytic activities. TppII adopts a toroidal shape of 28 Å large and 60 Å long revealed by electron microscopy data [15] and is considered as the eukaryotic counterpart of Tricorn. Finally, the DppA D-aminopeptidase from *Bacillus subtilis* was the only example of D-stereospecific self-compartmentalizing protease described so far [16]. DppA is a homodecameric particule with  $D_5$  symmetry only found in bacteria and archaea. Here, we report the structure of a novel self-compartmentalizing protease, the Pab87 protein from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. Pab87 is a serine protease belonging to the MEROPS S12 family and displays sequence similarity with penicillin-recognizing proteins, termed PRP (Figure 1). This proteolytic complex of 0.4 MDa adopts a  $D_4$  barrel-shaped oligomeric architecture with a central channel and an internal cavity. Furthermore, protein Pab87 is, like DppA, D-stereospecific as shown by activity assays and is proposed to play a role in the processing of environmental D-amino acid containing peptides.

## Results and Discussion

The Pab87 structure was determined at 2.2 Å resolution using the single-wavelength anomalous dispersion method with anom-



**Figure 1. Primary structure of the CubicO proteases.** Schematic representation of the CubicO protease sequences. The PRP (serine protease)  $\alpha$ / $\beta$  and all-helical regions are in light blue and cyan, respectively, the linking helix in orange and the lipocalin domain in salmon. Numbering refers to the Pab87 sequence. The three PRP conserved motifs are indicated in black, the three CubicO specific motifs in red. doi:10.1371/journal.pone.0004712.g001

alous signal of a high-phasing-power lanthanide derivative, the Lu-HPDO3A complex (Table 1) [17]. Pab87 is the first characterized member of a new self-compartmentalizing protease family, referred to as CubicO proteases as they display a cubic-shaped octameric form (Figure 2). The molecular mass of the macromolecular assembly was confirmed using analytical ultracentrifugation ( $391 \pm 10$  kDa). The crystal structure of the *P. abyssi* CubicO protease (Pab87) reveals a 422 symmetry, where the eight monomers are arranged in a barrel-shaped architecture. In the protease complex, monomers associate through a non-crystallographic four-fold axis to form two stacked tetramers that are themselves related by a non-crystallographic two-fold axis perpendicular to the latter. The C-terminal domains interdigitate and form a central ring of one third of the total height of the octamer, playing a central role in this oligomerization (Figure 2B). Moreover, cryo-electron microscopy data show that Pab87 adopts the same octameric structure in solution (Figure 2D).

### The CubicO protease family

CubicO protease sequences, found in six deep-sea archaea and one marine  $\gamma$ -proteobacterium, are constituted by two sequence domains, an N-terminal serine protease domain and a C-terminal domain without any detectable sequence similarity (Figure 1). The protease domain displays sequence similarity to PRP, notably the conserved catalytic motifs -SxxK-, -YxN- and -HxG-, and belongs to S12 protease family. The CubicO sequences can be characterized by three additional specific motifs, -YAEAF/L- and -FFYLN- located in the N-terminal domain, the latter extending the second PRP catalytic motif, and finally -ERY- found at the extreme C-terminus (Figure 1 and Figure S1).

With regard to the known energy-independent proteases, all CubicO sequences co-exist with at least one TET sequence. In the *P. abyssi* genome, four TET paralogs have been found. While TET aminopeptidases are present in organisms where Tricorn is not found, Tricorn is generally absent within the organisms that possess CubicO proteases. Concerning the D-aminopeptidase DppA, four of these seven organisms possess a DppA sequence. These observations highlight an intricate and complex distribution of either L or D-stereospecific self-compartmentalizing proteases among living organisms [11].

### Pab87 exopeptidase activities

As Pab87 is a S12 protease, the amino and carboxypeptidase activities were investigated. The aminopeptidase activity was tested using L- and D-Ala-*p*-nitroanilide compounds as substrates. Preliminary studies underlined the high  $K_m$  values for these substrates. Catalytic efficiency  $k_{cat}/K_m$  could be determined as a

function of temperature (Figure 3B). Clearly, the enzyme shows a preference for the substrate with a D configuration. At 90°C, the  $k_{cat}/K_m$  is  $3.92 \text{ M}^{-1} \text{ s}^{-1}$  and  $0.06 \text{ M}^{-1} \text{ s}^{-1}$  for respectively the D- and L-Ala-*p*-nitroanilide. The  $k_{cat}/K_m$  values increase drastically after 70°C. The maximum of activity was obtained at 90°C, the

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

Lu-HPDO3A derivative	
<b>Data collection</b>	
X-ray source	ESRF ID29
Wavelength (Å)	1.3403
Temperature (K)	100
Space group	<i>P</i> 1
Unit-cell parameters (Å, °)	a = 100.94, b = 105.63, c = 112.92, $\alpha$ = 72.17, $\beta$ = 66.51, $\gamma$ = 81.39
Resolution range (Å) <sup>a</sup>	30.0 - 2.2 (2.32 - 2.20)
No. of unique reflections	198,616
Multiplicity <sup>a</sup>	7.7 (7.8)
Completeness (%) <sup>a</sup>	95.7 (93.8)
$R_{sym}$ <sup>a, b</sup>	0.080 (0.289)
$\langle I/\sigma(I) \rangle$ <sup>a</sup>	17.4 (7.2)
<b>Refinement</b>	
Resolution range (Å)	30.0-2.2
$R_{work}/R_{free}$ <sup>c</sup>	19.6/24.1
Number of atoms	30,885
Protein	28,405
Lu-HPDO3A	112
Water	2,368
Average B factor	24.4
R.m.s <sup>d</sup> deviations from ideal values	
Bond lengths (Å)	0.007
Bond angles (°)	1.3

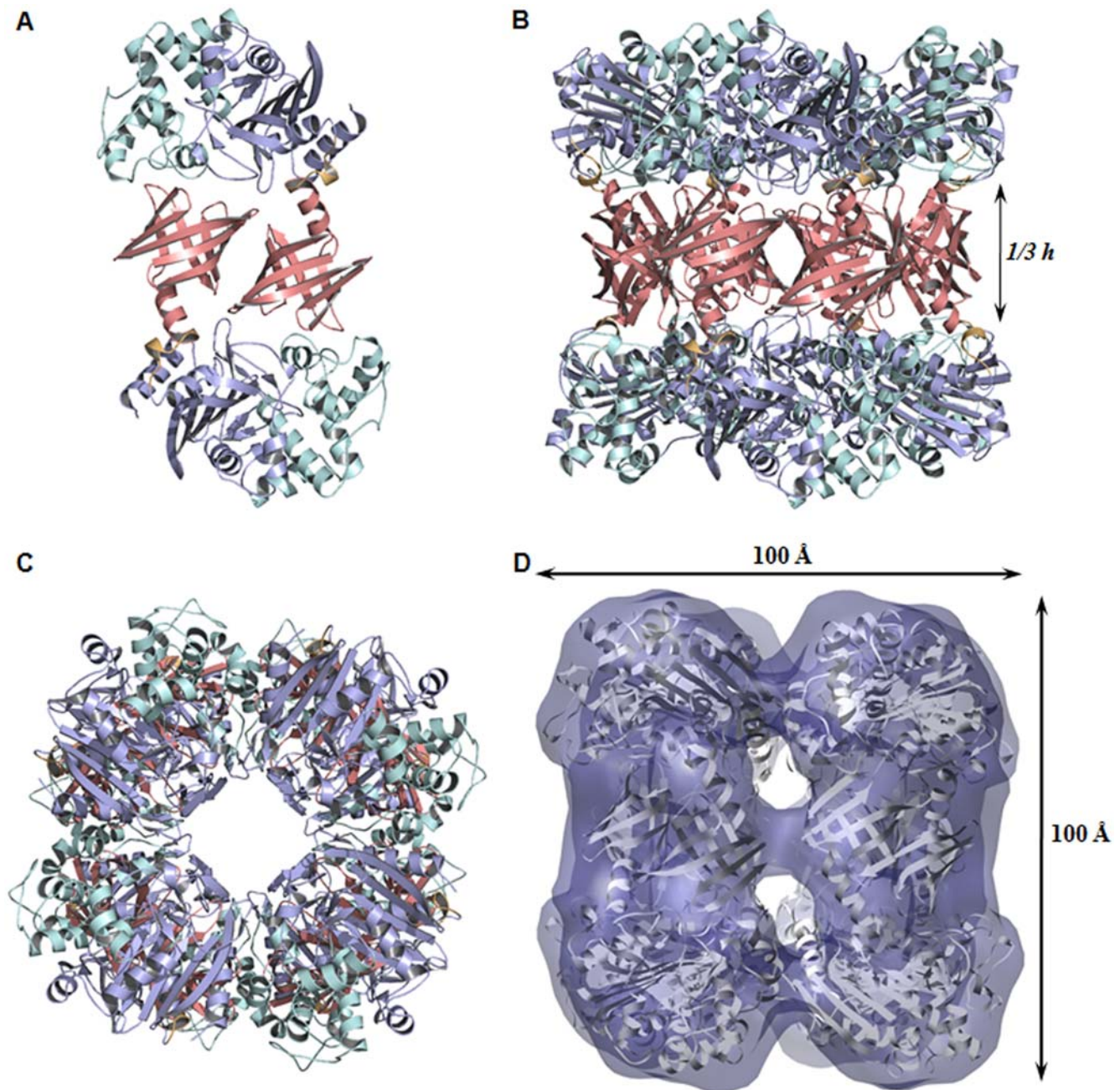
<sup>a</sup>Values in parentheses correspond to the highest-resolution shell.

<sup>b</sup> $R_{sym} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$ , where  $I_{h,i}$  is the *i*th observation of reflection **h** and  $\langle I_h \rangle$  is the weighted average intensity for all observations *i* of reflection **h**.

<sup>c</sup> $R_{cryst} = \sum |F_o - F_c| / \sum |F_o|$ , where  $F_o$  and  $F_c$  are observed and calculated amplitudes, respectively.  $R_{free}$  is calculated similarly using a test set of reflections. The test set (5% of reflections) is omitted in the refinement.

<sup>d</sup>Root mean square.

doi:10.1371/journal.pone.0004712.t001



**Figure 2. The CubicO particle.** Structural regions are color coded as in Figure 1. **A**, Ribbon representation of the head-to-tail dimer structure forming the edges of the cubic-shaped octamer. **B**, side view along and **C**, top view perpendicular to the four-fold axis of the octamer. The central ring formed by the C-terminal domains represents one third of the total height. **D**, Fitting of cryo-EM map shown in transparent solid surface with the X-ray structure.

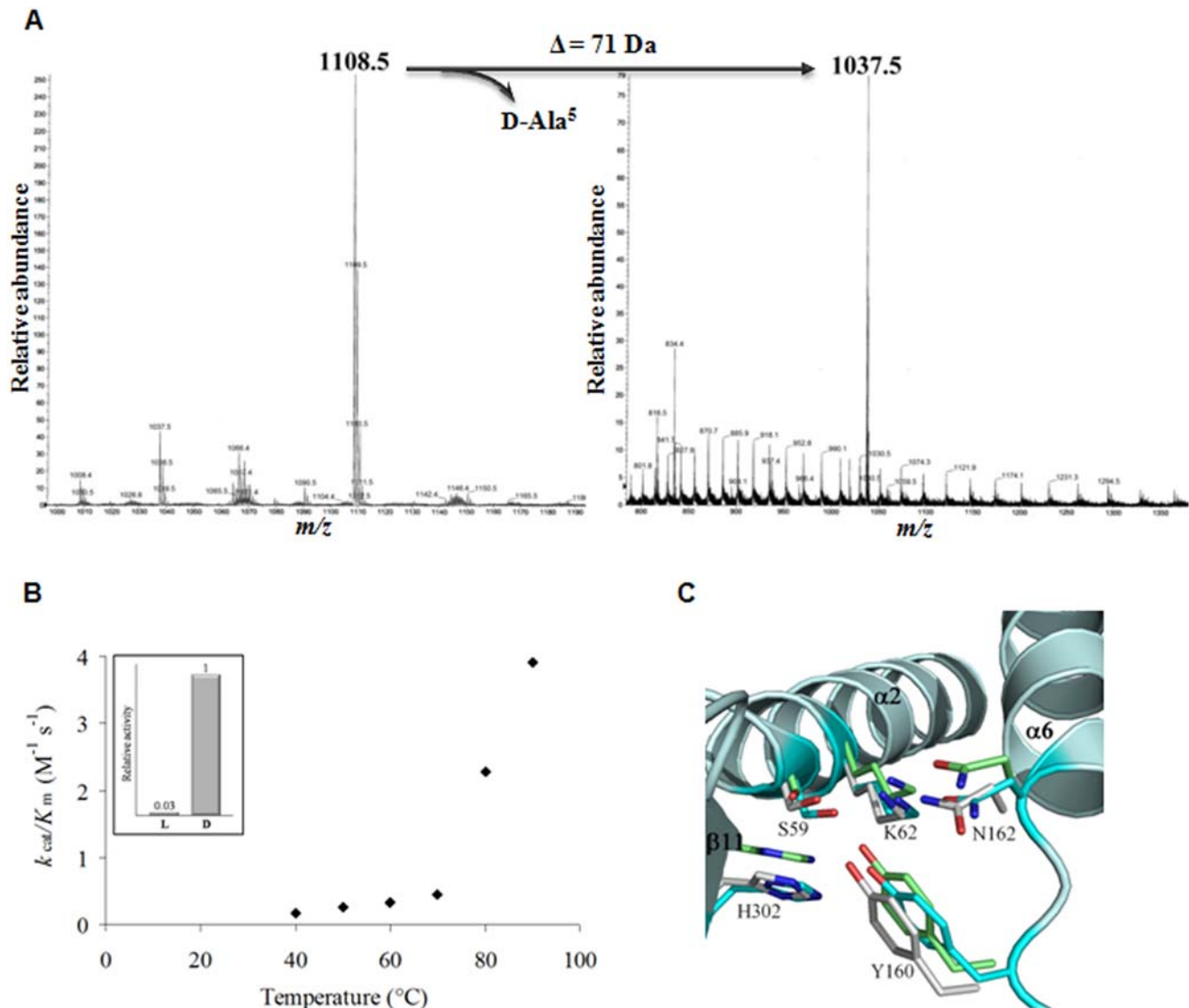
doi:10.1371/journal.pone.0004712.g002

maximum technically reachable temperature, underlying the hyperthermophilic properties of the enzyme. Moreover, DDCarboxypeptidase activity was observed on various mucopeptides. Mass spectrometry analyzes clearly show that Pab87 is able to hydrolyse the C-terminal D-alanine of pentapeptidic precursors with or without the sugar moiety from both Gram-negative and Gram-positive bacteria, at 37 and 90°C (Figure 3A). The D-stereospecificity observed for Pab87 is consistent with the activities of the majority of S12 proteases. Catalytic residues of *P. abyssi* CubicO protease perfectly superimpose to those of other S12

members, indicating similar catalytic pathways despite different substrate specificities (Figure 3C). In conclusion, Pab87 is a hyperthermophilic peptidase that acts on D-amino acid containing peptides at N- and C-termini.

#### A novel association of two structural domains

The 50.4 kDa monomer of Pab87 consists of two structural domains, an N-terminal PRP domain associated to a C-terminal lipocalin domain that plays a crucial role in the octamerization and in the active site compartmentalization (Figure 2A). The N-

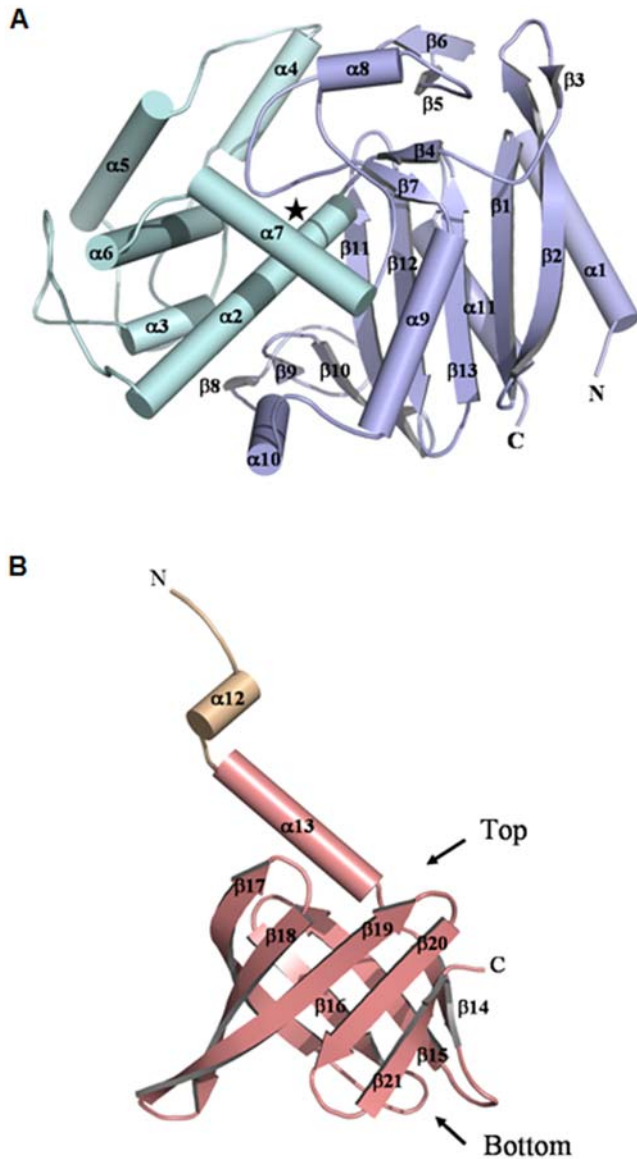


**Figure 3. Exopeptidase activities of the CubicO proteases. A**, MS analysis of the products of the carboxypeptidation reaction catalyzed by Pab87. Peaks at  $m/z$  1108.5 and 1037.5 correspond to the GM-penta-(Ala-Ala) substrate and to the GM-tetra-(Ala-Ala) product respectively. **B**, Effect of temperature on the Pab87 catalytic efficiency on D-Ala-*p*-nitroanilide. Box indicates the relative aminopeptidase activity at 90°C of Pab87 on L- and D-*p*-NA. **C**, Close-up view of the superimposed active sites of Pab87 (in cyan), *O. anthropi* DAP D-aminopeptidase (in white; PDB id 1E15) and of *Streptomyces* R61 DD-carboxypeptidase (in green; PDB id 3PTE). Only side chains of catalytic residues of the three structures with backbone of Pab87 are represented. Secondary structure elements are indicated in bold. doi:10.1371/journal.pone.0004712.g003

terminal domain structure follows the pattern of the known PRP structures, composed of two regions, one  $\alpha/\beta$  and one all-helical region with six helices. The  $\alpha/\beta$  region is discontinuous in sequence and consists of a main eight-stranded antiparallel  $\beta$ -sheet flanked by three helices. The C-terminal domain is connected to the N-terminal domain by a small helix, folded back onto the last helix of the N-terminal domain. The Pab87 C-terminal domain presents a typical lipocalin superfold [18] that consists of an  $\alpha$ -helix and an eight-stranded antiparallel  $\beta$ -barrel (Figure 4). This N-terminal  $\alpha$ -helix closes off the top of the barrel, in vicinity of the PRP domain. Like for all lipocalins, the interior of the barrel is coated with hydrophobic residues. The tunnel observed in the lipocalins is replaced by a small recess at the bottom of the calyx. As lipocalins were first discovered in eukaryotes, and more recently in bacteria [19], the Pab87 C-terminal domain provides the first example of an archaeal lipocalin.

### Structure of the macromolecular assembly

The 400 kDa octamer looks like a 100 Å side cube with a central hole along the four-fold axis and small entrances distributed around the equatorial plane of the octamer. The octamer can be considered as four head-to-tail dimers organized around the four-fold axis, forming the edges of the cube (Figure 2). Each monomer interacts with four other subunits of the octamer but main interactions are observed within the head-to-tail dimers, where monomers interact primarily with each other by shape recognition through key-lock type interaction (Figure 5A, B). Intersubunit contacts are formed by electrostatic and hydrogen-bonding interactions between the bottom of the lipocalin domain of the first monomer and the all-helical region of the PRP domain of the second monomer. Three bottom loops of the calyx embrace two helices of the N-terminal domain at the inside and outside face of the octamer. Interestingly, one of these helix-loop interactions



**Figure 4. Secondary structure assignment** for **A**, the N- and **B**, the C-terminal domains of the *P. abyssi* Pab87 monomer, color coded according to Figure 1 (the PRP  $\alpha/\beta$  and all-helical regions are in light blue and cyan, respectively, the linking helix in light orange and the lipocalin domain in salmon). The active site is located by a star in (A). doi:10.1371/journal.pone.0004712.g004

links spatially the first and the third CubicO conserved motifs through hydrogen bonding (Figure 5C, D).

Whereas the twelve subunits of the archaeal TET aminopeptidases are organized as a tetrahedron [12,13], the drosophila TppII protease adopts a toroidal shape [15], the six subunits of the Tricorn protease from *T. acidophilum* associate into rings with  $D_3$  symmetry [9], and the ten monomers of DppA D-aminopeptidase from *B. subtilis* oligomerize in a  $D_5$  symmetry [16], the structure of the *P. abyssi* CubicO protease reveals a novel giant protease architecture, a  $D_4$  barrel-shaped oligomer with a central channel and an internal cavity.

### The proteolytic central chamber

The 20 Å wide square-shaped channel, running along the four-fold axis through the entire complex, joins the equatorial two-fold

axes into a 60 Å wide central cavern formed by association of the eight C-terminal domains (Figure 6A). Each active site, facing the central cavern, lies between three domains, the PRP and the lipocalin domains of one monomer, and the second lipocalin domain of a head-to-tail dimer (Figure 6C). The inside volume of the cavity is about 80,000 Å<sup>3</sup>, large enough to accommodate small unfolded peptides (Figure 6), and reminds those of other self-compartmentalizing proteases known to bury their active sites in an inner cavity. In the *P. abyssi* CubicO protease, the proteolytic chamber is separated from the access channels by two crown-shaped surfaces at about 10 Å of each side of the equatorial plane. These crowns are formed by two loops of opposite monomers joining at the same height (Figure 6A). The shape and size of the cavity are conserved in all CubicO proteases as shown by the structure modelling of the six other members of the family (Figure S2). Variability mainly concerns the access to the active site through the large and flexible loops forming the inside crowns appearing to function as substrate sieves (Figure 7).

### Conclusion

The presence of such proteases with D-stereospecificity in archaea raises the question of their physiological function as these organisms generally lack D-amino acid containing peptides. All organisms that possess CubicO proteases are found near deep-sea hydrothermal vents. In these biotopes, archaea and bacteria live in intimate and complex symbiosis [20]. Growing bacteria continuously release mucopeptide fragments of their cell wall into the extracellular milieu that could serve as nutrients for other organisms living in the same habitat. As it has been shown for *Pyrococcus* species, heterotrophic organisms are able to utilize amino acids as sole carbon and nitrogen sources [21]. Analyses of the genomic context of the Pab87 gene revealed that it is surrounded by genes involved in the amino acid/peptide transport or metabolism, e.g. a PLP-dependent aminotransferase (PAB0086), a D-aminoacylase (PAB0090) and the dipeptide ABC transporter (PAB0091-95). Taken all together, we propose that CubicO proteases play a central role in proteolysis of these bacterial cell debris that contains D-amino acids.

### Materials and Methods

#### Multiple sequence alignment

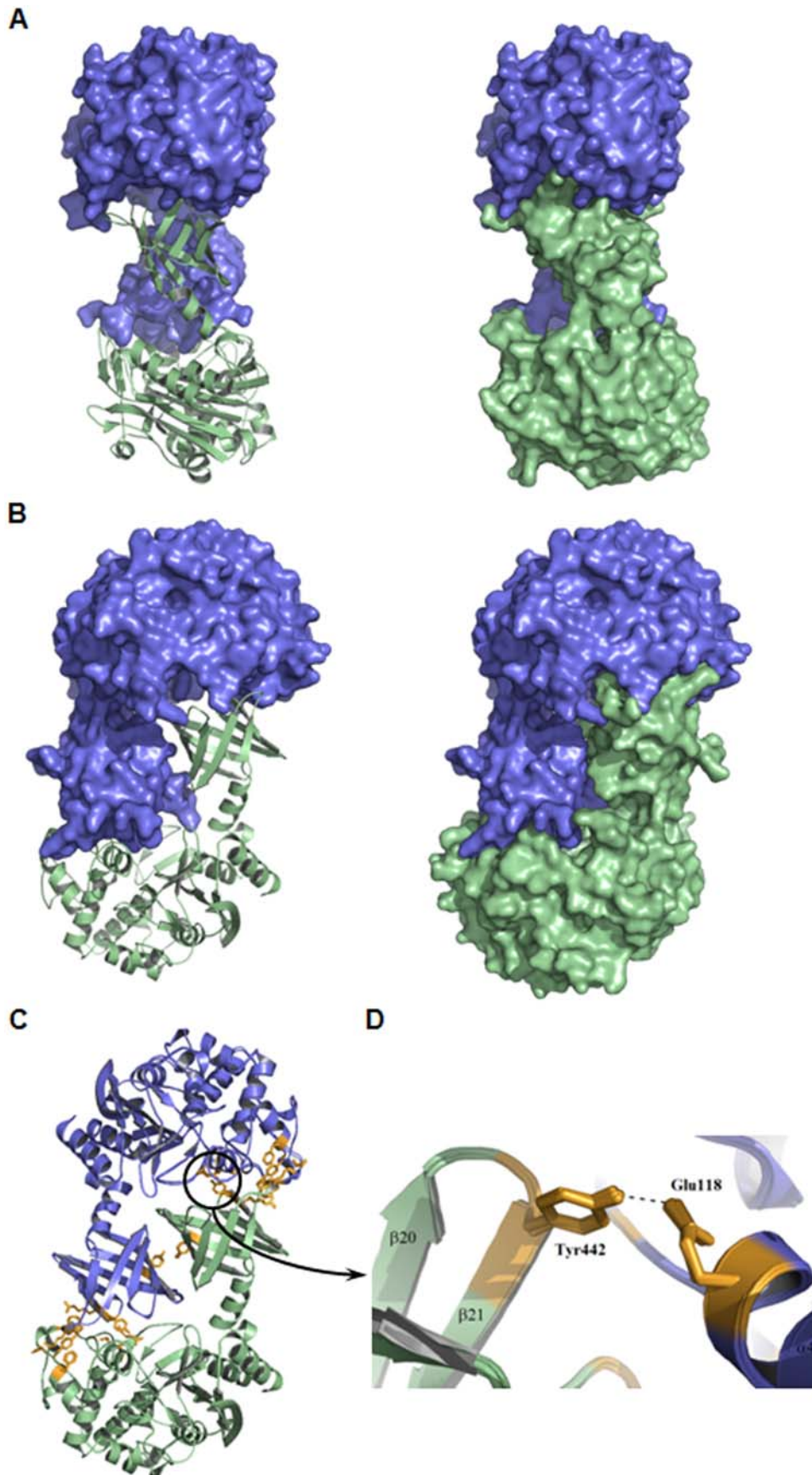
Pab87 CubicO homologs were detected using iterative searches of the non-redundant database of protein sequences at the NCBI with the PSI-BLAST program and the C-terminal domain of Pab87 as the query. The initial multiple alignment of complete sequences was performed using PipeAlign [22] and manually adjusted to maximize the sequence conservation.

#### Protein production and purification

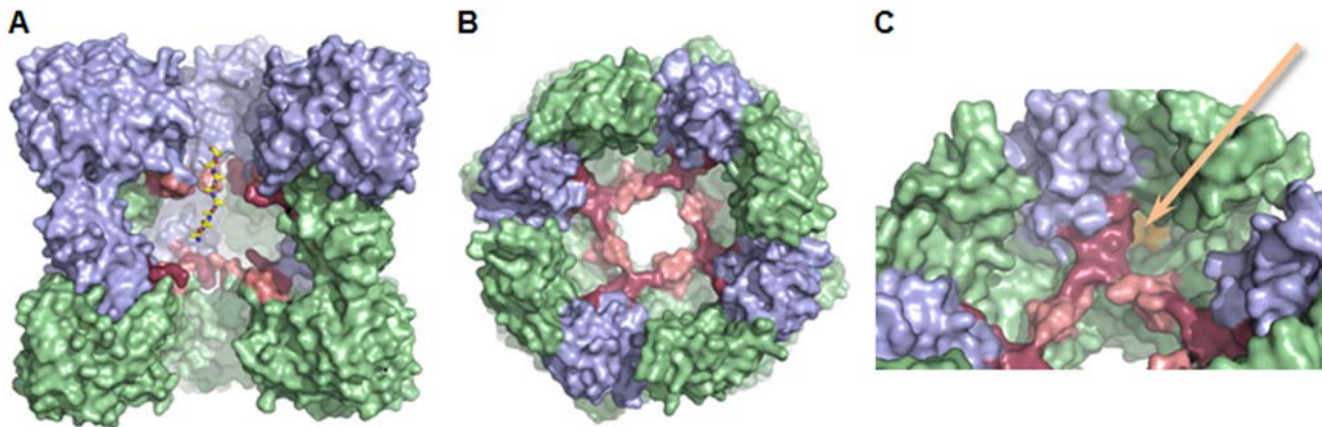
Production and purification of Pab87 were carried out as previously described [23]. Briefly, Pab87 was purified by heating at 90°C, and using anion exchange and size exclusion chromatographies. The protein was concentrated in 50 mM Tris-HCl buffer pH 8.5, 350 mM NaCl.

#### Analytical ultracentrifugation

Sedimentation velocity experiments were performed in a Beckman XL-A analytical ultracentrifuge using a double sector charcoal-Epon cell at 20°C and 40,000 rpm. Absorbance scans were taken at 280 nm every 1 min. The protein concentration was 10 μM in 150 mM Tris (pH 8.5), 350 mM NaCl. The program Sednterp [24] was used to calculate solvent density (1.01689 g/cm), solvent viscosity (0.01083 Poise) and partial specific volume



**Figure 5. Head-to-tail dimer key-lock interaction.** **A**, Side view of the head-to-tail dimer with the first monomer represented in blue surface and the second in green cartoon (left) and with both monomers represented in surface (right) highlighting the interdigitating of the monomers. **B**, Same figure as in (A) with a view rotated by 90° to the right. **C**, Residues involved in short distance interactions between the two monomers are highlighted in orange. **D**, Hydrogen bond between the two CubicO specific motifs -YAE<sup>118</sup>AF/L- and -ERY<sup>442</sup>-. doi:10.1371/journal.pone.0004712.g005



**Figure 6. The proteolytic chamber.** **A**, Open side view of the octamer. The front head-to-tail dimer has been removed. The top and bottom tetrameric rings are colored in blue and green, respectively. The  $\beta 5$ – $\beta 6$  and  $\beta 16$ – $\beta 17$  loops forming the two crown-shaped surfaces are colored in salmon and raspberry, respectively. To illustrate the size of the cavity, an Ala<sub>8</sub>-peptide was manually placed in extended conformation. **B**, Open top view of the octamer down the four-fold axis. For better clarity, the four N-terminal PRP domains of the top tetramer and the four C-terminal domain  $\beta 16$ – $\beta 17$  loops of the bottom tetramer have been removed. **C**, Same figure as in **(B)** tilted so that the active site entrance is visible. Catalytic pocket is highlighted in orange.

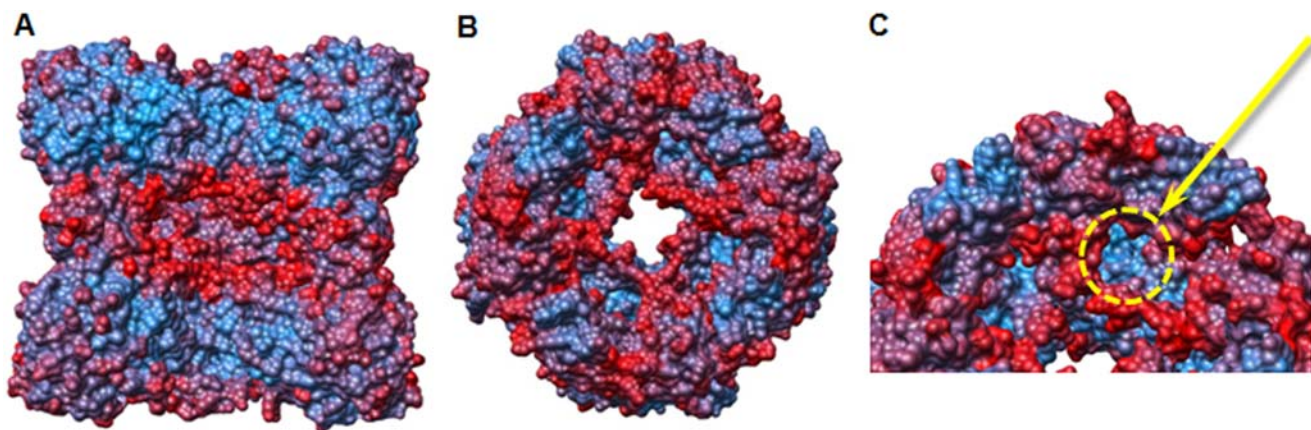
doi:10.1371/journal.pone.0004712.g006

(0.7499 ml/g) using the amino-acid composition. The sedimentation data were analyzed with the program Sedfit [25] using the continuous  $c(s)$  and  $c(M)$  distributions, and showed that 94% of the protein sedimented with a corrected sedimentation coefficient  $s_{20}^{\circ C,w} = 14.71 \pm 0.22$  S. The estimated frictional ratio ( $f/f_0 = 1.25$ ) is consistent with a globular shape of the particle. The molecular mass of the particle ( $391,420 \pm 10,670$  Da) is in very good agreement with the calculated molecular mass of the octamer (403,542 Da).

#### Cryo-electron microscopy

Pab87 was diluted to  $0.5 \text{ mg mL}^{-1}$  in 50 mM Tris-HCl buffer pH 8.5, 350 mM NaCl. 5  $\mu\text{L}$  of this solution was applied onto an electron microscopy grid covered with a holey carbon film. The grid was plunged into an ethane slush cooled with liquid nitrogen. The frozen hydrated specimen were observed at low temperature

on a cryo electron microscope (Tecnai F20, FEG cathode operated at 200 kV) and images were recorded under low-dose condition (less than  $2,000 \text{ electrons/nm}^2$ ) on Kodak SO163 photographic plates. Areas covered with individual molecules were recorded at a 50,000 times magnification. The micrographs were digitized with a  $5 \mu\text{m}$  raster size using a drum scanner (Primescan D7100, Heidelberg). The data was coarsened twice resulting in a final pixel spacing of 0.254 nm. Boxing and CTF phase correction of the 6,480 molecular images were performed in the EMAN software package. Image analysis was performed using the IMAGIC software package [26] (Image Science Software, Berlin, Germany) as described earlier [27]. The resolution of the final reconstructions was estimated to be around 14 Å from the Fourier shell correlation function obtained by comparing two reconstructions, generated by splitting randomly the data set in half and according to the 0.5 cut-off in the Fourier shell correlation curve



**Figure 7. Relative flexibility of the octamer.** The structure is colored by B-factors to illustrate flexibility (dynamic disorder). Red color reflects high flexibility, light blue reflects low flexibility relative to the B-factors of the overall complex. **A**, Open side view of the octamer. The front head-to-tail dimer has been removed. The  $\beta 5$ – $\beta 6$  and  $\beta 16$ – $\beta 17$  loops forming the two crown-shaped surfaces appear highly flexible. **B**, Open top view of the octamer down the four-fold axis. For better clarity, the four N-terminal PRP domains of the top tetramer and the four C-terminal domain  $\beta 16$ – $\beta 17$  loops of the bottom tetramer have been removed. **C**, Same figure as in **(B)** tilted so that the active site entrance is visible. Catalytic pocket is highlighted in yellow.

doi:10.1371/journal.pone.0004712.g007

(0.5 FSC criterion [28]). Figure 2D and Figure 7 were prepared with the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) [29] (<http://www.cgl.ucsf.edu/chimera>).

### Crystallization and diffraction data collection

Pab87 (7 mg mL<sup>-1</sup>) was mixed with lanthanide complex solution at required concentration to obtain Lu-HPDO3A final concentrations of 50, 100 and 200 mM. Pab87/Lu-HPDO3A complex was co-crystallized by hanging drop vapor diffusion method with 20 mM CaCl<sub>2</sub>, 30% MPD and 100 mM sodium acetate buffer pH 4.6 at 18°C. Diffraction data from a crystal obtained with Lu-HPDO3A concentration of 200 mM was used for SAD data collection at 2.2 Å resolution on ID29 beamline (ESRF, Grenoble, France). The wavelength was set to 1.3403 Å corresponding to the lutetium L<sub>III</sub> absorption edge ( $f'' \sim 28 \text{ e}^-$ ). Data were processed with XDS [30] and scaled with SCALA from the CCP4 suite [31]. The crystal has the symmetry of the space group *P1* with eight molecules per asymmetric unit (see Table 1 for crystallographic parameters).

### Structure determination and refinement

AutoSHARP [32] was used to locate the lutetium sites based on the Lu-derivative anomalous data and to generate initial SAD phases at 2.2 Å resolution with a figure of merit of 0.35. Phases further improved by density modification, with a figure of merit of 0.88, yielded a clearly interpretable map that was submitted to automatic building using BUCCANEER [33]. The resulting built model corresponding to 89% of the structure was completed using COOT [34] and refined with CNS [35]. The R-factor of the final model is 0.196 ( $R_{\text{free}} = 0.241$ ), using all data from 30.0 to 2.2 Å resolution (Table 1). The model contains amino acid residues 1–447 for chains A, B, F, G, H, 1–446 for chains C, D, E, 4 HPDO3A, 17 lutetium ions and 2,368 water molecules. The average root mean square deviation of the NCS-related subunits is 0.37 Å for the 446 superimposed C $\alpha$  atoms. Model quality and stereochemistry were checked using CNS and PROCHECK [36] (Table 1). Atomic coordinates and structure factors of protein Pab87 in complex with Lu-HPDO3A have been deposited in the Protein Data Bank with accession code 2QMI.

### Molecular modelling

The monomeric structure of the CubicO proteases from *Pyrococcus horikoshii*, *Caldivirga maquilgensis*, *Hyperthermus butylicus*, *Pyrobaculum aerophilum*, *Aeropyrum pernix* and marine  $\gamma$ -proteobacterium HTCC2207 were modelled by SWISS-MODEL [37] using the 3D structure of one Pab87 monomer from *Pyrococcus abyssi* as model (PDB id 2QMI). The octamers were generated using the same *D*<sub>4</sub> symmetry than the Pab87 octamer and optimized through rigid body minimization. Minimization process was pursued using simulated annealing and energy minimization with CNS [35]. The geometric quality of the models was assessed with Procheck [38]. The models were considered to be good with Verify3D [39], ProsaII [40] and Eval23D [41].

### Activity assays

The D- and L-aminopeptidase activities were detected by monitoring the formation of *p*-nitroaniline from D- or L-Ala-*p*-nitroanilide at 405 nm in a 20 mM Tris buffer (pH 8.0) from 40 to 90°C by 10°C steps [42].

The D-Ala-D-Ala carboxypeptidase activity was assayed on GlcNAc-MurNAc-L-Ala- $\gamma$ -D-Glu-L-(N<sup>6</sup>-D-Asp)Lys-D-Ala-D-Ala,

GlcNAc-MurNAc-L-Ala- $\gamma$ -D-Glu-L-(N<sup>6</sup>-L-Ala-L-Ala)Lys-D-Ala-D-Ala, GlcNAc-MurNAc-L-Ala- $\gamma$ -D-Glu-*meso*DAP-D-Ala-D-Ala and L-Ala- $\gamma$ -D-Glu-L-Lys-D-Ala-D-Ala. The assays were performed in a 10  $\mu$ L mixture containing protein Pab87 (5  $\mu$ g), Tris-HCl (20 mM, pH 8.0) and substrate (1 mM). The reaction mixtures were incubated for 90 min at 37 or 90°C, and desalted using a micro column (ZipTipC<sub>18</sub>; Millipore). The products of the reaction were analyzed by nanoelectrospray MS in the positive mode (Qstar Pulsar I; Applied Biosystems). The disaccharide-pentapeptide substituted by a D-aspartate residue (L-Lys<sup>3</sup>-D-Asp) was purified from the peptidoglycan of *Enterococcus faecium* M512 [43]. The disaccharide-pentapeptide substituted by an L-Ala-L-Ala side chain (L-Lys<sup>3</sup>-L-Ala-L-Ala) was purified from *Enterococcus faecalis* JH2-2 [44] and the disaccharide-pentapeptide containing *meso*-diaminopimelic acid (*meso*DAP<sup>3</sup>) was purified from *E. coli* strain ATCC 25113, a strain deleted for several DD-carboxypeptidases. The procedures used for peptidoglycan preparation, digestion with muramidases, and reduction of MurNAc to muramitol with sodium borohydride have been previously described for enterococci [43] and *E. coli* [45]. The resulting mucopeptides, disaccharide-tetrapeptides or tetrapeptide formed by the hydrolysis of the C-terminal D-Ala residue, were separated by RP-HPLC in acetonitrile gradients containing trifluoroacetic acid [43] and identified by mass spectrometry (MS). The concentration of the mucopeptides was estimated by amino acid analysis after acidic hydrolysis with a Hitachi autoanalyzer [46]. Pentapeptide L-Ala- $\gamma$ -D-Glu-L-Lys-D-Ala-D-Ala was purchased from Sigma.

### Supporting Information

**Figure S1** Structure-Corrected Sequence Alignment of CubicO protease family members. The sequence names are as following: pab87, *Pyrococcus abyssi* PAB0087, phori, *Pyrococcus horikoshii* PH0142/PH0143, cmaqu, *Caldivirga maquilgensis* Cmaq\_0116, paero, *Pyrobaculum aerophilum* PAE3237, hbuty, *Hyperthermus butylicus* Hbut\_1035, apern, *Aeropyrum pernix* Ap\_0338, gprot, marine  $\gamma$ -proteobacterium HTCC2207.  $\alpha$ -helices (cylinders) and  $\beta$ -strands (arrows) of *P. abyssi* Pab87 are aligned with the sequences and color coded according to Fig 1A (the peptidase  $\alpha/\beta$  and all-helical regions are in light blue and cyan, respectively, the linking helix in light orange and the lipocalin domain in salmon). Residues emphasized by black shading are 100% conserved, and gray shading represents 85% conservation in the multiple alignment of the 7 sequences. CubicO specific motifs are underlined by red stars (the YAEAF/L, FFYLN and ERY are located at residues 116–120, 158–162 and 440–442, respectively, in Pab87) and PRP specific motifs by green stars (the SXXK, YXN and HXG motifs are located at residues 59–62, 160–162 and 302–304, respectively, in Pab87).

Found at: doi:10.1371/journal.pone.0004712.s001 (0.09 MB PDF)

**Figure S2** Structure of the CubicO protease models from A, *Aeropyrum pernix*, B, *Caldivirga maquilgensis*, C, the marine  $\gamma$ -proteobacterium HTCC2207, D, *Hyperthermus butylicus*, E, *Pyrobaculum aerophilum*, F, *Pyrococcus abyssi*, G, *Pyrococcus horikoshii*. On left, open side view of the octamers. To visualize the internal cavity, the head-to-tail dimer in the front of the picture has been removed. The top and bottom tetrameric rings are colored in blue and green, respectively. The beta 5-beta 6 and beta 16-beta 17 loops forming the two crown-shaped surfaces are colored in salmon and raspberry, respectively. On right, ribbon representation of the octamers.

Found at: doi:10.1371/journal.pone.0004712.s002 (1.85 MB PDF)



## Acknowledgments

We thank S. Magnet for technical assistance in mass spectrometry experiments, M. Arthur for having kindly provided the mucopeptides, and P. L. Anelli and Bracco Imaging S.p.A. (Milano, Italy) for having kindly provided the Lu-HPDO3A complex. We acknowledge the European Synchrotron Radiation Facility for providing access to synchrotron radiation and D. Picot and L. Barucq for the collection of the Lu-data at beamline ID29, and S. Petrella for careful reading of the manuscript. A special thank to T. Terwilliger for helpful support in structure solving (consistent and similar results were obtained using the program PHENIX) and useful discussions.

## References

- Vandeputte-Rutten L, Gros P (2002) Novel proteases: common themes and surprising features. *Curr Opin Struct Biol* 12: 704–708.
- Lupas A, Flanagan JM, Tamura T, Baumeister W (1997) Self-compartmentalizing proteases. *Trends Biochem Sci* 22: 399–404.
- Barrett AJ, Rawlings ND (2007) ‘Species’ of peptidases. *Biol Chem* 388: 1151–1157.
- Bochtler M, Ditzel L, Groll M, Hartmann C, Huber R (1999) The proteasome. *Annu Rev Biophys Biomol Struct* 28: 295–317.
- Rohrwild M, Pfeifer G, Santarius U, Muller SA, Huang HC, et al. (1997) The ATP-dependent HslVU protease from *Escherichia coli* is a four-ring structure resembling the proteasome. *Nat Struct Biol* 4: 133–139.
- De Mot R, Nagy I, Walz J, Baumeister W (1999) Proteasomes and other self-compartmentalizing proteases in prokaryotes. *Trends Microbiol* 7: 88–92.
- Rasulova FS, Dergousova NI, Starkova NN, Melnikov EE, Rumsh LD, et al. (1998) The isolated proteolytic domain of *Escherichia coli* ATP-dependent protease Lon exhibits the peptidase activity. *FEBS Lett* 432: 179–181.
- Yao T, Cohen RE (1999) Giant proteases: beyond the proteasome. *Curr Biol* 9: R551–553.
- Brandstetter H, Kim JS, Groll M, Huber R (2001) Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature* 414: 466–470.
- Walz J, Koster AJ, Tamura T, Baumeister W (1999) Capsids of tricorn protease studied by electron cryomicroscopy. *J Struct Biol* 128: 65–68.
- Borissenko L, Groll M (2005) Crystal structure of TET protease reveals complementary protein degradation pathways in prokaryotes. *J Mol Biol* 346: 1207–1219.
- Franzetti B, Schoehn G, Hernandez JF, Jaquinod M, Ruigrok RW, et al. (2002) Tetrahedral aminopeptidase: a novel large protease complex from archaea. *EMBO J* 21: 2132–2138.
- Russo S, Baumann U (2004) Crystal structure of a dodecameric tetrahedral-shaped aminopeptidase. *J Biol Chem* 279: 51275–51281.
- Schoehn G, Vellieux FM, Asuncion Dura M, Receveur-Brechot V, Fabry CM, et al. (2006) An archaeal peptidase assembles into two different quaternary structures: A tetrahedron and a giant octahedron. *J Biol Chem* 281: 36327–36337.
- Rockel B, Peters J, Muller SA, Seyit G, Ringler P, et al. (2005) Molecular architecture and assembly mechanism of *Drosophila* tripeptidyl peptidase II. *Proc Natl Acad Sci U S A* 102: 10135–10140.
- Remaut H, Bompard-Gilles C, Goffin C, Frere JM, Van Becumen J (2001) Structure of the *Bacillus subtilis* D-aminopeptidase DppA reveals a novel self-compartmentalizing protease. *Nat Struct Biol* 8: 674–678.
- Girard E, Anelli PL, Vicat J, Kahn R (2003) High-phasing-power lanthanide derivatives: taking advantage of ytterbium and lutetium for optimized anomalous diffraction experiments using synchrotron radiation. *Acta Crystallogr D Biol Crystallogr* 59: 1877–1880.
- Grzyb J, Latowski D, Strzalka K (2006) Lipocalins - a family portrait. *J Plant Physiol* 163: 895–915.
- Bishop RE (2000) The bacterial lipocalins. *Biochim Biophys Acta* 1482: 73–83.
- DeLong EF (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci U S A* 89: 5685–5689.
- Kecha M, Benallaoua S, Touzel JP, Bonaly R, Duchiron F (2007) Biochemical and phylogenetic characterization of a novel terrestrial hyperthermophilic archaeon pertaining to the genus *Pyrococcus* from an Algerian hydrothermal hot spring. *Extremophiles* 11: 65–73.
- Plewniak F, Bianchetti L, Brelivet Y, Carles A, Chalmel F, et al. (2003) PipeAlign: A new toolkit for protein family analysis. *Nucleic Acids Res* 31: 3829–3832.
- Delfosse V, Hugonnet JE, Sougakoff W, Mayer C (2005) Cloning, purification, crystallization and preliminary crystallographic analysis of a penicillin-binding protein homologue from *Pyrococcus abyssi*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 61: 1006–1008.

## Author Contributions

Conceived and designed the experiments: VD CM. Performed the experiments: VD. Analyzed the data: VD CM. Wrote the paper: VD CM. Participated to structure solving: EG. Performed analytical ultracentrifugation experiments and analyzed the corresponding data: CB. Participated to activity assays experiments: MD. Participated to structure solving: MD. Participated in the writing of the paper: MD OP. Performed sequence analyzes: OP. Performed EM experiments: PS. Performed corresponding data analysis: PS.

- Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1992) Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge: Analytical Ultracentrifugation in Biochemistry and Polymer Science, The Royal Society of Chemistry, pp 90–125.
- Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78: 1606–1619.
- van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. *J Struct Biol* 116: 17–24.
- Jawhari A, Uhring M, De Carlo S, Crucifix C, Tocchini-Valentini G, et al. (2006) Structure and oligomeric state of human transcription factor TFIIIE. *EMBO Rep* 7: 500–505.
- Saxton WO, Baumeister W, Hahn M (1984) Three-dimensional reconstruction of imperfect two-dimensional crystals. *J Struct Biol* 13: 57–70.
- Petersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612.
- Kabsch W (1988) Automatic indexing of rotation diffraction patterns. *J Appl Cryst* 21: 67–71.
- Collaborative Computational Project 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50: 760–763.
- Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364: 215–230.
- Cowan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62: 1002–1011.
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126–2132.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54: 905–921.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26: 283–291.
- Kopp J, Schwede T (2004) The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res* 32: D230–234.
- Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res* 29: 221–222.
- Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in enzymology* 277: 396–404.
- Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* 17: 355–362.
- Gracy J, Chiche L, Sallantin J (1993) Improved alignment of weakly homologous protein sequences using structural information. *Protein Engineering, Design and Selection* 6: 821–829.
- Lottenberg R, Jackson CM (1983) Solution composition dependent variation in extinction coefficients for p-nitroaniline. *Biochim Biophys Acta* 742: 558–564.
- Arbeloa A, Hugonnet JE, Sentilhes AC, Josseaume N, Dubost L, et al. (2004) Synthesis of mosaic peptidoglycan cross-bridges by hybrid peptidoglycan assembly pathways in gram-positive bacteria. *J Biol Chem* 279: 41546–41556.
- Jacob AE, Hobbs SJ (1974) Conjugal transfer of plasmid-borne multiple antibiotic resistance in *Streptococcus faecalis* var. *zymogenes*. *J Bacteriol* 117: 360–372.
- Glauner B (1988) Separation and quantification of mucopeptides with high-performance liquid chromatography. *Anal Biochem* 172: 451–464.
- Mengin-Lecreux D, Falla T, Blanot D, van Heijenoort J, Adams DJ, et al. (1999) Expression of the *Staphylococcus aureus* UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:L-lysine ligase in *Escherichia coli* and effects on peptidoglycan biosynthesis and cell growth. *J Bacteriol* 181: 5909–5914.