BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

The first crystal structure of the peptidase domain of the U32 peptidase family

Magdalena Schacherl, Angelika A. M. Montada, Elena Brunstein and Ulrich Baumann*

Department of Chemistry, Institute of Biochemistry, University of Cologne, Otto-Fischer-Strasse 12-14, D-50674 Cologne, Germany. *Correspondence e-mail: ubaumann@uni-koeln.de

Received 18 August 2015
Accepted 15 October 2015

Edited by R. J. Read, University of Cambridge, England

Keywords: U32 peptidase; crystal structure; TIM barrel.

PDB reference: Mk0906, 5d88

Supporting information: this article has supporting information at journals.iucr.org/d

The U32 family is a collection of over 2500 annotated peptidases in the MEROPS database with unknown catalytic mechanism. They mainly occur in bacteria and archaea, but a few representatives have also been identified in eukarya. Many of the U32 members have been linked to pathogenicity, such as proteins from *Helicobacter* and *Salmonella*. The first crystal structure analysis of a U32 catalytic domain from *Methanopyrus kandleri* (gene *mk0906*) reveals a modified $(\beta\alpha)_8$ TIM-barrel fold with some unique features. The connecting segment between strands $\beta7$ and $\beta8$ is extended and helix $\alpha7$ is located on top of the C-terminal end of the barrel body. The protein exhibits a dimeric quaternary structure in which a zinc ion is symmetrically bound by histidine and cysteine side chains from both monomers. These residues reside in conserved sequence motifs. No typical proteolytic motifs are discernible in the three-dimensional structure, and biochemical assays failed to demonstrate proteolytic activity. A tunnel in which an acetate ion is bound is located in the C-terminal part of the β-barrel. Two hydrophobic grooves lead to a tunnel at the C-terminal end of the barrel in which an acetate ion is bound. One of the grooves binds to a *Strep*-Tag II of another dimer in the crystal lattice. Thus, these grooves may be binding sites for hydrophobic peptides or other ligands.

1. Introduction

Proteolytic enzymes control a plethora of biological processes including food uptake and digestion, blood clotting, protein quality control, signalling and apoptosis (López-Otín & Bond, 2008; Neurath, 1986). In most genomes, some 2–5% of all genes encode proteases (Puente *et al.*, 2003). These are usually classified according to their catalytic mechanism and key residues as serine (S), cysteine (C), aspartic (A), metallo- (M) and N-terminal nucleophile/threonine (T) peptidases plus a few others. Of the more than 490 000 entries in the MEROPS database (Rawlings *et al.*, 2011), over 478 000 belong to these five classes. For some 7000 entries the catalytic mechanism is not known and these proteins are designated 'U'. One set of these is the U32 family, a collection of about 2700 homologous proteins of mostly bacterial and archaeal origin, although a few members are present in eukarya, such as protozoa, plants and animals, and two U32 peptidase homologues are even found in bacteriophages. They all contain the so-called U32 peptidase domain that can be assigned to various U32 subfamilies, and frequently some other domains of unknown function. Their molecular weights typically range from 26 to 70 kDa. The larger ones are multidomain proteins that contain further domains of unknown function in addition to the assigned U32 peptidase domain.

PrtC from *Porphyromonas gingivalis* was the first U32 peptidase for which a proteolytic activity for type I collagen and azocoll was reported (Kato *et al.*, 1992). For *Helicobacter*

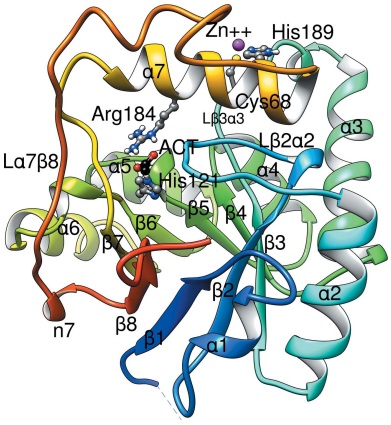


Table 1
Data collection and processing.

Values in parentheses are for the outer shell.

	$KAu(CN)_2$ derivative	Native
Data collection		
Diffraction source	Rigaku MicroMax-007 HF	X06DA, Swiss Light Source
Wavelength (Å)	1.5418	1.000
Temperature (K)	100	100
Detector	MAR345	PILATUS 2M
Space group	$P3_221$	$P3_221$
Unit-cell parameters (Å, °)	$a = b = 79.1$, $c = 89.3$, $\alpha = \beta = 90$, $\gamma = 120$	$a = b = 79.2$, $c = 89.2$, $\alpha = \beta = 90$, $\gamma = 120$
Resolution range (Å)	20.0–2.43 (2.58–2.43)	50.0–1.66 (1.71–1.66)
Total No. of reflections	167210 (21826)	375790 (58542)
No. of unique reflections†	23222 (3495)	73564 (11854)
Completeness (%)	98.8 (93.8)	99.6 (98.8)
Multiplicity	7.2 (6.2)	5.1 (4.9)
$\langle I/\sigma(I)\rangle$	26.7 (6.0)	29.7 (3.1)
R_{meas}‡	0.079 (0.392)	0.029 (0.498)
Overall B factor from Wilson plot (Å^2)		26.7
Phasing		
No. of sites	6	
Resolution range (Å)	20.0–2.43	
FOM	0.40	
Refinement		
PDB code		5d88
Resolution range (Å)		39.58–1.66 (1.72–1.66)
Completeness (%)		99.6
σ Cutoff		−3.0
No. of reflections, working set		72080
No. of reflections, test set		1466
Final R_{cryst}		0.133 (0.215)
Final R_{free}		0.152 (0.273)
No. of non-H atoms		
Protein		1976
Zn^{2+}		1
Acetate		4
Water		161
Total		2142
R.m.s. deviations§		
Bonds (Å)		0.009
Angles (°)		0.99
B factors (Å^2)		
Average		38.6
Protein		38.1
Zn^{2+}		23.7
Acetate		29.6
Water		45.1
MolProbity geometry analysis¶		
Ramachandran plot		
Most favoured (%)		97.6
Allowed (%)		2.4
Rotamer outliers (%)		0.5
Clashscore		1.3

† Friedel pairs are counted as separate reflections. ‡ R_{meas} defined as by Diederichs & Karplus (1997). § Ideal values are according to Engh & Huber (1991). ¶ Chen *et al.* (2010).

pylori HP0169 proteolytic activity against collagen and an essential function in gastric colonization have been reported (Kavermann *et al.*, 2003). Furthermore, a specific collagen I activity has beem reported for the protein A0A098L2Q3 from *Geobacillus thermoleovorans* (Jasilionis *et al.*, 2012). In the first two cases, the proteolytic activities of both proteins were reported to be sensitive to EDTA treatment, thus leading to the conclusion of a metalloprotease-type activity, while in the latter Zn^{2+} ions increased the thermostability. However, none of the typical metalloprotease, for example HE*XX*H, or collagen-binding sequence motifs is present in any of the U32 peptidases. On the other hand, the MEROPS identifiers U32.001 to U32.003 share at least an E-*x*-F-x_2-G-(SA)-(LIVM)-C-x_4-G-*x*-C-*x*-(LIVM)-S sequence motif that is indicative of these subfamilies. However, the function of these amino acids has not been investigated.

Many of the U32 members are implicated in pathogenicity, such as the *yegQ* gene of *Salmonella enterica* (Wu *et al.*, 2002; Xiong *et al.*, 2011) or the operon-encoded proteins YprA and YprB from *Yersina ruckeri*, which increase virulence in a fish model (Navais *et al.*, 2014). Furthermore, a gene encoding a U32 member was found on a plasmid conferring antibiotic resistance in a *Providencia stuartii* strain (McGann *et al.*, 2012).

To date, no three-dimensional structure has been determined for the U32 peptidase domain of any member. The only structural information available is for a small 10 kDa nonpeptidase domain corresponding to the C-terminal 90 amino acids of the U32.002 protein from *G. thermoleovorans* (Trillo-Muyo *et al.*, 2013), revealing an antiparallel eight-stranded β-barrel (PDB entry 4he6) which is similar to the fMet–tRNA^{fMet} binding domain of bacterial IF2. The function of this domain is unknown, but it occurs additionally to the U32 peptidase domain in other members of the U32 family, for example at the C-terminus of the abovementioned *P. gingivalis* PrtC. This protein has a polypeptide chain of about 330 amino acids that fold into an amino-terminal U32 peptidase domain, followed by this IF2-like domain.

In order to gain insight into the three-dimensional structure of the U32 peptidase domain and possibly into the novel proteolytic catalytic mechanism, we undertook the structure determination of Mk0906, a U32 peptidase from *Methanopyrus kandleri*, one of the most thermophilic organisms. Mk0906 has a rather distant relationship to the U32 members mentioned above and is a rather unique protein, with the closest homologues exhibiting a sequence identity of below 40% in a core region of about 150 amino acids out of 234. It basically consists of only the U32 peptidase domain, but lacks the abovementioned sequence motif. Our studies reveal a TIM-barrel fold with some irregularities and a lack of structural motifs suitable for catalysing peptide-bond hydrolysis.

2. Materials and methods

2.1. Protein expression, purification and crystallization

The pJExpress411 vector containing a codon-optimized *mk0906* gene (DNA2.0, USA) bearing an N-terminal thrombin-cleavable *Strep*-Tag II (the sequence of the whole construct is given in the Supporting Information) was transformed into *Escherichia coli* BL21 (DE3) cells by heat shock, plated on LB agar plus 50 µg ml^{-1} kanamycin plates and incubated overnight at 37°C. Pre-cultures grown overnight from a single colony at 37°C were used to inoculate 1 l expression culture (LB medium plus 50 µg ml^{-1} kanamycin) to an optical density at 600 nm (OD_{600}) of 0.1. Gene expression was induced with 0.5 m*M* isopropyl *β*-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6 and cell growth was continued for a further 5 h at 37°C. The cells were harvested by centrifugation at 7000*g* and 4°C for 20 min. The cell pellets were washed with TBS buffer (20 m*M* Tris pH 7.5, 200 m*M* NaCl), pelleted again and stored at −80°C until use.

A cell pellet from 1 l culture was solubilized in TBS with 10 µg ml^{-1} DNaseI (AppliChem, Germany). The cells were lysed on ice/water by sonication for 15 min at 30% amplitude (Vibra-Cell VCX500, Sonics, Switzerland). The cell debris was pelleted at 10 000*g* at 4°C for 10 min and the supernatant was cleared by ultracentrifugation at 165 000*g* at 4°C for 30 min. The supernatant was loaded onto a 1 ml *Strep*-Tactin column (IBA, Germany) and washed with TBS. Protein was eluted with TBS containing 2.5 m*M* desthiobiotin. The protein was concentrated and further purified on a HiLoad Superdex 200 16/600 column (GE Healthcare, Germany) equilibrated with TBS supplemented with 2 m*M* DTT. The protein concentration was determined at 280 nm using a molar extinction coefficient of 35 410 M^{-1} cm^{-1}. The yield was 8 mg of protein per litre of culture.

Crystals were grown by the sitting-drop vapour-diffusion technique. Protein (10 mg ml^{-1}) was mixed with precipitant in a 1:1 ratio to form 2 µl drops, which were then equilibrated at 20°C against the cryocondition 2.5 *M* sodium acetate pH 7.0. Crystals were harvested in nylon loops and directly flash-cooled in liquid nitrogen.

2.2. X-ray diffraction data collection, structure solution and refinement

Initial data were collected in-house employing a Rigaku MicroMax-007 HF rotating-anode generator equipped with confocal mirrors (VariMax HF, Rigaku, Japan) and a MAR345 image-plate detector (marXperts, formerly MAR Research, Hamburg, Germany) at 100 K using an Oxford Cryostream 700 (Oxford Cryosystems, Oxford, England). Native and derivative data were processed using *XDS* (Kabsch, 2010). The structure was solved by the SAD method using a gold derivative that was prepared by soaking crystals with 1 m*M* $KAu(CN)_2$ for 1 h. Heavy-atom positions were determined and phases were computed using the *PHENIX* package (Adams *et al.*, 2011; Read & McCoy, 2011). Initially, automated model building and refinement was performed using *ARP*/*wARP* (Langer *et al.*, 2008) and *phenix.refine* (Afonine *et al.*, 2012) using the in-house data. High-resolution native data and peak data at the zinc edge were subsequently collected on beamline X06DA of the Swiss Light Source at the Paul Scherrer Institute, Villigen, Switzerland using a PILATUS 2M detector (Dectris, Switzerland), followed by further refinement against the high-resolution data. All data-collection and refinement statistics are given in Table 1.

2.3. Proteolytic assays

2.3.1. Collagen-cleavage assay. 5 µg rat-tail collagen type I (BD Biosciences, Oakpark, USA) was incubated with 200 or 50 µ*M* (referring to a monomer) freshly purified *Strep*-tagged MK0906 in TBS buffer plus 5 m*M* calcium chloride (Fluka, Germany) in a volume of 200 µl. The total volume was distributed equally into 1.5 ml test tubes and incubated for 0 min, 3 h and 16 h at 37°C. As a positive control, *Clostridium histolyticum* collagenase from the EnzCheck Gelatinase/Collagenase Assay Kit (Molecular Probes, USA) was used at a final concentration of 0.05 U per 200 µl. After the incubation time, 5× SDS loading buffer was added to the samples, followed by incubation at 95°C for 5 min and a subsequent centrifugation step (20 000*g*, 30 s). The samples were applied onto 4–20% Mini-PROTEAN TGX Precast Protein Gels (Bio-Rad, Germany) and the gels were stained with Coomassie Blue.

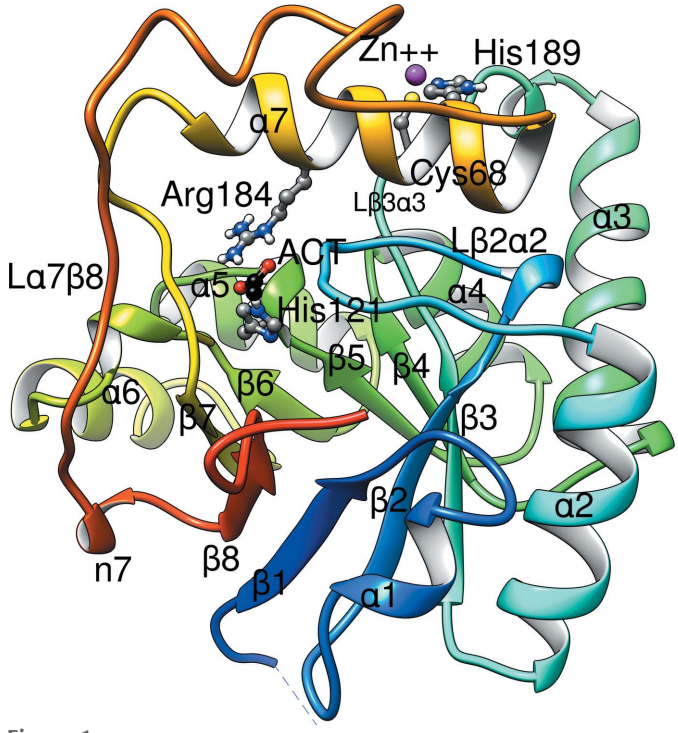


Figure 1
Overall fold of the Mk0906 monomer. A rainbow-coloured cartoon is shown with the N-terminus in blue and the C-terminus in red. Secondary structures are labelled. The N-terminal *Strep*-Tag II is not shown. The bound acetate ion (ACT) together with two of its interacting residues (Arg184 and His121) is depicted in ball-and-stick representation. The Zn^{2+} ion (magenta) that is important for the dimeric quaternary structure is shown together with its ligands Cys68 and His189. This figure was produced with *Chimera* (Pettersen *et al.*, 2004).

2.3.2. BODIPY protease assay. The protease activity was determined using the EnzCheck Protease Assay Kit (Molecular Probes, Thermo Fisher Scientific, Germany). Red-fluorescent BODIPY TR-X stock solution (1 mg ml^{-1}) was prepared according to the manual. For the assay, a 10 µg ml^{-1} working solution was prepared by diluting the BODIPY TR-X stock solution in 1× digestion buffer (10 m*M* Tris–HCl pH 7.8, 0.1 m*M* sodium azide). The working solution was mixed in a 1:1 ratio in white 384-microwell plates (HiBase, flat bottom and medium binding, Greiner, Germany) with different concentrations of either trypsin (positive control) or freshly purified *Strep*-tagged MK0906 in a final volume of 20 µl. Incubation was performed for up to 4 h in the dark at 37°C. After the incubation time, fluorescence was recorded using a Synergy HT plate reader (Biotek; excitation 590 ± 20 nm, emission 645 ± 40 nm). Triplicates were recorded for each condition and were corrected for background readings.

3. Results and discussion

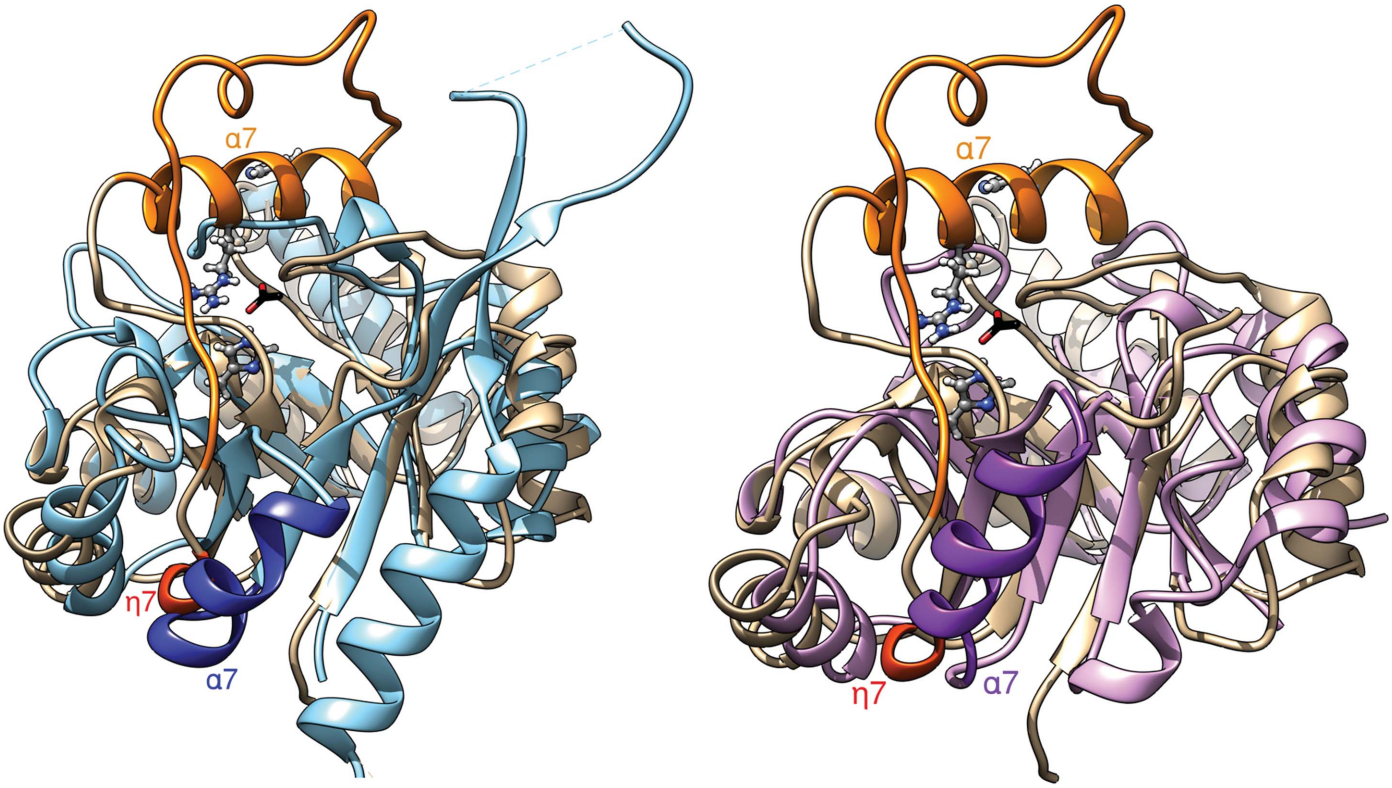
3.1. Overall structure

Heterologous protein expression yielded good amounts of homogenous protein with a purity of >98%. The expressed protein carries an N-terminal *Strep*-Tag II and a thrombin cleavage site, adding 20 amino acids in front of the native amino-terminal methionine residue. The purified protein crystallized readily in the trigonal space group $P3_221$. Phases were determined by SAD employing a gold derivative and our in-house X-ray equipment. High-resolution data were collected at the Swiss Light Source and refinement at 1.66 Å resolution resulted in good statistics (Table 1). Owing to the high redundancy and rather strong anomalous signal, Friedel pairs were kept separate and the zinc ion in the structure was treated as an anomalous group in *phenix.refine*. In the crystal lattice a stable dimer is clearly discernible, which will be discussed later. The N-terminal *Strep*-Tag II mediates a crystal contact between two different, nontouching dimers over a distance of more than 50 Å in the lattice.

The monomer exhibits a $(\beta\alpha)_8$ TIM-barrel fold with helix α8 missing and approximate dimensions of 40 × 30 × 30 Å (Fig. 1).

Not unsurprisingly, a *DALI* search of the PDB revealed a large number of similar structures with the TIM-barrel fold, for example the copper homeostasis protein CutCm (Zhu *et al.*, 2005; PDB entry 1x7i) with a *Z*-score of 14.7 (Fig. 2*a*) and the phosphoribosylanthranilate isomerase (Thoma *et al.*, 2000; PDB entry 1dl3) with a *Z*-score of 13.9 (Fig. 2*b*).

However, there are distinct features in the Mk0906 structure. Helix α8 is missing, and the segment connecting β7 and



(*a*) (*b*)

Figure 2
Overlay of Mk0906 with related structures. (*a*) Overlay with the copper homeostasis protein CutCm from *Shigella flexneri* depicted in blue (PDB entry 1x7i; Zhu *et al.*, 2005). (*b*) Superposition with the phosphoriboslyanthranilate isomerase from *Thermotoga maritima* shown in pink (PDB entry 1dl3; Thoma *et al.*, 2000). Mk0906 is shown in wheat. Helix α7 of all proteins is shown in a slightly different colour (orange in Mk0906, dark blue in CutCm and purple in phosphoriboslyanthranilate isomerase). The 3_{10}-helix η7 of Mk0906 is also shown in a bolder colour, illustrating its position, which is usually taken by helix α7 in canonical TIM-barrel structures.

β8 is quite long (>40 amino acids; Fig. 3). Within this stretch, helix α7 (residues 179–194) is displaced from its usual position at the lateral barrel side at the front in Fig. 1 to the top, where it traverses the C-terminal side of the barrel (Fig. 1). Also, the loop connecting helix α7 and β8 (termed Lα7β8) is very long (residues 195–223) and contains a 3_{10}-helix (residues 218–221, denoted η7 in Fig. 3) that occupies part of the space where helix α7 is usually located in canonical TIM-barrel structures. Together with the protruding loop Lβ2α2 (residues 28–42), loop Lα7β8 and helix α7 form a cap on the C-terminal side of the barrel. An extensive hydrogen-bonding network connects loop Lα7β8 and helix α7. The completely conserved Arg209 is totally buried and forms two salt bridges with Asp177 and Glu183, with the latter also being engaged in a salt bridge with Arg211. Furthermore, helix α7 and loop Lα7β8 make major contributions to the dimer interface, where they mediate a number of nonpolar contacts and hydrogen bonds.

3.2. The large dimer interface contains a conserved Zn^{2+} ion

Two monomers related by a crystallographic twofold axis form a tight dimer with a buried interface of about 3250 Å^2 (Fig. 4). Besides involving α7 and Lα7β8, the interface is mainly formed by the loop Lβ3α3 (residues 64–75) and the N-terminal end of helix α4. The loop Lβ3α3 carries a ^{68}CUG*XX*HL^{74} motif that is discernible from the sequence alignment in Fig. 3, where U represents a nonpolar amino acid. The cysteine residue (Cys68) of this motif binds to a Zn^{2+} ion, which is further coordinated by His189, which is located on helix α7 and is part of the ^{185}NUFSH^{189} motif. The other two ligands are the symmetry-equivalent residues Cys68′ and His189′ of the other monomer. The Zn^{2+}-binding signature motifs CUG*XX*HL and NUFSH are conserved in the nearest archaeal homologues (Fig. 3), and even some bacterial sequences show these motifs (Supplementary Fig. S1).

The zinc ion is located on one side of the interface. On the other side there is a network of aromatic side chains that interact tightly with each other (Fig. 4*b*): Tyr104, Phe175, Phe181, Phe182 and Tyr205 are located within the interface. Some contacts are of the 'offset-stacked' π–π type (Phe175–Phe175'), while others more closely resemble the 'edge-to-face' approach (Phe181–Phe181′). Additionally, the aromatic rings interact with other groups, for example alkyl CH groups (Phe182–Pro103′).

In addition to the zinc-mediated coordinative bonds and the hydrophobic contacts, there are ten hydrogen bonds across the interface. For example, the side chain of the conserved Asn185, which is located on helix α7, is involved in a hydrogen-

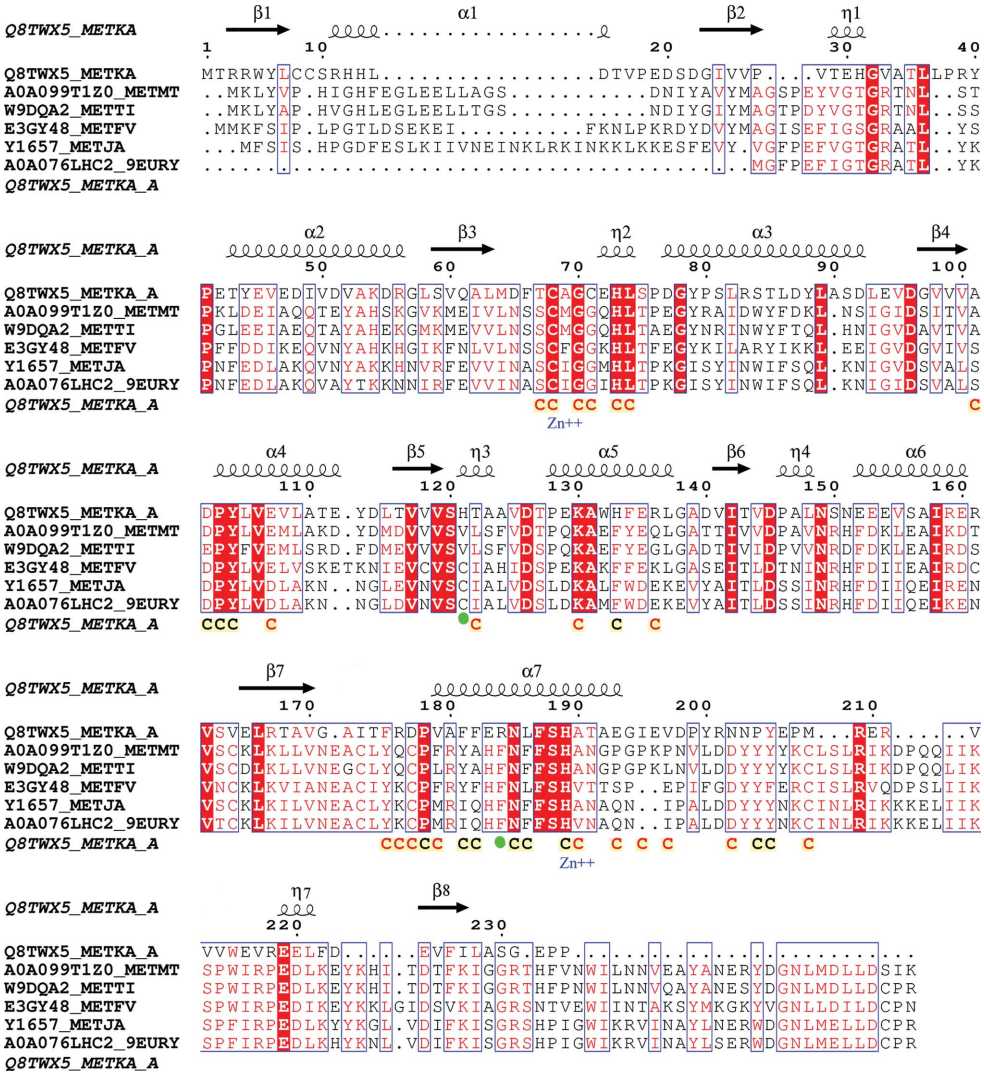


Figure 3
Sequence alignment of Mk0906. A0A099T1Z0_METMT, *Methanococcoides methylutens*, 39% sequence identity for 155 overlapping amino acids; W9DQA2_METTI, *Methanolobus tindarius*, 34% sequence identity for 150 overlapping amino acids; E3GY48_METFV, *Methanothermus fervidus*, 38% sequence identity for 149 overlapping amino acids; Y1657_METJA, *Methanocaldococcus jannaschii*, 33% sequence identity for 175 overlapping amino acids; A0A076LHC2_9EURY, *Methanocaldococcus* sp., 30% sequence identity for 175 overlapping amino acids. Secondary-structure elements from the Mk0906 structure are shown at the top; a 'C' at the bottom denotes a contact in the dimer interface and those in red type are involved in hydrogen bonds across the dimer interface. The zinc-binding residues (Zn^{++}) and those interacting with the acetate carboxylate group (green dots) are indicated. This figure was produced with *ESPript* (Gouet *et al.*, 2003).

bonding network with its symmetry-equivalent counterpart (two hydrogen bonds *via* side-chain $NH_2\cdots O{=}C$ interaction) as well as with the main-chain carbonyl O atom of Thr67 from the other subunit. The ammonium group of Lys130 forms a hydrogen bond to the main-chain carbonyl O atom of Phe175′, and Tyr205 donates a hydrogen bond to Asp102′.

3.3. A putative binding site for apolar ligands interacts with the *Strep*-Tag II

Screenings for proteolytic activity either by general proteolytic assays such as BODIPY casein or by specific collagenase/gelatinase tests such as collagen/gelatin digestion monitored by SDS–PAGE all turned all out negative, despite being performed under various conditions (varying the pH and the temperature) and employing a vast excess of Mk0906. Some representative experiments are shown in Supplementary Fig. S2.

Inspection of the three-dimensional structure did not reveal any motifs that are characteristic of hydrolases, for example catalytic triads or dyads, metal sites *etc.* Furthermore, no three-dimensional arrangements of amino acids suitable for the activation of a water molecule could be discerned. Therefore, we looked for potential active-site pockets in the Mk0906 structure that are not necessarily related to peptide hydrolysis. An acetate ion originating from the crystallization buffer is clearly defined in the electron-density map. It is bound at the end of a deep and narrow pocket by Arg184, which is located on helix $\alpha7$ (Fig. 5). Because the active centres of TIM-barrel enzymes are located on the C-terminal side of the β-barrel, this position is a reasonable candidate for an active centre. Arg184 is deeply buried between helix $\alpha7$ and blocks a continuous tunnel running from the entrance site lined by Ala101 and Ala172 to the exit located close to Thr122. Its guanidinium head group forms hydrogen bonds to the carboxylate of the acetate ion located at the entrance site and water molecules situated at the exit of the tunnel, as well as to the main-chain carbonyl O atom of Ile173. The acetate receives further hydrogen bonds from the main-chain amide NH groups of Leu36 and Ile173. His121 is situated just beneath the acetate ion. It sits on loop L$\beta5\alpha5$ on a short 3_{10}-helical segment and donates a hydrogen bond to the carboxylate group of the acetate. Interestingly, both residues are not conserved in the closest homologues (Fig. 3).

In front of the tunnel entrance there is an extended hydrophobic groove (termed Groove I) lined by Leu36, Leu37, Pro38, Ile173, Leu228, Ala229, Val213 and Trp215 and extending in the north–south direction in Fig. 6. At the ω-site of this groove, *i.e.* opposite the carboxylate-binding tunnel, the groove is covered by an arc formed by the C-terminus, namely Ser230, Gly213 and Glu232, Pro233 and Pro234. This arc might be flexible and open and close upon the binding of larger hydrophobic ligands.

Interestingly, this site is also part of the interaction surface for the *Strep*-Tag II of another monomer (belonging to another dimer) in the crystal lattice. The N-terminal extension introduced by the tag and the thrombin site (−20)MGS-WSHPQFEKSSGLVPRGS(−1), where negative numbers refer to the artificially

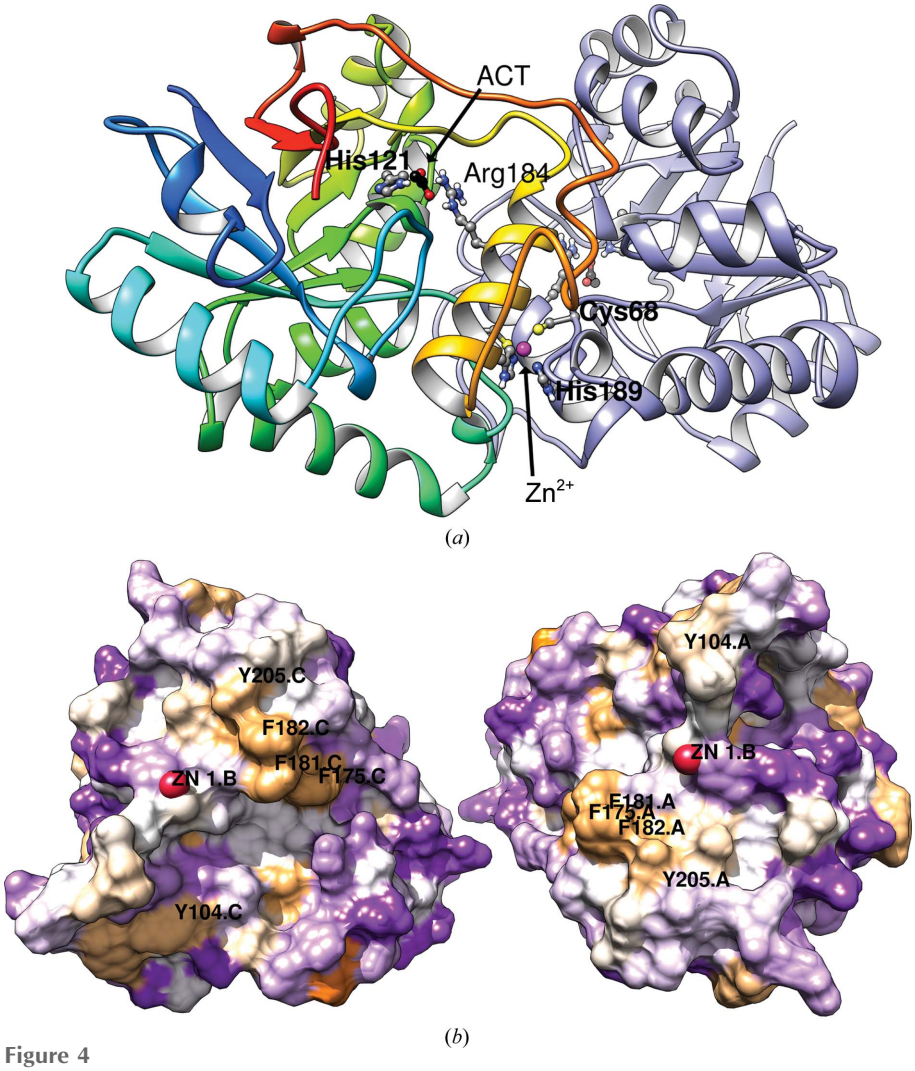


Figure 4
Dimeric quaternary structure of Mk0906 and the dimer interface. (*a*) Ribbon diagram of the dimer. One of the two protomers is coloured as in Fig. 1 and the other is in slate blue. The zinc-coordinating residues and the acetate ion together with the interacting residues are labelled. (*b*) A view of the dimerization surfaces of both monomers. The dimer is reconstituted by rotating both monomers by approximately 90° around the vertical to each other and superposing the zinc ion shown as a red sphere as the reference point. The surface is coloured according to polarity applying the Wimley–White hydrophobicity scale (Wimley & White, 1996) ranging from the most negative values (polar residues, depicted in purple) through 0 (white) to the most positive values (apolar residues, orange).

introduced amino acids, is well defined in the electron-density map from Gly(−19) to Ser(−9) (Fig. 6*c*), thus comprising all eight amino acids from the *Strep*-Tag II (underlined) at the N-terminus and lacking only the starting methionine, which may have been post-translationally cleaved off. The residues Ser(−8) to Arg(−3) are disordered. The visible amino acids of the tag are bound to another mainly hydrophobic groove (Groove II) from a symmetry-related molecule in the crystal lattice. Groove II branches off from the acetate-binding tunnel in a southeast direction in Fig. 6(*a*) and lies between the C-terminal end of loop Lβ2α2 and the N-terminal part of loop Lα7β8. In the N-terminal tag segment, residues Gly(−19) to Ser(−9) build an antiparallel β-hairpin, with amino acids Ser(−18) to His(−15) forming a type I tight turn. The indole side chain of Trp(−17) is deeply buried close to the ω-site in Groove I. The peptide extends from His(−15) in an extended β-conformation along another hydrophobic groove running in a southeast direction from the acetate-binding tunnel in Fig. 6(*b*), with Pro(−14) and Phe(−12) sitting in hydrophobic pockets at the bottom of Groove II.

4. Discussion

The first structure of of a U32 peptidase domain reveals a TIM-barrel fold with some variations.

A striking feature of the Mk0906 structure is the dimeric quaternary structure, which is accomplished, in addition to other interactions, by a Zn^{2+} ion coordinating symmetrically to residues from both monomers. The sequence motifs carrying the zinc ligands and most of the residues forming the dimer interface are conserved in other homologous proteins, although the sequence identities are only between 30 and 40% and the biological and also the biochemical functions of these proteins may be very different.

There are no typical three-dimensional arrangements visible that resemble protease motifs nor are there constellations that could serve as water-activating entities. While TIM-barrel enzymes are very versatile and catalyse many different reactions, including hydrolytic events, they have never been implicated as proteolytic enzymes, and their hydrolysis of any amide bond has not yet been reported. Also, taking into account the negative proteolytic assays, we believe that the annotation of at least Mk0906 as a peptidase should possibly be re-evaluated. Mk0906 is very distant from *P. ginigivalis* PrtC, the proteotypic U32 peptidase, with a sequence identity of less than 20% (Supplementary Fig. S3). PrtC essentially consists of an U32 domain followed by an eight-stranded β-barrel C-terminal domain, as reported for *G. thermoleovorans* (Trillo-Muyo *et al.*, 2013; PDB entry 4he6). It should be noted that the few conserved amino acids indicated in the alignment appear to serve a structural function rather than a catalytic function.

Our structure reveals a putative binding site for an amphiphilic molecule with a carboxylate group or a similarly negatively charged moiety at one end and a mainly hydrophobic tail. This putative binding site consists of two grooves that are largely lined by hydrophobic amino acids and where a crystal contact is made to the *Strep*-Tag II of a crystallographically related dimer. This could be a binding site for peptide segments with hydrophobic side chains but also for other hydrophobic ligands.

In summary, our crystal structure analysis reveals a TIM-barrel fold for the U32 peptidase domain. The large sequence diversity from other, reportedly proteolytically active U32

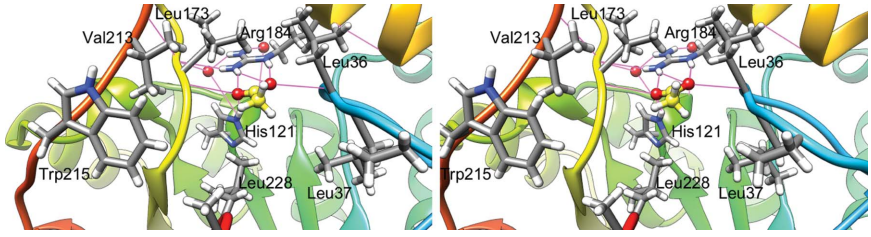


Figure 5
The acetate-binding site. Close-up wall-eyed stereoview of the interactions of the acetate ion and the residues lining the hydrophobic cleft (Groove I). Hydrogen bonds are indicated as purple lines.

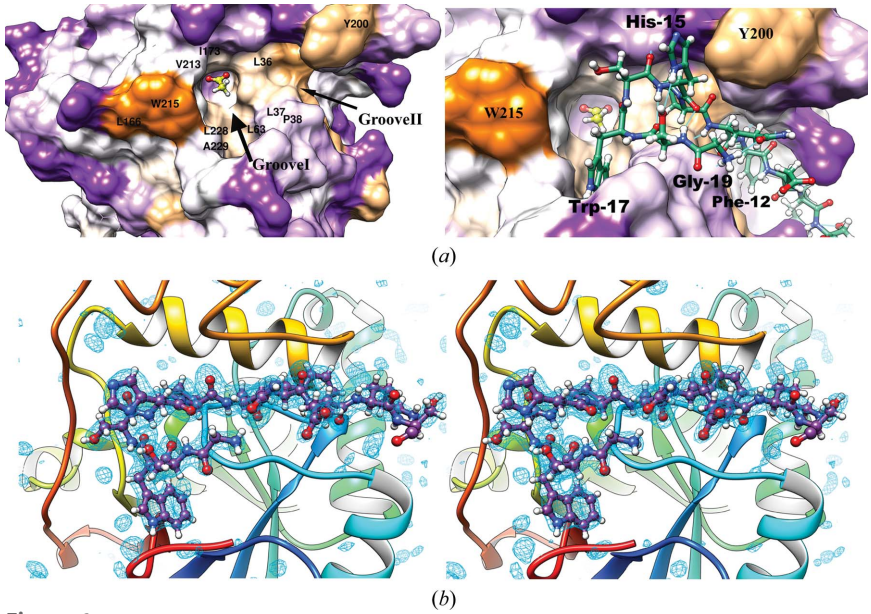


Figure 6
A putative active centre. (*a*) The pocket with the bound acetate ion (yellow C atoms) is shown. The surface colouring according to hydrophobicity as in Fig. 4(*b*) (polar in purple through white to apolar in orange) showing the extended hydrophobic grooves. (*b*) The *Strep*-Tag II mostly bound to Groove II is depicted in ball-and-stick representation with green C atoms and with some residues labelled in bold. A simulated-annealing OMIT $F_o - F_c$ map at the *Strep*-Tag II is contoured at 3.0σ and shown as a cyan mesh in a divergent-view stereo representation.

members together with the absence of active-site motifs and a negative result in proteolytic activities could possibly imply a biochemical function that differs from proteolysis for Mk0906. It should also be stressed that the published proteolytic activities of PrtC and *H. pylori* HP0169 have never been confirmed in the literature. Interestingly, for at least some members of the U32 peptidase family a role in anaerobic or microaerobic growth has been suggested, but without further characterization of the function of these proteins (Filiatrault *et al.*, 2006; Navais *et al.*, 2014).

5. Related literature

The following reference is cited in the Supporting Information for this article: Katoh *et al.* (2005).

Acknowledgements

Data-collection experiments were performed on beamline X06DA of the Swiss Light Source (SLS), Paul Scherrer Institute, Villigen, Switzerland. The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement No. 283570 (BioStruct-X). The project was supported by the University of Cologne and grant INST 216/682-1 FUGG from the German Research Council.

References

Adams, P. D. *et al.* (2011). *Methods*, **55**, 94–106.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Mol. Biol.* **4**, 269–275.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Filiatrault, M. J., Picardo, K. F., Ngai, H., Passador, L. & Iglewski, B. H. (2006). *Infect. Immun.* **74**, 4237–4245.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
Jasilionis, A., Kaupinis, A., Ger, M., Valius, M., Chitavichius, D. & Kuisiene, N. (2012). *Cent. Eur. J. Biol.* **7**, 587–595.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kato, T., Takahashi, N. & Kuramitsu, H. K. (1992). *J. Bacteriol.* **174**, 3889–3895.
Katoh, K., Kuma, K., Toh, H. & Miyata, T. (2005). *Nucleic Acids Res.* **33**, 511–518.
Kavermann, H., Burns, B. P., Angermuller, K., Odenbreit, S., Fischer, W., Melchers, K. & Haas, R. (2003). *J. Exp. Med.* **197**, 813–822.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
López-Otín, C. & Bond, J. S. (2008). *J. Biol. Chem.* **283**, 30433–30437.
Mc Gann, P., Hang, J., Clifford, R. J., Yang, Y., Kwak, Y. I., Kuschner, R. A., Lesho, E. P. & Waterman, P. E. (2012). *Antimicrob. Agents Chemother.* **56**, 1673–1679.
Navais, R., Méndez, J., Pérez-Pascual, D., Cascales, D. & Guijarro, J. A. (2014). *Virulence*, **5**, 619–624.
Neurath, H. (1986). *J. Cell. Biochem.* **32**, 35–49.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Puente, X. S., Sánchez, L. M., Overall, C. M. & López-Otín, C. (2003). *Nature Rev. Genet.* **4**, 544–558.
Rawlings, N. D., Barrett, A. J. & Bateman, A. (2011). *Nucleic Acids Res.* **40**, D343–D350.
Read, R. J. & McCoy, A. J. (2011). *Acta Cryst.* D**67**, 338–344.
Thoma, R., Hennig, M., Sterner, R. & Kirschner, K. (2000). *Structure*, **8**, 265–276.
Trillo-Muyo, S., Jasilionis, A., Domagalski, M. J., Chruszcz, M., Minor, W., Kuisiene, N., Arolas, J. L., Solà, M. & Gomis-Rüth, F. X. (2013). *Acta Cryst.* D**69**, 464–470.
Wimley, W. C. & White, S. H. (1996). *Nature Struct. Mol. Biol.* **3**, 842–848.
Wu, M. T., Carlson, S. A. & Meyerholz, D. K. (2002). *Microb. Pathog.* **33**, 279–287.
Xiong, N., Brewer, M. T., Anderson, K. L., Weeks, K. E. & Carlson, S. A. (2011). *Microb. Pathog.* **51**, 230–232.
Zhu, D.-Y., Zhu, Y.-Q., Huang, R.-H., Xiang, Y., Yang, N., Lu, H.-X., Li, G.-P., Jin, Q. & Wang, D.-C. (2005). *Proteins*, **58**, 764–768.