

# Crystal Structure and Activity Studies of the C11 Cysteine Peptidase from *Parabacteroides merdae* in the Human Gut Microbiome\*

Received for publication, November 27, 2015, and in revised form, February 4, 2016. Published, JBC Papers in Press, March 3, 2016, DOI 10.1074/jbc.M115.706143

 Karen McLuskey<sup>‡1</sup>,  Jaspreet S. Grewal<sup>‡§1</sup>, Debanu Das<sup>||</sup>, Adam Godzik<sup>||\*\*</sup>, Scott A. Lesley<sup>||##5§</sup>,  
 Ashley M. Deacon<sup>||</sup>, Graham H. Coombs<sup>||</sup>, Marc-André Elsliger<sup>||##</sup>, Ian A. Wilson<sup>||##2</sup>, and  Jeremy C. Mottram<sup>##53</sup>

From the <sup>‡</sup>Wellcome Trust Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, United Kingdom, the <sup>§</sup>Department of Biology, Centre for Immunology and Infection, University of York, Wentworth Way, Heslington, York YO10 5DD, United Kingdom, the <sup>||</sup>Joint Center for Structural Genomics, the <sup>||</sup>Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, the <sup>##</sup>Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, the <sup>||</sup>Center for Research in Biological Systems, University of California, San Diego, La Jolla, California, 92093, the <sup>\*\*</sup>Program on Bioinformatics and Systems Biology, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California 92037, the <sup>§§</sup>Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, and the <sup>||</sup>Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow G4 0RE, United Kingdom

Clan CD cysteine peptidases, a structurally related group of peptidases that include mammalian caspases, exhibit a wide range of important functions, along with a variety of specificities and activation mechanisms. However, for the clostripain family (denoted C11), little is currently known. Here, we describe the first crystal structure of a C11 protein from the human gut bacterium, *Parabacteroides merdae* (PmC11), determined to 1.7-Å resolution. PmC11 is a monomeric cysteine peptidase that comprises an extended caspase-like  $\alpha/\beta/\alpha$  sandwich and an unusual C-terminal domain. It shares core structural elements with clan CD cysteine peptidases but otherwise structurally differs from the other families in the clan. These studies also revealed a well ordered break in the polypeptide chain at Lys<sup>147</sup>, resulting in a large conformational rearrangement close to the active site. Biochemical and kinetic analysis revealed Lys<sup>147</sup> to be an intramolecular processing site at which cleavage is required for full activation of the enzyme, suggesting an autoinhibitory mechanism for self-preservation. PmC11 has an acidic binding pocket and a preference for basic substrates, and accepts substrates with Arg and Lys in P1 and does not require Ca<sup>2+</sup> for activity. Collectively, these data provide insights into the mechanism and activity of PmC11 and a detailed framework for studies on C11 peptidases from other phylogenetic kingdoms.

Cysteine peptidases play crucial roles in the virulence of bacterial and other eukaryotic pathogens. In the MEROPS peptidase database (1), clan CD contains groups (or families) of cysteine peptidases that share some highly conserved structural elements (2). Clan CD families are typically described using the name of their archetypal, or founding, member and also given an identification number preceded by a “C,” to denote cysteine peptidase. Although seven families (C14 is additionally split into three subfamilies) have been described for this clan, crystal structures have only been determined from four: legumain (C13) (3), caspase (C14a) (4), paracaspase (C14b(P)) (5), metacaspase (C14b(M)) (6), gingipain (C25) (7), and the cysteine peptidase domain (CPD) of various toxins (C80) (8). No structural information is available for clostripain (C11), separase (C50), or PrtH-peptidase (C85).

Clan CD enzymes have a highly conserved His/Cys catalytic dyad and exhibit strict specificity for the P<sub>1</sub> residue of their substrates. However, despite these similarities, clan CD forms a functionally diverse group of enzymes: the overall structural diversity between (and at times within) the various families provides these peptidases with a wide variety of substrate specificities and activation mechanisms. Several members are initially expressed as proenzymes, demonstrating self-inhibition prior to full activation (2).

The archetypal and arguably most notable family in the clan is that of the mammalian caspases (C14a), although clan CD members are distributed throughout the entire phylogenetic kingdom and are often required in fundamental biological processes (2). Interestingly, little is known about the structure or function of the C11 proteins, despite their widespread distribution (1) and its archetypal member, clostripain from *Clostridium histolyticum*, first reported in the literature in 1938 (9). Clostripain has been described as an arginine-specific peptidase with a requirement for Ca<sup>2+</sup> (10) and loss of an internal nonapeptide for full activation; lack of structural information on the family appears to have prohibited further investigation.

\* This work was supported by the Medical Research Council Grant MR/K019384, Wellcome Trust Grants 091790 and 104111, and National Institutes of Health Grant U54 GM094586 (JCSG). The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and National Institutes of Health (NIH), National Center for Research Resources, Biomedical Technology Program Grant P41RR001209, and the NIGMS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIGMS or NIH. The authors declare that they have no conflicts of interest with the contents of this article.

✂ Author's Choice—Final version free via Creative Commons CC-BY license. The atomic coordinates and structure factors (code 3UWS) have been deposited in the Protein Data Bank (<http://wwpdb.org/>).

<sup>1</sup> Both authors contributed equally to this article.

<sup>2</sup> To whom correspondence may be addressed. E-mail: wilson@scripps.edu.

<sup>3</sup> To whom correspondence may be addressed. E-mail: jeremy.mottram@york.ac.uk.

As part of an ongoing project to characterize commensal bacteria in the microbiome that inhabit the human gut, the structure of C11 peptidase, PmC11, from *Parabacteroides merdae* was determined using the Joint Center for Structural Genomics (JCSG)<sup>4</sup> HTP structural biology pipeline (11). The structure was analyzed, and the enzyme was biochemically characterized to provide the first structure/function correlation for a C11 peptidase.

## Experimental Procedures

Cloning, expression, purification, crystallization, and structure determination of PmC11 were carried out using standard JCSG protocols (11) as follows.

**Cloning**—Clones were generated using the polymerase incomplete primer extension (PIPE) cloning method (12). The gene encoding PmC11 (SP5111E) was amplified by polymerase chain reaction (PCR) from *P. merdae* genomic DNA using *Pfu-Turbo* DNA polymerase (Stratagene), using I-PIPE primers that included sequences for the predicted 5' and 3' ends (shown below). The expression vector, pSpeedET, which encodes an amino-terminal tobacco etch virus protease-cleavable expression and purification tag (MGSDKIHSHHHHENLYFQ/G), was PCR amplified with V-PIPE (Vector) primers. V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *Escherichia coli* GeneHogs (Invitrogen) competent cells were transformed with the I-PIPE/V-PIPE mixture and dispensed on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. The plasmid encoding the full-length protein was deposited in the PSI:Biological Materials Repository at the DNASU plasmid repository (PmCD00547516). For structure determination, to obtain soluble protein using the PIPE, method the gene segment encoding residues Met<sup>1</sup>-Asn<sup>22</sup> was deleted because these residues were predicted to correspond to a signal peptide using SignalP (13).

**Protein Expression and Selenomethionine Incorporation**—The expression plasmid for the truncated PmC11 construct was transformed into *E. coli* GeneHogs competent cells and grown in minimal media supplemented with selenomethionine and 30  $\mu\text{g ml}^{-1}$  of kanamycin at 37 °C using a GNF fermentor (14). A methionine auxotrophic strain was not required as selenomethionine is incorporated via the inhibition of methionine biosynthesis (15, 16). Protein expression was induced using 0.1% (w/v) L-arabinose and the cells were left to grow for a further 3 h at 37 °C. At the end of the cell culture, lysozyme was added to all samples to a final concentration of 250  $\mu\text{g ml}^{-1}$  and the cells were harvested and stored at  $-20$  °C, until required. Primers used in this section are as follows: I-PIPE (forward): CTGTACTTCCAGGCGAGACTCCGGAACCCCGGACA-ACCCGC; I-PIPE (reverse): AATTAAGTCGCGTTATTCAT-

AAACTGCCTTATACCAGCCGAC; V-PIPE (forward): TAA-CGCGACTTAATTAACCTCGTTTAAACGGTCTCCAGC; and V-PIPE (reverse): GCCCTGGAAGTACAGGTTTTCG-TGATGATGATGATGAT.

**Protein Purification for Crystallization**—Cells were resuspended, homogenized, and lysed by sonication in 40 mM Tris (pH 8.0), 300 mM NaCl, 10 mM imidazole, and 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (Lysis Buffer 1) containing 0.4 mM MgSO<sub>4</sub> and 1  $\mu\text{l}$  of 250 unit/ $\mu\text{l}^{-1}$  of benzonase (Sigma). The cell lysate was then clarified by centrifugation (32,500  $\times g$  for 25 min at 4 °C) before being passed over Ni<sup>2+</sup>-chelating resin equilibrated in Lysis Buffer 1 and washed in the same buffer supplemented with 40 mM imidazole and 10% (v/v) glycerol. The protein was subsequently eluted in 20 mM Tris (pH 8.0), 150 mM NaCl, 10% (v/v) glycerol, 1 mM TCEP, and 300 mM imidazole, and the fractions containing the protein were pooled.

To remove the His tag, PmC11 was exchanged into 20 mM Tris (pH 8.0), 150 mM NaCl, 30 mM imidazole, and 1 mM TCEP using a PD-10 column (GE Healthcare), followed by incubation with 1 mg of His-tagged tobacco etch virus protease per 15 mg of protein for 2 h at room temperature and subsequent overnight incubation at 4 °C. The sample was centrifuged to remove any precipitated material (13,000  $\times g$  for 10 min at 4 °C) and the supernatant loaded onto Ni<sup>2+</sup>-chelating resin equilibrated with 20 mM Tris (pH 8.0), 150 mM NaCl, 30 mM imidazole, and 1 mM TCEP and washed with the same buffer. The flow-through and wash fractions were collected and concentrated to 13.3 mg ml<sup>-1</sup> using Amicon Ultra-15 5K centrifugal concentrators (Millipore).

**Crystallization and Data Collection**—PmC11 was crystallized using the nanodroplet vapor diffusion method using standard JCSG crystallization protocols (11). Drops were comprised of 200 nl of protein solution mixed with 200 nl of crystallization solution in 96-well sitting-drop plates, equilibrated against a 50- $\mu\text{l}$  reservoir. Crystals of PmC11 were grown at 4 °C in mother liquor consisting of 0.2 M NH<sub>4</sub>H<sub>2</sub>PO<sub>4</sub>, 20% PEG-3350 (JCSG Core Suite I). Crystals were flash cooled in liquid nitrogen using 10% ethylene glycol as a cryoprotectant prior to data collection and initial screening for diffraction was carried out using the Stanford Automated Mounting system (17) at the Stanford Synchrotron Radiation Lightsources (SSRL, Menlo Park, CA). Single wavelength anomalous dispersion data were collected using a wavelength of 0.9793 Å, at the Advanced Light Source (ALS, beamline 8.2.2, Berkeley, CA) on an ADSC Quantum 315 CCD detector. The data were indexed and integrated with XDS (18) and scaled using XSCALE (18). The diffraction data were indexed in space group P2<sub>1</sub> with  $a = 39.11$ ,  $b = 108.68$ ,  $c = 77.97$  Å, and  $\beta = 94.32^\circ$ . The unit cell contained two molecules in the asymmetric unit resulting in a solvent content of 39% (Matthews' coefficient ( $V_m$ ) of 2.4 Å<sup>3</sup> Da<sup>-1</sup>).

**Structure Determination**—The PmC11 structure was determined by the single wavelength anomalous dispersion method using an x-ray wavelength corresponding to the peak of the selenium K edge. Initial phases were derived using the autoSHARP interface (19), which included density modification with SOLOMON (20). Good quality electron density was obtained at 1.7-Å resolution, allowing an initial model to be

<sup>4</sup> The abbreviations used are: JCSG, Joint Center for Structural Genomics; PIPE, polymerase incomplete primer extension; TCEP, Tris(2-carboxyethyl)phosphine; AMC, 7-amino-4-methylcoumarin; PDB, Protein Data Bank; Bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; Z, benzyloxycarbonyl; FMK, fluoromethyl ketone; CTD, C-terminal domain; Bz-R-AMC, benzoyl-L-Arg-4-methylcoumaryl-7-amide; Z-GGR-AMC, benzyloxycarbonyl-Gly-Gly-Arg-AMC; BOC-VLK-AMC, *t*-butyloxycarbonyl-Val-Leu-Lys.

**TABLE 1**

**Crystallographic statistics for PDB code 3UWS**

Values in parentheses are for the highest resolution shell.

Data collection	
Wavelength (Å)	0.9793
Space group	P2 <sub>1</sub>
Unit cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å); β°	39.11, 108.68, 77.97; β = 94.32°
Resolution range (Å)	28.73–1.70 (1.79–1.70)
Unique reflections	70,913
<i>R</i> <sub>merge</sub> <sup>a</sup> on <i>I</i> (%)	10.2 (49.0)
<i>R</i> <sub>meas</sub> <sup>b</sup> on <i>I</i> (%)	11.0 (52.7)
<i>R</i> <sub>pim</sub> <sup>c</sup> on <i>I</i> (%)	4.1 (19.2)
<i>I</i> / <i>σ<sub>i</sub></i>	15.6 (4.6)
Wilson B (Å <sup>2</sup> )	15.9
Completeness (%)	99.6 (99.8)
Multiplicity	7.3 (7.5)
Model and refinement	
Reflections (total/test)	70,883/3,577
<i>R</i> <sub>cryst</sub> / <i>R</i> <sub>free</sub> <sup>d</sup> (%)	14.3/17.5
No. protein residues/atoms	700/5612
No. of water/EDO molecules	690/7
ESU <sup>e</sup> based on <i>R</i> <sub>free</sub> (Å)	0.095
B-values (Å <sup>2</sup> )	
Average isotropic B (overall)	20.0
Protein overall	18.8
All main/side chains	16.7/20.8
Solvent/EDO	29.4/35.6
RMSD <sup>g</sup>	
Bond lengths (Å)	0.01
Bond angles (°)	1.6
Ramachandran analysis (%)	
Favored regions	97.0
Allowed regions	3.0
Outliers	0.0

$$^a R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

$$^b R_{\text{meas}} = \frac{\sum_{hkl} [N(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

$$^c R_{\text{pim}} (\text{precision-indicating } R_{\text{merge}}) = \frac{\sum_{hkl} [(1/(N-1))^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|]}{\sum_{hkl} \sum_i I_i(hkl)}$$

where *n* is the multiplicity of reflection *hkl*, and *I<sub>i</sub>*(*hkl*) and *I*(*hkl*) are the intensity of the *i*th measurement and the average intensity of reflection *hkl*, respectively (44).

$$^d R_{\text{cryst}} \text{ and } R_{\text{free}} = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$$

for reflections in the working and test sets, respectively, where *F<sub>obs</sub>* and *F<sub>calc</sub>* are the observed and calculated structure-factor amplitudes, respectively. *R<sub>free</sub>* is the same as *R<sub>cryst</sub>* but for 5% of the total reflections chosen at random and omitted from structural refinement.

<sup>e</sup> ESU is the estimated standard uncertainties of atoms.

<sup>f</sup> The average isotropic B includes TLS and residual B components.

<sup>g</sup> RMSD, root-mean-square deviation.

obtained by automated model building with ARP/wARP (21). Model completion and refinement were iteratively performed with Coot (22) and REFMAC (23, 24) to produce a final model with an *R*<sub>cryst</sub> and *R*<sub>free</sub> of 14.3 and 17.5%, respectively. The refinement included experimental phase restraints in the form of Hendrickson-Lattman coefficients, TLS refinement with one TLS group per molecule in the asymmetric unit, and NCS restraints. The refined structure contains residues 24–375 and 28–375 for the two molecules in the crystallographic asymmetric unit. Structural validation was carried using the JCSG Quality Control Server that analyzes both the coordinates and data using a variety of structural validation tools to confirm the stereochemical quality of the model (ADIT (25), MOLPROBITY (26), and WHATIF 5.0 (27)) and agreement between model and data (SGCHECK (28) and RESOLVE (29)). All of the main-chain torsion angles were in the allowed regions of the Ramachandran plot and the MolProbity overall clash score for the structure was 2.09 (within the 99th percentile for its resolution). The atomic coordinates and structure factors for PmC11 have been deposited in the Protein Data Bank (PDB) with the accession code 3UWS. Data collection, model, and refinement statistics are reported in Table 1.

**Structural Analysis**—The primary sequence alignment with assigned secondary structure was prepared using CLUSTAL

OMEGA (30) and ALINE (31). The topology diagram was produced with TOPDRAW (32) and all three-dimensional structural figures were prepared with PyMol (33) with the electrostatic surface potential calculated with APBS (18) and contoured at ±5 kT/e. Architectural comparisons with known structures revealed that PmC11 was most structurally similar to caspase-7, gingipain-K, and legumain (PDB codes 4hq0, 4tkx, and 4aw9, respectively). The statistical significance of the structural alignment between PmC11 and both caspase-7 and gingipain-K is equivalent (Z-score of 9.2) with legumain giving a very similar result (Z-score of 9.1). Of note, the β-strand topology of the CDP domains of *Clostridium difficile* toxin B (family C80; TcdB; PDB code 3pee) is identical to that observed in the PmC11 β-sheet, but the Z-score from DaliLite was notably less at 7.6. It is possible that the PmC11 structure is more closely related to the C80 family than other families in clan CD, and appear to reside on the same branch of the phylogenetic tree based on structure (2).

**Protein Production for Biochemical Assays**—The PmCD00547516 plasmid described above was obtained from the PSI:Biological Materials Repository and used to generate a cleavage site mutant *PmC11*<sup>K147A</sup> and an active-site mutant *PmC11*<sup>C179A</sup> using the QuikChange Site-directed Mutagenesis kit (Stratagene) as per the manufacturer's instructions using the following primers: K147A mutant (forward): CAGAATAAGCTGG-CAGCGTTCCGGACAGGACG, and K147A mutant (reverse): CGTCCTGTCCGAACGCTGCCAGCTTATTCTG; C179A mutant (forward): CCTGTTTCGATGCCGCTACATGGCA-AGC, and C179A mutant (reverse): GCTTGCCATGTAGGC-GGCATCGAACAGG. The expression plasmids containing *PmC11* were transformed into *E. coli* BL21 Star (DE3) and grown in Luria-Bertani media containing 30 μg ml<sup>-1</sup> of kanamycin at 37 °C until an optical density (600 nm) of ~0.6 was reached. L-Arabinose was added to a final concentration of 0.2% (w/v) and the cells incubated overnight at 25 °C.

Compared with the protein production for crystallography, a slightly modified purification protocol was employed for biochemical assays. Initially, the cells were resuspended in 20 mM sodium phosphate (pH 7.5), 150 mM NaCl (Lysis Buffer 2) containing an EDTA-free protease inhibitor mixture (cComplete, Roche Applied Science). Cells were disrupted by three passages (15 KPSI) through a One-Shot cell disruptor (Constant Systems) followed by centrifugation at 20,000 × *g* for 20 min at 4 °C. The supernatant was collected and sterile-filtered (0.2 μm) before being applied to a 5-ml HisTrap HP column (GE Healthcare) equilibrated in Lysis Buffer 2 containing 25 mM imidazole, and the protein was eluted in the same buffer containing 250 mM imidazole. The peak fractions were pooled and buffer exchanged into the assay buffer (20 mM Tris, 150 mM NaCl, pH 8.0) using a PD-10 column. When required, purified PmC11 was concentrated using Vivaspin 2 30-K centrifugal concentrators (Sartorius). Protein concentration was routinely measured using Bradford's reagent (Bio-Rad) with a BSA standard.

**Fluorogenic Substrate Activity Assays**—The release of the fluorescent group AMC (7-amino-4-methylcoumarin) from potential peptide substrates was used to assess the activity of PmC11. Peptidase activity was tested using 20 μg of PmC11 and

100  $\mu\text{M}$  substrate (unless otherwise stated) in assay buffer to a final reaction volume of 200  $\mu\text{l}$  and all samples were incubated (without substrate) at 37  $^{\circ}\text{C}$  for 16 h prior to carrying out the assay. The substrate and plate reader were brought to 37  $^{\circ}\text{C}$  for 20 min prior to the addition of the PmC11 and samples prepared without PmC11 were used as blanks (negative controls). The curves were plotted using the blank-corrected fluorescence units against the time of acquisition (in min). The assays were carried out in black 96-well flat-bottomed plates (Greiner). AMC fluorescence was measured using a PHERAstar FS plate reader (BMG Labtech) with excitation and emission wavelengths of 355 and 460 nm, respectively.

To investigate the substrate specificity of PmC11, substrates Z-GGR-AMC, Bz-R-AMC, Z-GP-AMC, Z-HGP-AMC, Ac-DEVD-AMC (all Bachem), BOC-VLK-AMC, and BOC-K-AMC (both PeptaNova) were prepared at 100 mM in 100% dimethyl sulfoxide. The amount of AMC (micromoles) released was calculated by generating an AMC standard curve (as described in Ref. 34) and the specific activity of PmC11 was calculated as picomoles of AMC released per min per mg of the protein preparation.

The reaction rates ( $V_{\text{max}}$ ) and  $K_m$  values were determined for mutants PmC11<sup>K147A</sup> and PmC11<sup>C179A</sup> by carrying out the activity assay at varying concentrations of Bz-R-AMC between 0 and 600  $\mu\text{M}$ . The blank-corrected relative fluorescence units were plotted against time (min) with  $\Delta\text{FU}/T$  giving the reaction rate. The  $K_m$  and  $V_{\text{max}}$  of PmC11 and PmC11<sup>K147A</sup> against an R-AMC substrate were determined from the Lineweaver-Burk plot as described (34), calculated using GraphPad Prism6 software. All experiments were carried out in triplicate.

**Effect of VRPR-FMK on PmC11**—To test the effect of the inhibitor on the activity of PmC11, 25  $\mu\text{M}$  Z-VRPR-FMK (100 mM stock in 100% dimethyl sulfoxide, Enzo Life Sciences), 20  $\mu\text{g}$  of PmC11, 100  $\mu\text{M}$  R-AMC substrate, 1 mM EGTA were prepared in the assay buffer and the activity assay carried out as described above. A gel-shift assay, to observe Z-VRPR-FMK binding to PmC11, was also set up using 20  $\mu\text{g}$  of PmC11, 25  $\mu\text{M}$  inhibitor, 1 mM EGTA in assay buffer. The reactions were incubated at 37  $^{\circ}\text{C}$  for 20 min before being stopped by the addition SDS-PAGE sample buffer. Samples were analyzed by loading the reaction mixture on a 10% NuPAGE BisTris gel using MES buffer.

**Effect of Cations on PmC11**—The enzyme activity of PmC11 was tested in the presence of various divalent cations:  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Zn}^{2+}$ , and  $\text{Cu}^{2+}$ . The final concentration of the salts ( $\text{MgSO}_4$ ,  $\text{CaCl}_2$ ,  $\text{MnCl}_2$ ,  $\text{CoCl}_2$ ,  $\text{FeSO}_4$ ,  $\text{ZnCl}_2$ , and  $\text{CuSO}_4$ ) was 1 mM and the control was set up without divalent ions but with addition of 1 mM EGTA. The assay was set up using 20 mg of PmC11, 1 mM salts, 100  $\mu\text{M}$  R-AMC substrate, and the assay buffer, and incubated at 37  $^{\circ}\text{C}$  for 16 h. The activity assay was carried out as described above.

**Size Exclusion Chromatography**—Affinity-purified PmC11 was loaded onto a HiLoad 16/60 Superdex 200 gel filtration column (GE Healthcare) equilibrated in the assay buffer. The apparent molecular weight of PmC11 was determined from calibration curves based on protein standards of known molecular weights.

**Autoprocessing Profile of PmC11**—Autoprocessing of PmC11 was evaluated by incubating the enzyme at 37  $^{\circ}\text{C}$  and removing samples at 1-h intervals from 0 to 16 h and placing into SDS-PAGE loading buffer to stop the processing. Samples were then analyzed on a 4–12% NuPAGE (Thermo Fisher) Novex BisTris gel run in MES buffer.

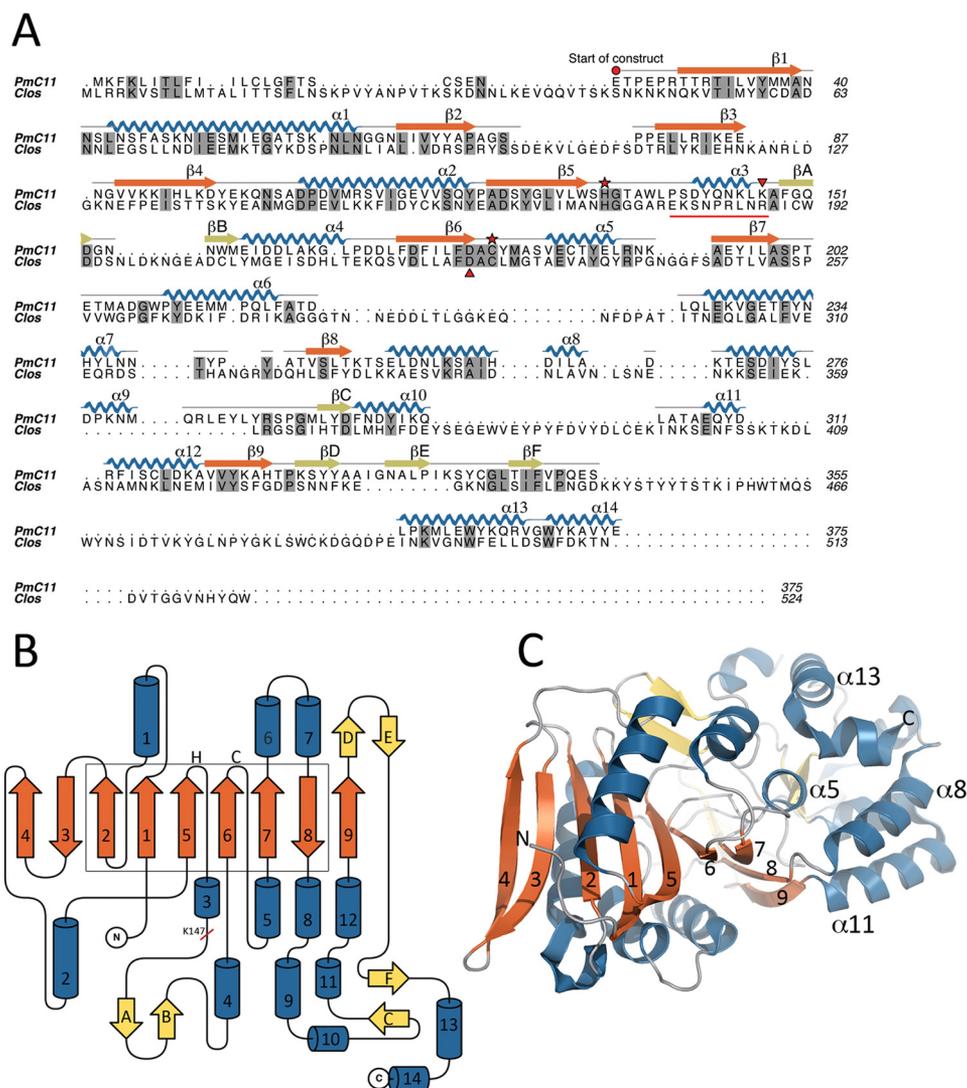
**Autoprocessing Cleavage Site Analysis**—To investigate whether processing is a result of intra- or inter-molecular cleavage, the PmC11<sup>C179A</sup> mutant was incubated with increasing concentrations of activated PmC11 (0, 0.1, 0.2, 0.5, 1, 2, and 5  $\mu\text{g}$ ). The final assay volume was 40  $\mu\text{l}$  and the proteins were incubated at 37  $^{\circ}\text{C}$  for 16 h in the PmC11 assay buffer. To stop the reaction, NuPAGE sample buffer was added to the protein samples and 20  $\mu\text{l}$  was analyzed on 10% NuPAGE Novex BisTris gel using MES buffer. These studies revealed no apparent cleavage of PmC11<sup>C179A</sup> by the active enzyme at low concentrations of PmC11 and that only limited cleavage was observed when the ratio of active enzyme (PmC11: PmC11<sup>C179A</sup>) was increased to ~1:10 and 1:4.

## Results

**Structure of PmC11**—The crystal structure of the catalytically active form of PmC11 revealed an extended caspase-like  $\alpha/\beta/\alpha$  sandwich architecture comprised of a central nine-stranded  $\beta$ -sheet, with an unusual C-terminal domain (CTD), starting at Lys<sup>250</sup>. A single cleavage was observed in the polypeptide chain at Lys<sup>147</sup> (Fig. 1, A and B), where both ends of the cleavage site are fully visible and well ordered in the electron density. The central nine-stranded  $\beta$ -sheet ( $\beta 1$ – $\beta 9$ ) of PmC11 consists of six parallel and three anti-parallel  $\beta$ -strands with 4  $\uparrow$  3  $\downarrow$  2  $\uparrow$  1  $\uparrow$  5  $\uparrow$  6  $\uparrow$  7  $\downarrow$  8  $\downarrow$  9  $\uparrow$  topology (Fig. 1A) and the overall structure includes 14  $\alpha$ -helices with six ( $\alpha 1$ – $\alpha 2$  and  $\alpha 4$ – $\alpha 7$ ) closely surrounding the  $\beta$ -sheet in an approximately parallel orientation. Helices  $\alpha 1$ ,  $\alpha 7$ , and  $\alpha 6$  are located on one side of the  $\beta$ -sheet with  $\alpha 2$ ,  $\alpha 4$ , and  $\alpha 5$  on the opposite side (Fig. 1A). Helix  $\alpha 3$  sits at the end of the loop following  $\beta 5$  (L5), just preceding the Lys<sup>147</sup> cleavage site, with both L5 and  $\alpha 3$  pointing away from the central  $\beta$ -sheet and toward the CTD, which starts with  $\alpha 8$ . The structure also includes two short  $\beta$ -hairpins ( $\beta A$ – $\beta B$  and  $\beta D$ – $\beta E$ ) and a small  $\beta$ -sheet ( $\beta C$ – $\beta F$ ), which is formed from two distinct regions of the sequence ( $\beta C$  precedes  $\alpha 11$ ,  $\alpha 12$  and  $\beta 9$ , whereas  $\beta F$  follows the  $\beta D$ – $\beta E$  hairpin) in the middle of the CTD (Fig. 1B).

The CTD of PmC11 is composed of a tight helical bundle formed from helices  $\alpha 8$ – $\alpha 14$  and includes strands  $\beta C$  and  $\beta F$ , and  $\beta$ -hairpin  $\beta D$ – $\beta E$ . The CTD sits entirely on one side of the enzyme interacting only with  $\alpha 3$ ,  $\alpha 5$ ,  $\beta 9$ , and the loops surrounding  $\beta 8$ . Of the interacting secondary structure elements,  $\alpha 5$  is perhaps the most interesting. This helix makes a total of eight hydrogen bonds with the CTD, including one salt bridge (Arg<sup>191</sup>-Asp<sup>255</sup>) and is surrounded by the CTD on one side and the main core of the enzyme on the other, acting like a linchpin holding both components together (Fig. 1C).

**Structural Comparisons**—PmC11 is, as expected, most structurally similar to other members of clan CD with the top hits in a search of known structures being caspase-7, gingipain-K, and legumain (PDB codes 4hq0, 4tkx, and 4aw9, respectively) (Table 2). The C-terminal domain is unique to PmC11 within



**FIGURE 1. Crystal structure of a C11 peptidase from *P. merdae*.** A, primary sequence alignment of PmC11 (Uniprot ID A7A9N3) and clostripain (Uniprot ID P09870) from *C. histolyticum* with identical residues highlighted in gray shading. The secondary structure of PmC11 from the crystal structure is mapped onto its sequence with the position of the PmC11 catalytic dyad, autocatalytic cleavage site (Lys<sup>147</sup>), and S<sub>1</sub> binding pocket Asp (Asp<sup>177</sup>) highlighted by a red star, a red downturned triangle, and a red upturned triangle, respectively. Connecting loops are colored gray, the main  $\beta$ -sheet is in orange, with other strands in olive,  $\alpha$ -helices are in blue, and the nonapeptide linker of clostripain that is excised upon autocleavage is underlined in red. Sequences around the catalytic site of clostripain and PmC11 align well. B, topology diagram of PmC11 colored as in A except that additional (non-core)  $\beta$ -strands are in yellow. Helices found on either side of the central  $\beta$ -sheet are shown above and below the sheet, respectively. The position of the catalytic dyad (H, C) and the processing site (Lys<sup>147</sup>) are highlighted. Helices (1–14) and  $\beta$ -strands (1–9 and A–F) are numbered from the N terminus. The core caspase-fold is highlighted in a box. C, tertiary structure of PmC11. The N and C termini (N and C) of PmC11 along with the central  $\beta$ -sheet (1–9), helix  $\alpha$ 5, and helices  $\alpha$ 8,  $\alpha$ 11, and  $\alpha$ 13 from the C-terminal domain, are all labeled. Loops are colored gray, the main  $\beta$ -sheet is in orange, with other  $\beta$ -strands in yellow, and  $\alpha$ -helices are in blue.

**TABLE 2**

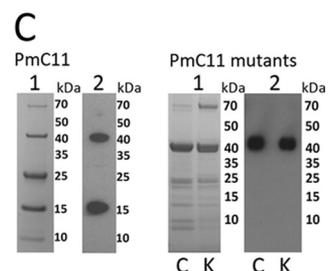
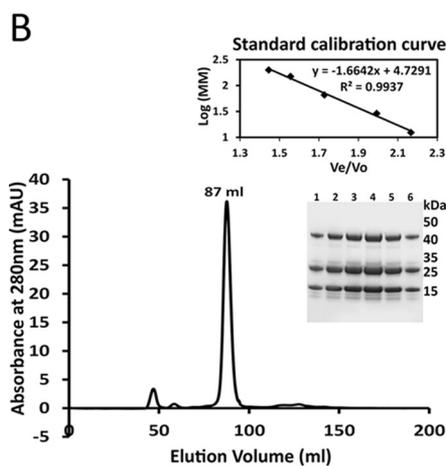
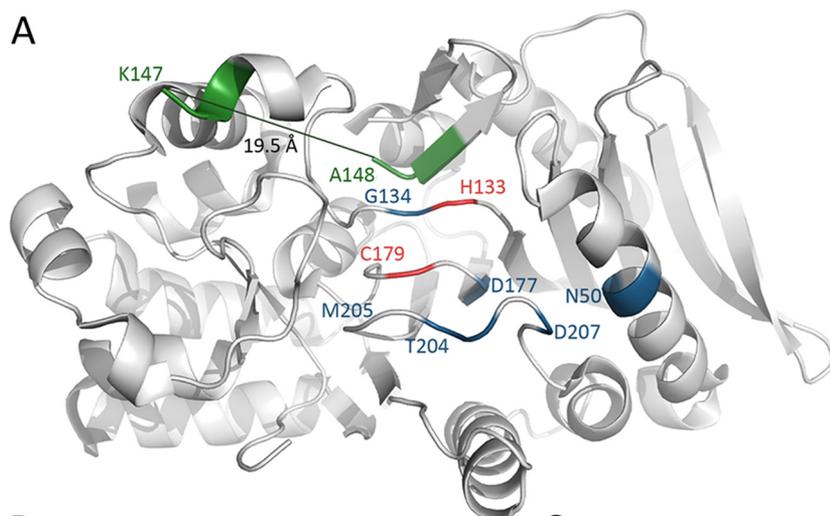
**Summary of PDBeFOLD (45) superposition of structures found to be most similar to PmC11 in the PDB based on DaliLite (46)**

The results are ordered in terms of structural homology ( $Q_H$ ), where %SSE<sup>PC-X</sup> is the percentage of the SSEs in the PmC11 that can be identified in the target X (where X = caspase-7 (47), legumain (3), gingipain (48), and TcdB-CPD (49)); %SSE<sup>X-PC</sup> is the percentage of SSEs in X (as above) that can be identified in PmC11 (as above); % sequence ID is the percentage sequence identity after structural alignment;  $N_{align}$  is the number of matched residues; and r.m.s. deviation the root mean squared deviation on the C $\alpha$  positions of the matched residues.

Enzyme	Family	PDB code	$Q^H$	Z-score	%SSE <sup>PC-X</sup>	%SSE <sup>X-PC</sup>	% Seq. ID	$N_{align}$	RMSD (Å)	$N_{Strands}$
PmC11	C11	3UWS	1.00	33.4	100	100	100	352	0.00	9
Caspase-7	C14A	4HQ0	0.16	4.3	38	79	14	162	3.27	6
Legumain	C13	4AW9	0.13	5.5	31	53	13	161	2.05	6
TcdB-CPD	C80	3PEE	0.10	4.9	28	50	12	138	3.18	9
Gingipain	C25	4TKX	0.07	5.4	28	27	12	153	2.97	10

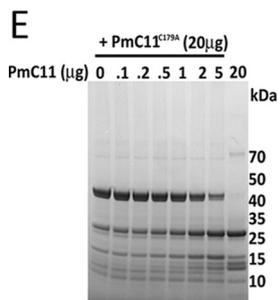
clan CD and structure comparisons for this domain alone does not produce any hits in the PDB (DaliLite, PDBeFold), suggesting a completely novel fold. As the archetypal and arguably most well studied member of clan CD, the caspases were used

as the basis to investigate the structure/function relationships in PmC11, with caspase-7 as the representative member. Six of the central  $\beta$ -strands in PmC11 ( $\beta$ 1– $\beta$ 2 and  $\beta$ 5– $\beta$ 8) share the same topology as the six-stranded  $\beta$ -sheet found in caspases,



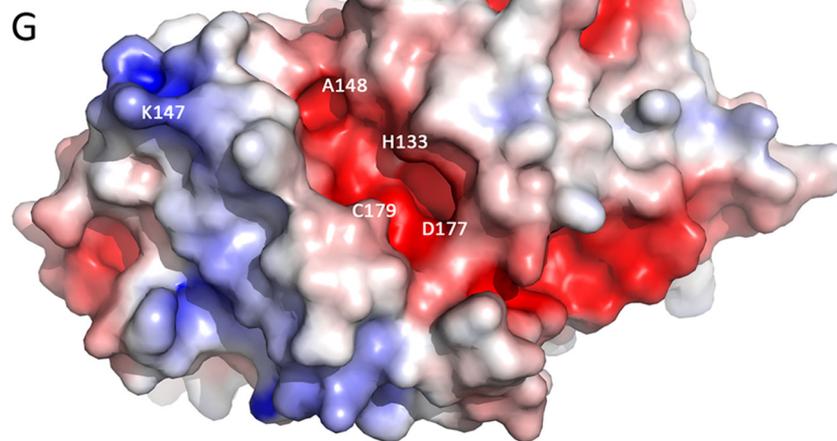
**D**

PmC11 <sup>K147A</sup>	
V <sub>max</sub> (FU sec <sup>-1</sup> )	323 +/- 124
K <sub>m</sub> (μM)	301 +/- 239
PmC11	
V <sub>max</sub> (FU sec <sup>-1</sup> )	4312 +/- 43
K <sub>m</sub> (μM)	274 +/- 6



**F**

Substrate	Specific activity (pmol mg <sup>-1</sup> min <sup>-1</sup> ) +/- S.D.
Z-R-AMC	1587 +/- 182
Z-GGR-AMC	540 +/- 145
BOC-VLK-AMC	560 +/- 55



## P. merdae C11 Cysteine Peptidase

with strands  $\beta 3$ ,  $\beta 4$ , and  $\beta 9$  located on the outside of this core structure (2) (Fig. 1B, box). His<sup>133</sup> and Cys<sup>179</sup> were found at locations structurally homologous to the caspase catalytic dyad, and other clan CD structures (2), at the C termini of strands  $\beta 5$  and  $\beta 6$ , respectively (Figs. 1, A and B, and 2A). A multiple sequence alignment of C11 proteins revealed that these residues are highly conserved (data not shown).

Five of the  $\alpha$ -helices surrounding the  $\beta$ -sheet of PmC11 ( $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 4$ ,  $\alpha 6$ , and  $\alpha 7$ ) are found in similar positions to the five structurally conserved helices in caspases and other members of clan CD, apart from family C80 (2). Other than its more extended  $\beta$ -sheet, PmC11 differs most significantly from other clan CD members at its C terminus, where the CTD contains a further seven  $\alpha$ -helices and four  $\beta$ -strands after  $\beta 8$ .

**Autoprocessing of PmC11**—Purification of recombinant PmC11 (molecular mass = 42.6 kDa) revealed partial processing into two cleavage products of 26.4 and 16.2 kDa, related to the observed cleavage at Lys<sup>147</sup> in the crystal structure (Fig. 2A). Incubation of PmC11 at 37 °C for 16 h, resulted in a fully processed enzyme that remained as an intact monomer when applied to a size-exclusion column (Fig. 2B). The single cleavage site of PmC11 at Lys<sup>147</sup> is found immediately after  $\alpha 3$ , in loop L5 within the central  $\beta$ -sheet (Figs. 1, A and B, and 2A). The two ends of the cleavage site are remarkably well ordered in the crystal structure and displaced from one another by 19.5 Å (Fig. 2A). Moreover, the C-terminal side of the cleavage site resides near the catalytic dyad with Ala<sup>148</sup> being 4.5 and 5.7 Å from His<sup>133</sup> and Cys<sup>179</sup>, respectively. Consequently, it appears feasible that the helix attached to Lys<sup>147</sup> ( $\alpha 3$ ) could be responsible for steric autoinhibition of PmC11 when Lys<sup>147</sup> is covalently bonded to Ala<sup>148</sup>. Thus, the cleavage would be required for full activation of PmC11. To investigate this possibility, two mutant forms of the enzyme were created: PmC11<sup>C179A</sup> (a catalytically inactive mutant) and PmC11<sup>K147A</sup> (a cleavage-site mutant). Initial SDS-PAGE and Western blot analysis of both mutants revealed no discernible processing occurred as compared with active PmC11 (Fig. 2C). The PmC11<sup>K147A</sup> mutant enzyme had a markedly different reaction rate ( $V_{\max}$ ) compared with WT, where the reaction velocity of PmC11 was 10 times greater than that of PmC11<sup>K147A</sup> (Fig. 2D). Taken together, these data reveal that PmC11 requires processing at Lys<sup>147</sup> for optimum activity.

To investigate whether processing is a result of intra- or intermolecular cleavage, the PmC11<sup>C179A</sup> mutant was incubated with increasing concentrations of processed and acti-

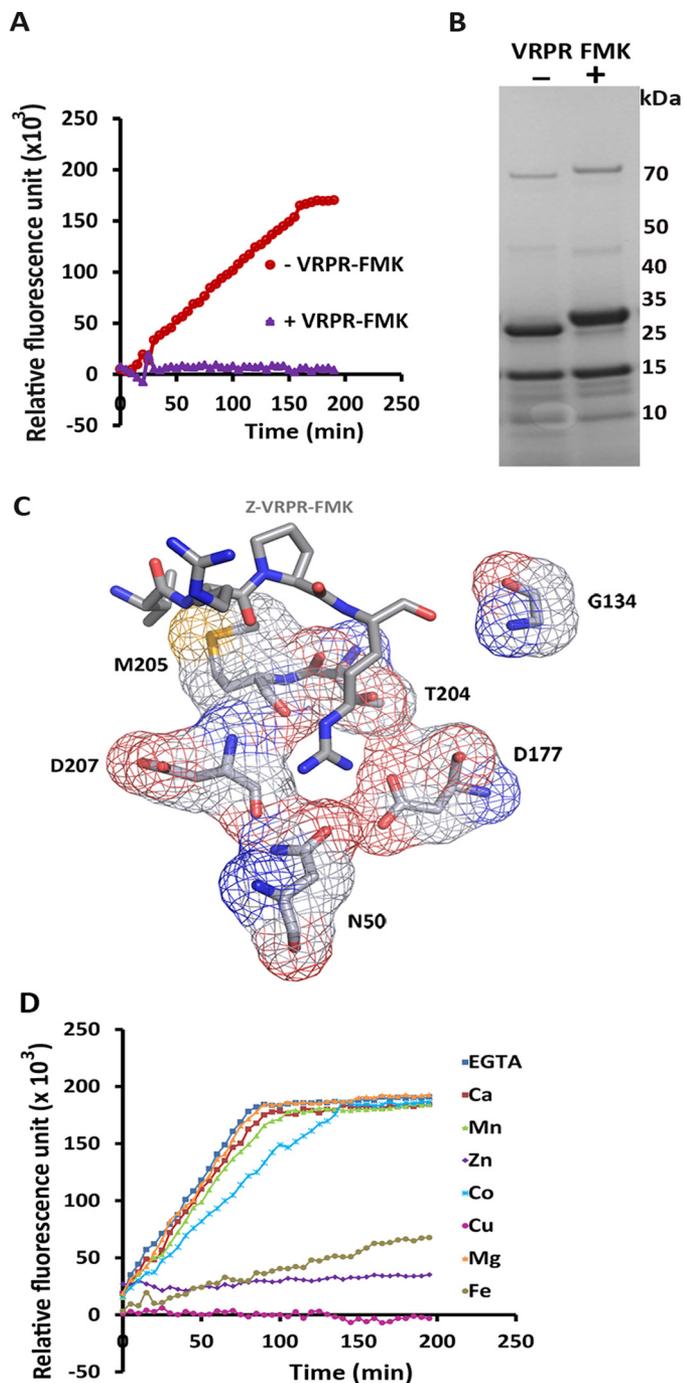
vated PmC11. These studies revealed that there was no apparent cleavage of PmC11<sup>C179A</sup> by the active enzyme at low concentrations of PmC11 and that only limited cleavage was observed when the ratio of active enzyme (PmC11: PmC11<sup>C179A</sup>) was increased to ~1:10 and 1:4, with complete cleavage observed at a ratio of 1:1 (Fig. 2E). This suggests that cleavage of PmC11<sup>C179A</sup> was most likely an effect of the increasing concentration of PmC11 and intermolecular cleavage. Collectively, these data suggest that the pro-form of PmC11 is autoinhibited by a section of L5 blocking access to the active site, prior to intramolecular cleavage at Lys<sup>147</sup>. This cleavage subsequently allows movement of the region containing Lys<sup>147</sup> and the active site to open up for substrate access.

**Substrate Specificity of PmC11**—The autocatalytic cleavage of PmC11 at Lys<sup>147</sup> (sequence KLKΛA) demonstrates that the enzyme accepts substrates with Lys in the P<sub>1</sub> position. The substrate specificity of the enzyme was further tested using a variety of fluorogenic substrates. As expected, PmC11 showed no activity against substrates with Pro or Asp in P<sub>1</sub> but was active toward substrates with a basic residue in P<sub>1</sub> such as Bz-R-AMC, Z-GGR-AMC, and BOC-VLK-AMC. The rate of cleavage was ~3-fold greater toward the single Arg substrate Bz-R-AMC than for the other two (Fig. 2F) and, unexpectedly, PmC11 showed no activity toward BOC-K-AMC. These results confirm that PmC11 accepts substrates containing Arg or Lys in P<sub>1</sub> with a possible preference for Arg.

The catalytic dyad of PmC11 sits near the bottom of an open pocket on the surface of the enzyme at a conserved location in the clan CD family (2). The PmC11 structure reveals that the catalytic dyad forms part of a large acidic pocket (Fig. 2G), consistent with a binding site for a basic substrate. This pocket is lined with the potential functional side chains of Asn<sup>50</sup>, Asp<sup>177</sup>, and Thr<sup>204</sup> with Gly<sup>134</sup>, Asp<sup>207</sup>, and Met<sup>205</sup> also contributing to the pocket (Fig. 2A). Interestingly, these residues are in regions that are structurally similar to those involved in the S<sub>1</sub> binding pockets of other clan CD members (shown in Ref. 2).

Because PmC11 recognizes basic substrates, the tetrapeptide inhibitor Z-VRPR-FMK was tested as an enzyme inhibitor and was found to inhibit both the autoprocessing and activity of PmC11 (Fig. 3A). Z-VRPR-FMK was also shown to bind to the enzyme: a size-shift was observed, by SDS-PAGE analysis, in the larger processed product of PmC11 suggesting that the inhibitor bound to the active site (Fig. 3B). A structure overlay of PmC11 with the MALT1-paracaspase (MALT1-P), in complex with Z-VRPR-FMK (35), revealed that the PmC11

**FIGURE 2. Biochemical and structural characterization of PmC11.** A, ribbon representation of the overall structure of PmC11 illustrating the catalytic site, cleavage site displacement, and potential S<sub>1</sub> binding site. The overall structure of PmC11 is shown in gray, looking down into the catalytic site with the catalytic dyad in red. The two ends of the autolytic cleavage site (Lys<sup>147</sup> and Ala<sup>148</sup>, green) are displaced by 19.5 Å (thin black line) from one another and residues in the potential substrate binding pocket are highlighted in blue. B, size exclusion chromatography of PmC11. PmC11 migrates as a monomer with a molecular mass around 41 kDa calculated from protein standards of known molecular weights. Elution fractions across the major peak (1–6) were analyzed by SDS-PAGE on a 4–12% gel in MES buffer. C, the active form of PmC11 and two mutants, PmC11<sup>C179A</sup> (C) and PmC11<sup>K147A</sup> (K), were examined by SDS-PAGE (lane 1) and Western blot analysis using an anti-His antibody (lane 2) show that PmC11 autoprocesses, whereas mutants, PmC11<sup>C179A</sup> and PmC11<sup>K147A</sup>, do not show autoprocessing *in vitro*. D, cysteine peptidase activity of PmC11.  $K_m$  and  $V_{\max}$  of PmC11 and K147A mutant were determined by monitoring change in the fluorescence corresponding to AMC release from Bz-R-AMC. Reactions were performed in triplicate and representative values  $\pm$  S.D. are shown. E, intermolecular processing of PmC11<sup>C179A</sup> by PmC11. PmC11<sup>C179A</sup> (20  $\mu$ g) was incubated overnight at 37 °C with increasing amounts of processed PmC11 and analyzed on a 10% SDS-PAGE gel. Inactive PmC11<sup>C179A</sup> was not processed to a major extent by active PmC11 until around a ratio of 1:4 (5  $\mu$ g of active PmC11). A single lane of 20  $\mu$ g of active PmC11 (labeled 20) is shown for comparison. F, activity of PmC11 against basic substrates. Specific activity is shown  $\pm$  S.D. from three independent experiments. G, electrostatic surface potential of PmC11 shown in a similar orientation, where blue and red denote positively and negatively charged surface potential, respectively, contoured at  $\pm 5$  kT/e. The position of the catalytic dyad, one potential key substrate binding residue Asp<sup>177</sup>, and the ends of the cleavage site Lys<sup>147</sup> and Ala<sup>148</sup> are indicated.



**FIGURE 3. PmC11 binds and is inhibited by Z-VRPR-FMK and does not require  $\text{Ca}^{2+}$  for activity.** *A*, PmC11 activity is inhibited by Z-VRPR-FMK. Cleavage of Bz-R-AMC by PmC11 was measured in a fluorometric activity assay with (+, purple) and without (-, red) Z-VRPR-FMK. The relative fluorescence units of AMC released are plotted against time (min) ( $n = 3$ ;  $\pm$ S.D.). *B*, gel-shift assay reveals that Z-VRPR-FMK binds to PmC11. PmC11 was incubated with (+) or without (-) Z-VRPR-FMK and the samples analyzed on a 10% SDS-PAGE gel. A size shift can be observed in the larger processed product of PmC11 (26.1 kDa). *C*, PmC11 with the Z-VRPR-FMK from the MALT1-paracaspase (MALT1-P) superimposed. A three-dimensional structural overlay of Z-VRPR-FMK from the MALT1-P complex onto PmC11. The position and orientation of Z-VRPR-FMK was taken from superposition of the PmC11 and MALT1\_P structures and indicates the presumed active site of PmC11. Residues surrounding the inhibitor are labeled and represent potentially important binding site residues, labeled in *black* and shown in an atomic representation. Carbon atoms are shown in *gray*, nitrogen in *blue*, and oxygen in *red*. *D*, divalent cations do not increase the activity of PmC11. The cleavage of Bz-R-AMC by PmC11 was measured in the presence of the cations  $\text{Ca}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Cu}^{2+}$ ,  $\text{Mg}^{2+}$ , and  $\text{Fe}^{3+}$  with EGTA as a negative control, and

dyad sits in a very similar position to that of active MALT1-P and that Asn<sup>50</sup>, Asp<sup>177</sup>, and Asp<sup>207</sup> superimpose well with the principal MALT1-P inhibitor binding residues (Asp<sup>365</sup>, Asp<sup>462</sup>, and Glu<sup>500</sup>, respectively (VRPR-FMK from MALT1-P with the corresponding PmC11 residues from the structural overlay is shown in Fig. 1D), as described in Ref. 5). Asp<sup>177</sup> is located near the catalytic cysteine and is conserved throughout the C11 family, suggesting it is the primary  $S_1$  binding site residue. In the structure of PmC11, Asp<sup>207</sup> resides on a flexible loop pointing away from the  $S_1$  binding pocket (Fig. 3C). However, this loop has been shown to be important for substrate binding in clan CD (2) and this residue could easily rotate and be involved in substrate binding in PmC11. Thus, Asn<sup>50</sup>, Asp<sup>177</sup>, and Asp<sup>207</sup> are most likely responsible for the substrate specificity of PmC11. Asp<sup>177</sup> is highly conserved throughout the clan CD C11 peptidases and is thought to be primarily responsible for substrate specificity of the clan CD enzymes, as also illustrated from the proximity of these residues relative to the inhibitor Z-VRPR-FMK when PmC11 is overlaid on the MALT1-P structure (Fig. 3C).

**Comparison with Clostripain**—Clostripain from *C. histolyticum* is the founding member of the C11 family of peptidases and contains an additional 149 residues compared with PmC11. A multiple sequence alignment revealed that most of the secondary structural elements are conserved between the two enzymes, although they are only ~23% identical (Fig. 1A). Nevertheless, PmC11 may be a good model for the core structure of clostripain.

The primary structural alignment also shows that the catalytic dyad in PmC11 is structurally conserved in clostripain (36) (Fig. 1A). Unlike PmC11, clostripain has two cleavage sites (Arg<sup>181</sup> and Arg<sup>190</sup>), which results in the removal of a nonapeptide, and is required for full activation of the enzyme (37) (highlighted in Fig. 1A). Interestingly, Arg<sup>190</sup> was found to align with Lys<sup>147</sup> in PmC11. In addition, the predicted primary  $S_1$ -binding residue in PmC11 Asp<sup>177</sup> also overlays with the residue predicted to be the P1 specificity determining residue in clostripain (38) (Asp<sup>229</sup>, Fig. 1A).

As studies on clostripain revealed addition of  $\text{Ca}^{2+}$  ions are required for full activation, the  $\text{Ca}^{2+}$  dependence of PmC11 was examined. Surprisingly,  $\text{Ca}^{2+}$  did not enhance PmC11 activity and, furthermore, other divalent cations,  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Zn}^{2+}$ , and  $\text{Cu}^{2+}$ , were not necessary for PmC11 activity (Fig. 3D). In support of these findings, EGTA did not inhibit PmC11 suggesting that, unlike clostripain, PmC11 does not require  $\text{Ca}^{2+}$  or other divalent cations, for activity.

## Discussion

The crystal structure of PmC11 now provides three-dimensional information for a member of the clostripain C11 family of cysteine peptidases. The enzyme exhibits all of the key structural elements of clan CD members, but is unusual in that it has a nine-stranded central  $\beta$ -sheet with a novel C-terminal

relative fluorescence measured against time (min). The addition of cations produced no improvement in activity of PmC11 when compared in the presence of EGTA, suggesting that PmC11 does not require metal ions for proteolytic activity. Furthermore,  $\text{Cu}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Zn}^{2+}$  appear to inhibit PmC11.

## *P. merdae* C11 Cysteine Peptidase

domain. The structural similarity of PmC11 with its nearest structural neighbors in the PDB is decidedly low, overlaying better with six-stranded caspase-7 than any of the other larger members of the clan (Table 2). The substrate specificity of PmC11 is Arg/Lys and the crystal structure revealed an acidic pocket for specific binding of such basic substrates. In addition, the structure suggested a mechanism of self-inhibition in both PmC11 and clostripain and an activation mechanism that requires autoprocessing. PmC11 differs from clostripain in that it does not appear to require divalent cations for activation.

Several other members of clan CD require processing for full activation including legumain (39), gingipain-R (40), MARTX-CPD (8), and the effector caspases, *e.g.* caspase-7 (41). To date, the effector caspases are the only group of enzymes that require cleavage of a loop within the central  $\beta$ -sheet. This is also the case in PmC11, although the cleavage loop is structurally different to that found in the caspases and follows the catalytic His (Fig. 1A), as opposed to the Cys in the caspases.

All other clan CD members requiring cleavage for full activation do so at sites external to their central sheets (2). The caspases and gingipain-R both undergo intermolecular (*trans*) cleavage and legumain and MARTX-CPD are reported to perform intramolecular (*cis*) cleavage. In addition, several members of clan CD exhibit self-inhibition, whereby regions of the enzyme block access to the active site (2). Like PmC11, these structures show preformed catalytic machinery and, for a substrate to gain access, movement and/or cleavage of the blocking region is required.

The structure of PmC11 gives the first insight into this class of relatively unexplored family of proteins and should allow important catalytic and substrate binding residues to be identified in a variety of orthologues. Indeed, insights gained from an analysis of the PmC11 structure revealed the identity of the *Trypanosoma brucei* PNT1 protein as a C11 cysteine peptidase with an essential role in organelle replication (42). The PmC11 structure should provide a good basis for structural modeling and, given the importance of other clan CD enzymes, this work should also advance the exploration of these peptidases and potentially identify new biologically important substrates.

---

**Author Contributions**—K. M., J. S. G., D. D., I. A. W., and J. C. M. designed the research; K. M., J. S. G., and D. D. performed the research; K. M., J. S. G., D. D., G. H. C., A. S., M. A. E., and J. C. M. analyzed the data; A. G., S. A. L., A. M. D., M. A. E., and I. A. W. supervised various components of the JCSG structural genomics pipeline; M. K. G., A. G., S. A. L., A. M. D., and M. A. E. contributed reagents, materials, and analysis tools; and K. M., J. S. G., G. H. C., M. A. E., I. A. W., and J. C. M. wrote the paper.

---

**Acknowledgments**—We thank all members of the JCSG high throughput Structural Biology pipeline for their contribution to this work. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the United States Department of Energy under Contract DE-AC02-05CH11231. Genomic DNA from *P. merdae* was extracted from cells (ATCC number 43184) obtained from the American Type Culture Collection (ATCC).

---

## References

1. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **40**, D343–350
2. McLuskey, K., and Mottram, J. C. (2015) Comparative structural analysis of the caspase family with other clan CD cysteine peptidases. *Biochem. J.* **466**, 219–232
3. Dall, E., and Brandstetter, H. (2013) Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10940–10945
4. Walker, N. P., Talanian, R. V., Brady, K. D., Dang, L. C., Bump, N. J., Ferenz, C. R., Franklin, S., Ghayur, T., Hackett, M. C., and Hammill, L. D. (1994) Crystal structure of the cysteine protease interleukin-1 $\beta$ -converting enzyme: a (p20/p10)<sub>2</sub> homodimer. *Cell* **78**, 343–352
5. Yu, J. W., Jeffrey, P. D., Ha, J. Y., Yang, X., and Shi, Y. (2011) Crystal structure of the mucosa-associated lymphoid tissue lymphoma translocation 1 (MALT1) paracaspase region. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 21004–21009
6. McLuskey, K., Rudolf, J., Proto, W. R., Isaacs, N. W., Coombs, G. H., Moss, C. X., and Mottram, J. C. (2012) Crystal structure of a *Trypanosoma brucei* metacaspase. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7469–7474
7. Eichinger, A., Beisel, H. G., Jacob, U., Huber, R., Medrano, F. J., Banbula, A., Potempa, J., Travis, J., and Bode, W. (1999) Crystal structure of gingipain R: an Arg-specific bacterial cysteine proteinase with a caspase-like fold. *EMBO J.* **18**, 5453–5462
8. Lupardus, P. J., Shen, A., Bogoy, M., and Garcia, K. C. (2008) Small molecule-induced allosteric activation of the *Vibrio cholerae* RTX cysteine protease domain. *Science* **322**, 265–268
9. Kocholaty, W., Weil, L., and Smith, L. (1938) Proteinase secretion and growth of *Clostridium histolyticum*. *Biochem. J.* **32**, 1685–1690
10. Kembhavi, A. A., Buttler, D. J., Rauber, P., and Barrett, A. J. (1991) Clostripain: characterization of the active site. *FEBS Lett.* **283**, 277–280
11. Elsliger, M. A., Deacon, A. M., Godzik, A., Lesley, S. A., Wooley, J., Wüthrich, K., and Wilson, I. A. (2010) The JCSG high-throughput structural biology pipeline. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1137–1142
12. Klock, H. E., and Lesley, S. A. (2009) The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol. Biol.* **498**, 91–103
13. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786
14. Lesley, S. A., Kuhn, P., Godzik, A., Deacon, A. M., Mathews, I., Kreuzsch, A., Spraggon, G., Klock, H. E., McMullan, D., Shin, T., Vincent, J., Robb, A., Brinen, L. S., Miller, M. D., McPhillips, T. M., Miller, M. A., Scheibe, D., Canaves, J. M., Guda, C., Jaroszewski, L., Selby, T. L., Elsliger, M. A., Wooley, J., Taylor, S. S., Hodgson, K. O., Wilson, I. A., Schultz, P. G., and Stevens, R. C. (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11664–11669
15. Doublé, S. (2007) Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol.* **363**, 91–108
16. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124
17. Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M., and Phizackerley, R. P. (2002) An automated system to mount cryo-cooled protein crystals on a synchrotron beam line, using compact sample cassettes and a small-scale robot. *J. Appl. Crystallogr.* **35**, 720–726
18. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
19. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230
20. Abrahams, J. P., and Leslie, A. G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 30–42
21. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for x-ray crystallography using ARP/

- WARP version 7. *Nat. Protoc.* **3**, 1171–1179
22. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
  23. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
  24. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367
  25. Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M., and Westbrook, J. D. (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1833–1839
  26. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
  27. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56
  28. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) SFHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 191–205
  29. Terwilliger, T. (2004) SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52
  30. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539
  31. Bond, C. S., and Schüttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 510–512
  32. Bond, C. S. (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311–312
  33. DeLano, W. (2002) *The PyMOL Molecular Graphic System*, version 1.2r3pre, Schrodinger, LLC, San Carlos, CA
  34. McLuskey, K., Moss, C. X., and Mottram, J. C. (2014) Purification, characterization, and crystallization of *Trypanosoma metacaspases*. *Methods Mol. Biol.* **1133**, 203–221
  35. Wiesmann, C., Leder, L., Blank, J., Bernardi, A., Melkko, S., Decock, A., D'Arcy, A., Villard, F., Erbel, P., Hughes, N., Freuler, F., Nikolay, R., Alves, J., Bornancin, F., and Rénatus, M. (2012) Structural determinants of MALT1 protease activity. *J. Mol. Biol.* **419**, 4–21
  36. Ullmann, D., and Jakubke, H. D. (1994) The specificity of clostripain from *Clostridium histolyticum*: mapping the S' subsites via acyl transfer to amino acid amides and peptides. *Eur. J. Biochem.* **223**, 865–872
  37. Witte, V., Wolf, N., and Dargatz, H. (1996) Clostripain linker deletion variants yield active enzyme in *Escherichia coli*: a possible function of the linker peptide as intramolecular inhibitor of clostripain automaturation. *Curr. Microbiol.* **33**, 281–286
  38. Labrou, N. E., and Rigden, D. J. (2004) The structure-function relationship in the clostripain family of peptidases. *Eur. J. Biochem.* **271**, 983–992
  39. Li, D. N., Matthews, S. P., Antoniou, A. N., Mazzeo, D., and Watts, C. (2003) Multistep autoactivation of asparaginyl endopeptidase *in vitro* and *in vivo*. *J. Biol. Chem.* **278**, 38980–38990
  40. Mikolajczyk, J., Boatright, K. M., Stennicke, H. R., Nazif, T., Potempa, J., Bogyo, M., and Salvesen, G. S. (2003) Sequential autolytic processing activates the zymogen of Arg-gingipain. *J. Biol. Chem.* **278**, 10458–10464
  41. Denault, J. B., and Salvesen, G. S. (2003) Human caspase-7 activity and regulation by its N-terminal peptide. *J. Biol. Chem.* **278**, 34042–34050
  42. Grewal, J. S., McLuskey, K., Das, D., Myburgh, E., Wilkes, J., Brown, E., Lemgruber, L., Gould, M. K., Burchmore, R. J., Coombs, G. H., Schnauffer, A., and Mottram, J. C. (2016) PNT1 is a C11 cysteine peptidase essential for replication of the Trypanosome kinetoplast. *J. Biol. Chem.* **291**, 9492–9500
  43. Weiss, M. S., Metzner, H. J., and Hilgenfeld, R. (1998) Two non-proline cis peptide bonds may be important for factor XIII function. *FEBS Lett.* **423**, 291–296
  44. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269–275
  45. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268
  46. Holm, L., and Park, J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**, 566–567
  47. Kang, H. J., Lee, Y. M., Bae, K. H., Kim, S. J., and Chung, S. J. (2013) Structural asymmetry of procaspase-7 bound to a specific inhibitor. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1514–1521
  48. Gorman, M. A., Seers, C. A., Michell, B. J., Feil, S. C., Huq, N. L., Cross, K. J., Reynolds, E. C., and Parker, M. W. (2015) Structure of the lysine specific protease Kgp from *Porphyromonas gingivalis*, a target for improved oral health. *Protein Sci.* **24**, 162–166
  49. Shen, A., Lupardus, P. J., Gersch, M. M., Puri, A. W., Albrow, V. E., Garcia, K. C., and Bogyo, M. (2011) Defining an allosteric circuit in the cysteine protease domain of *Clostridium difficile* toxins. *Nat. Struct. Mol. Biol.* **18**, 364–371

**Crystal Structure and Activity Studies of the C11 Cysteine Peptidase from  
*Parabacteroides merdae* in the Human Gut Microbiome**

Karen McLuskey, Jaspreet S. Grewal, Debanu Das, Adam Godzik, Scott A. Lesley,  
Ashley M. Deacon, Graham H. Coombs, Marc-André Elsliger, Ian A. Wilson and  
Jeremy C. Mottram

*J. Biol. Chem.* 2016, 291:9482-9491.

doi: 10.1074/jbc.M115.706143 originally published online March 3, 2016

---

Access the most updated version of this article at doi: [10.1074/jbc.M115.706143](https://doi.org/10.1074/jbc.M115.706143)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 48 references, 12 of which can be accessed free at  
<http://www.jbc.org/content/291/18/9482.full.html#ref-list-1>