View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by White Rose Research Online

This is a repository copy of *Crystal Structure and Activity Studies of the C11 Cysteine Peptidase from Parabacteroides merdae in the Human Gut Microbiome.*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/97254/

Version: Accepted Version

Article:
McLuskey, Karen, Grewal, Jaspreet, Das, Debanu et al. (7 more authors) (2016) Crystal Structure and Activity Studies of the C11 Cysteine Peptidase from Parabacteroides merdae in the Human Gut Microbiome. Journal of Biological Chemistry. pp. 9482-9491. ISSN 1083-351X

https://doi.org/10.1074/jbc.M115.706143

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

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

Karen McLuskey[1*], Jaspreet S. Grewal[1,2*], Debanu Das[3,4], Adam Godzik[3,6,7], Scott A. Lesley[3,5,8], Ashley M. Deacon[3,4], Graham H. Coombs[9], Marc-André Elsliger[3,5], Ian A. Wilson[3,5] and Jeremy C. Mottram[1,2]

[1]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, UK
[2] Centre for Immunology and Infection, Department of Biology, University of York, Wentworth Way, Heslington, York, YO10 5DD, UK
[3]Joint Center for Structural Genomics, http://www.jcsg.org
[4]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA94025, USA
[5] Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
[6]Center for Research in Biological Systems, University of California, San Diego, La Jolla, California 92093
[7]Program on Bioinformatics and Systems Biology, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California 92037
[8]Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California 92121
[9] Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, G4 0RE, UK

Running Head: *Parabacteroides merdae* C11 cysteine peptidase

Keywords: crystal structure, protease

Correspondence: wilson@scripps.edu; jeremy.mottram@york.ac.uk

* equal contribution

ABSTRACT

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 α/β/α 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 K147, resulting in a large conformational rearrangement close to the active site. Biochemical and kinetic analysis revealed K147 to be an intramolecular processing site at which cleavage is required for full activation of the enzyme, suggesting an auto-inhibitory mechanism for self-preservation. PmC11 has an acidic binding pocket and a preference for basic substrates, and accepts both Arg and Lys substrates and does not require Ca^{2+} 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.

INTRODUCTION

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. Whilst 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_1 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 wide spread distribution (1) and its archetypal member, clostripain from *Clostridium histolyticum,* being first reported in the literature in 1938 (9). Clostripain has been described as an arginine-specific peptidase with a requirement for Ca^{2+} (10) and loss of an internal nonapeptide for full activation; lack of structural information on the family appears to have prohibited further investigation.

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 JCSG HTP structural biology pipeline (11) and now has been analysed 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 (http://www.jcsg.org) 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 *Parabacteroides merdae* genomic DNA using *PfuTurbo* 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 (TEV) protease-cleavable expression and purification tag (MGSDKIHHHHHHENLYFQ/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:Biology Materials Repository at the DNASU plasmid repository (PmCD00547516). For structure determination, using the PIPE method for soluble protein expression, the gene segment encoding residues M1-N22 was deleted since these residues were predicted to correspond to a signal peptide using SignalP (13).

Protein Expression and Selenomethionine–incorporation (SeMet) – The expression plasmid for the truncated PmC11construct was transformed into *E. coli* GeneHogs competent cells and grown in minimal media

supplemented with SeMet and 30 μg mL^{-1} kanamycin at 37 °C using a GNF fermentor (14). A methionine auxotrophic strain was not required as SeMet is incorporated via the inhibition of methionine biosynthesis (15,16). Protein expression was induced using 0.1% (w/v) L-arabinose and the cells 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 μg mL^{-1} and the cells harvested and stored at -20 °C, until required. Primers used in this section are as follows:

I-PIPE(forward):
CTGTACTTCCAGGGCGAGACTCCGGAACCCCGGACAACCCGC,

I-PIPE (reverse):
AATTAAGTCGCGTTATTCATAAACTGCCTTATACCAGCCGAC,

V-PIPE (forward):
TAACGCGACTTAATTAACTCGTTTAAACGGTCTCCAGC,

V-PIPE (reverse):
GCCCTGGAAGTACAGGTTTTCGTGATGATGATGATGAT.

Protein purification for crystallisation. 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 $MgSO_4$ and 1 μL of 250 U μL^{-1} Benzonase (Sigma). The cell lysate was then clarified by centrifugation (32,500 g for 25 min at 4 °C) before being passed over Ni^{2+}–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 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 TEV 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 g for 10 min at 4°C) and the supernatant loaded onto Ni^{2+}–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^{-1} using Amicon Ultra-15 5K centrifugal concentrators (Millipore).

Crystallisation and Data Collection. PmC11 was crystallized using the nanodroplet vapour diffusion method using standard JCSG crystallization protocols (11). Drops were comprised of 200 nl protein solution mixed with 200 nl crystallization solution in 96-well sitting-drop plates, equilibrated against a 50 μl reservoir. Crystals of PmC11 were grown at 4°C in mother liquor consisting of 0.2 M $NH_4H_2PO_4$, 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 (SAM) (17) at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA). Single wavelength anomalous dispersion (SAD) diffraction 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 $P2_1$ with a = 39.11, b = 108.68, c = 77.97 Å and β = 94.32°. The unit cell contained two molecules in the asymmetric unit (ASU) resulting in a solvent content of 39% (Matthews' coefficient (Vm) of 2.4 $Å^3$ Da^{-1}).

Structure Determination—The PmC11 structure was determined by the SAD 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 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_{cryst} and R_{free} 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 ASU, and NCS restraints. The refined structure contains residues

24–375 and 28–375 for the two molecules in the crystallographic ASU. Structural validation was carried using the JCSG Quality Control Server (http://smb.slac.stanford.edu/jcsg/QC/) that analyses 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, www.wwpdb.org) 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 3D structural figures were prepared with PyMol (33) with the electrostatic surface potential additionally calculated with APBS (18) contoured at ±5 kT/e. Architectural comparisons with known structures revealed that PmC11 was most structurally similar to caspase-7, gingipain-K and legumain (PBD IDs 4hq0, 4tkx, and 4awb, respectively). The statistical significance of the structural alignment between PmC11 and both caspase-7 and gingipain-K was reported to be 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 ID 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:Biology Materials Repository and used to generate a cleavage site mutant PmC11^{K147A} and an active-site mutant PmC11^{C179A} using the QuikChange Site-directed Mutagenesis kit (Stratagene) as per the manufacturer's instructions using the following primers:
K147A mutant (forward):
CAGAATAAGCTGGCAGCGTTCGGACAGG ACG,
K147A mutant (reverse):
CGTCCTGTCCGAACGCTGCCAGCTTATTC TG:
C179A mutant (forward):
CCTGTTCGATGCCGCCTACATGGCAAGC:
C179A mutant (reverse):
GCTTGCCATGTAGGCGGCATCGAACAGG.
The expression plasmids containing PmC11 were transformed into *E. coli* BL21 Star (DE3) and grown in Luria-Bertani media containing 30 μg mL^{-1} 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 to the protein production for crystallography, a slightly modified purification protocol was employed. 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 cocktail (cOmplete, Roche). 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 (BioRad) with a BSA standard.

Fluorogenic substrate activity assays—The release of the fluorescent group AMC (7-Amino 4-Methyl Coumarin) from potential peptide substrates was used to assess the activity of PmC11. Peptidase activity was tested using 20 μg PmC11 and 100 μM substrate (unless otherwise stated) in assay buffer to a final reaction volume of 200 μL and all samples were

incubated (without substrate) at 37 °C for 16 h prior to carrying out the assay. The substrate and plate reader were brought to 37 °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 nm and of 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% DMSO. The amount of AMC (µmoles) released was calculated by generating an AMC standard curve (as described (34)) and the specific activity of PmC11 was calculated as pmoles of AMC released per min per mg of the protein preparation.

The reaction rates (Vmax) and Km values were determined for mutants $PmC11^{K147A}$ and $PmC11^{C179A}$ by carrying out the activity assay at varying concentrations of Bz-R-AMC between 0 and 600 µM. The blank-corrected relative fluorescence units were plotted against time (min) with ΔFU/T giving the reaction rate. The Km and Vmax of PmC11 and $PmC11^{K147A}$ 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 ± SD.

Effect of VRPR-FMK on PmC11— To test the effect of the inhibitor on the activity of PmC11, 25 µM of Z-VRPR-FMK (100mM stock in 100% DMSO, Enzo Life Sciences), 20 µg PmC11, 100 µM 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 µg PmC11, 25 µM inhibitor, 1 mM EGTA in assay buffer. The reactions were incubated at 37 °C for 20 min before being stopped by the addition SDS/PAGE sample buffer. Samples were analysed by loading the reaction mixture on a 10% NuPAGE Bis-Tris gel using MES buffer.

Effect of cations on PmC11—The enzyme activity of PmC11 was tested in presence of various divalent cations: Mg^{2+}, Ca^{2+}, Mn^{2+}, Co^{2+}, Fe^{2+}, Zn^{2+} and Cu^{2+}. The final concentration of the salts ($MgSO_4$, $CaCl_2$, $MnCl_2$, $CoCl_2$, $FeSO_4$, $ZnCl_2$, and $CuSO_4$) was 1 mM and a control was set up without divalent ions but with addition of 1 mM EGTA as the control. The assay was set up using 20 mg PmC11, 1 mM salts, 100 µM R-AMC substrate, and the assay buffer, and incubated at 37°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.

Auto-processing profile of PmC11—Auto-processing of PmC11 was evaluated by incubating the enzyme at 37°C and removing samples at one hour intervals for from 0 to 16 hours and placing into SDS/PAGE loading buffer to stop the processing. Samples were then analysed on a 4-12% NuPAGE (Thermo Fischer) Novex Bis-Tris gel run in MES buffer.

Auto-processing cleavage site analysis— To investigate whether processing is a result of intra or inter-molecular cleavage the $PmC11^{C179A}$ mutant was incubated with increasing concentrations of activated PmC11 (0, 0.1, 0.2, 0.5, 1, 2 and 5 µg). The final assay volume was 40 µL and the proteins were incubated at 37°C for 16 h in the PmC11 assay buffer. To stop the reaction, NuPAGE sample buffer was added to the protein samples and 20 µL was analysed on 10% NuPAGE Novex Bis-Tris gel using MES buffer. These studies revealed no apparent cleavage of $PmC11^{C179A}$ by the active enzyme at low concentrations of PmC11 and that only limited cleavage was observed when the ratio of active enzyme (PmC11: $PmC11^{C179A}$) 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 α/β/α sandwich architecture comprised of a central nine-stranded β-sheet, with an unusual C-terminal domain (CTD), starting at residue K250. A single cleavage was observed in the polypeptide chain at K147 (Figs 1A, B), where both ends of the cleavage site are fully visible and well-ordered in the electron density. The central nine-stranded

β-sheet (β1–β9) of PmC11 consists of six parallel and three anti-parallel β-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 α-helices with six of these (α1–α2 and α4–α7) closely surrounding the β-sheet in an approximately parallel orientation. Helices α1, α7and α6 are located on one side of the β-sheet with α2, α4 and α5 on the opposite side (Fig 1A). Helix α3 sits at the end of the loop following β5 (L5), just preceding the K147 cleavage site, with both L5 and α3 pointing away from the central β-sheet and towards the CTD, which starts with α8. The structure also includes two short β-hairpins (βA–βB and βD–βE) and a small β-sheet (βC–βF), which is formed from two distinct regions of the sequence (βC precedes α11, α12 and β9, while βF follows the βD-βE hairpin) in the middle of the CTD (Fig. 1B).

The CTD of PmC11 is composed of a tight helical bundle formed from helices α8–α14 and includes strands βC and βF, and β-hairpin βD–βE. The CTD sits entirely on one side of the enzyme interacting only with α3, α5, β9 and the loops surrounding β8. Of the interacting secondary structure elements (SSEs), α5 is perhaps the most interesting. This helix makes a total of eight hydrogen bonds with the CTD, including one salt bridge (R191–D255) 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 (PBD IDs 4hq0, 4tkx, and 4awb, respectively) (Table 2). The C-terminal domain is unique to PmC11 within 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 β-strands in PmC11 (β1–β2 and β5–β8) share the same topology as the six-stranded β-sheet found in caspases, with strands β3, β4 and β9 located on the outside of this core structure (2) (Fig. 1B, box). H133 and C179 were found at locations structurally homologous to the caspase catalytic dyad, and other clan CD structures (2), at the C-termini of strands β5 and β6, respectively (Figs 1A, B and 2A). Additionally, a multiple sequence alignment of C11 proteins revealed that these residues are also highly conserved (data not shown), revealing the location of the C11 catalytic dyad.

Five of the α-helices surrounding the β-sheet of PmC11 (α1, α2, α4, α6 and α7) 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 β-sheet, PmC11 differs most significantly from other clan CD members at its C-terminus, where the CTD contains a further seven α-helices and four β-strands after β8.

Auto-Processing of PmC11—Purification of recombinant PmC11 (MW = 42.6 kDa) revealed partial processing into two cleavage products of 26.4 and 16.2 kDa, related to the observed cleavage at K147 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 K147 is found immediately after α3, in loop L5 within the central β-sheet (Figs. 1A, 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 to the catalytic dyad with A148 being 4.5 and 5.7 Å from H133 and C179, respectively. Consequently, it appears feasible that the helix attached to K147 (α3) could be responsible for steric auto-inhibition of PmC11 as K147 moves to form a covalent bond with A148. Prior to cleavage this would limit access to the active site; thus, the cleavage would be required for full activation of PmC11. To investigate this possibility, two mutant forms of the enzyme were created: $PmC11^{C179A}$ (a catalytically inactive mutant) and $PmC11^{K147A}$ (a cleavage-site mutant). Initial SDS-PAGE and western blot analysis of both mutants revealed no discernible processing occurred as compared to active PmC11 (Fig. 2C). The $PmC11^{K147A}$ mutant enzyme had a markedly different reaction rate (Vmax) compared to WT, where the reaction velocity of PmC11 was 10 times

greater than that of PmC11^{K147A} (Fig. 2D). Taken together, these data reveal that PmC11 requires processing at K147 for optimum activity.

To investigate whether processing is a result of intra or inter-molecular cleavage, the PmC11^{C179A} mutant was incubated with increasing concentrations of processed and activated PmC11. These studies revealed that there was no apparent cleavage of PmC11^{C179A} by the active enzyme at low concentrations of PmC11 and that only limited cleavage was observed when the ratio of active enzyme (PmC11: PmC11^{C179A}) was increased to ~1:10 and 1:4, with complete cleavage observed at a ratio of 1:1 (Fig. 2E). This suggests that the limited cleavage was most likely an effect of the increasing concentration of PmC11 and intra-molecular cleavage. Collectively, these data suggest that a pro-form of PmC11 is auto-inhibited by a section of L5 blocking access to the active site, prior to intra-molecular cleavage at K147. This cleavage subsequently allows movement of the region containing K147 and the active site to open up for substrate access.

Substrate specificity of PmC11—The autocatalytic cleavage of PmC11 at K147 (sequence KLK^A) demonstrates that the enzyme accepts substrates with Lys in the P_1 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_1 but was active towards substrates with a basic residue in P_1 such as Benzoyl-L-Arg-4-methylcoumaryl-7-amide (Bz-R-AMC), benzyloxycarbonyl-Gly-Gly-Arg-AMC (Z-GGR-AMC) and t-Butyloxycarbonyl-Val-Leu-Lys (BOC-VLK-AMC). The rate of cleavage was ~3-fold greater toward the single Arg substrate Bz-R-AMC than the other two (Fig. 2F) and, unexpectedly, PmC11 showed no activity towards BOC-K-AMC. These results confirm that PmC11 accepts substrates containing Arg or Lys in P_1 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 PmC11structure 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 N50, D177 and T204 with the side chains of G134, D207 and M205 also contributing to the pocket (Fig. 2A). Interestingly, these residues are in regions that are structurally similar to those involved in the S_1 binding pockets of other clan CD members (shown in (2)).

Since PmC11 recognises basic substrates, the tetra-peptide inhibitor Z-VRPR-FMK was tested as an enzyme inhibitor and was found to inhibit both the auto-processing 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-paracacaspase (MALT1-P), in complex with Z-VRPR-FMK (35), revealed that the PmC11 dyad sits in a very similar position to that of active MALT1-P and that N50, D177 and D207 superimpose well with the principal MALT1-P inhibitor binding residues (D365, D462 and E500, respectively (VRPR-FMK from MALT1-P with the corresponding PmC11 residues from the structural overlay is shown in Fig. 1D), as described in (5)). D177 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, D207 resides on a flexible loop pointing away from the S1 binding pocket (Fig. 1D). 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, N50, D177 and D207 are most likely responsible for the substrate specificity of PmC11. D177 is highly conserved throughout the clan CD C11 peptidases and is thought to be primarily responsible for the 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 overlayed on the MALT1-P structure (Fig. 3C).

Comparison with clostripain—Clostripain is the founding member of the C11 peptidases and contains an additional 149 residues compared to 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 (R181 and R190), which results in the removal of a nonapeptide, and is required for full activation of the enzyme (37) (highlighted in Fig. 1A). Interestingly, R190 was found to align with K147 in PmC11. In addition, the predicted primary S1-binding residue in PmC11 D177 also overlays with the residue predicted to be the P1 specificity determining residue in clostripain (38) (D229, Fig. 1A).

As studies on clostripain revealed addition of Ca^{2+} ions are required for full activation, the Ca^{2+} dependency of PmC11 was examined. Surprisingly, Ca^{2+} did not enhance PmC11 activity and, furthermore, other divalent cations, Mg^{2+}, Mn^{2+}, Co^{2+}, Fe^{2+}, Zn^{2+} and Cu^{2+}, were necessary for PmC11 activity (Fig. 3D). In support of these findings, EGTA did not inhibit PmC11 suggesting that, unlike clostripain, PmC11 does not require 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 β-sheet with a novel C-terminal domain. The structural similarity of PmC11 with its nearest structural neighbours in the PDB is decidedly low, overlaying better with six-stranded caspase-7 than any of the other larger members of the clan (Table 3). 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 auto-processing. PmC11 differs from clostripain in that is 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 β-sheet. This is also the case in PmC11, although the cleavage loop is structurally different to that found in the caspases, it follows the catalytic His (Fig. 1A), as opposed to the Cys as 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 have both been shown to undergo inter-molecular (*trans*) cleavage and legumain and MARTX-CPD are reported to perform intra-molecular (*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 modelling 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.

ACKNOWLEDGEMENTS. We thank all members of the JCGS high throughput structural biology pipeline for their contribution to this work. This work was supported by the Medical Research Council (MR/K019384), the Wellcome Trust (091790, 104111) and NIH U54 GM094586 (JCSG). The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health (NIH), National Center for Research Resources, Biomedical Technology Program (P41RR001209), and the National Institute of General Medical Sciences (NIMGS). The Advanced Light Source is

supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIGMS or NIH. Genomic DNA from *P. merdae* was extracted from cells (ATCC #43184) obtained from the American Type Culture Collection (ATCC).

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: KM, JSG DD, IAW, JCM designed research; KM, JSG, DD performed research; KM, JSG, DD, GHC, AS, MAE, and JCM analysed data; AG, SAL, AMD, MAE and IAW supervised various components of the JCSG structural genomics pipeline; MKG, AG, SAL, AMD and MAE contributed reagents/materials/analysis tools; and KM, JSG, GHC, MAE, IAW, and JCM wrote the paper.

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., Hammill, L. D., and et al. (1994) Crystal structure of the cysteine protease interleukin-1 beta-converting enzyme: a (p20/p10)2 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., Bogyo, 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., Buttle, 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., Wuthrich, 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., Kreusch, 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. Doublie, 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., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242
24. Murshudov, G. N., Skubak, 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, 29
28. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) SFCHECK: 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., Soding, 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 Schuttelkopf, 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 (Schrodinger, LLC, San

Carlos, CA), version 1.2r3pre.
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 Renatus, 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., Schnaufer A., Mottram,J.C. (2016) PNT1 is a C11 Cysteine Peptidase Essential for Replication of the Trypanosome Kinetoplast. *J.Biol.Chem.*
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

FIGURE LEGENDS

FIGURE 1. Crystal structure of a C11 peptidase from *Parabacteroides merdae. (A)* Primary sequence alignment of PmC11 (Uniprot ID: A7A9N3) and clostripain (Uniprot ID: P09870) with identical residues highlighted in grey shading. The secondary structure of PmC11 from the crystal structure is mapped onto its sequence with the position of the PmC11 catalytic dyad, auto-catalytic cleavage site (K147), and S_1 binding pocket Asp (D211) highlighted by a red star, a red downturned triangle, and a red upturned triangle, respectively. Connecting loops are colored gray, the main β-sheet is in orange, with other strands in olive, α- helices are in blue and the nonapeptide linker of clostripain that is excised upon auto-cleavage is underlined in red. Sequences around the catalytic site of clostripain and PmC11 align well. *(B)* Topology diagram of PmC11 colored as in B except that additional (non-core) β-strands are in yellow. Helices found on either side of the central β-sheet are shown above and below the sheet, respectively. The position of the catalytic dyad (H, C) and the processing site (K147) are highlighted. Helices (1–14) and β-strands (1–9 and A–F) are numbered from the N-terminus. The core caspase fold is highlighted in the grey box. *(C)* Tertiary structure of PmC11. The N and C termini (N, C) of PmC11 along with the central β-sheet (1-9), helix α5, and helices α8, α11 and α13 from the C-terminal domain, are all labelled. Loops are colored gray, the main β-sheet is in orange, with other β-strands in yellow, and α-helices are in blue.

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_1 binding site. The overall structure of PmC11 is shown in grey, looking down into the catalytic site with the catalytic dyad in red, The two ends of the autolytic cleavage site (K147 and A148, 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 $PmC11^{C179A}$ (C) and $PmC11^{K147A}$ (K) were examined by SDS-PAGE (lane 1) and Western Blot analysis using an anti-His antibody (lane 2) show that PmC11 auto-processes, whereas mutants, $PmC11^{C179A}$ and $PmC11^{K147A}$, do not show auto-processing *in vitro*. *(D)* Cysteine peptidase activity of PmC11. Km and Vmax 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 ± S.D. are shown. *(E)* Intramolecular processing of $PmC11^{C179A}$ by PmC11. $PmC11^{C179A}$ (20 µg) was incubated overnight at 37°C with

increasing amounts of processed PmC11 and analysed on a 10% SDS-PAGE gel. Inactive $PmC11^{C179A}$ was not processed to a major extent by active PmC11 until around a ratio of 1:4 (5 µg of active PmC11). A single lane of 20 µg of active PmC11 (labelled 20) is shown for comparison. (F) Activity of PmC11 against basic substrates. Specific activity is shown ± S.D. from three independent experiments. *(G)* Electrostatic surface potential of PmC11 shown in a similar orientation as Fig. S4, where blue and red denote positively and negatively charged surface potential, respectively, contoured at ±5 kT/e. The position of the catalytic dyad, one potential key substrate binding residue D177, and the ends of the cleavage site K147 and A148 are indicated.

FIGURE 3. *PmC11 binds and is inhibited by Z-VRPR-FMK and does not require 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; ±SD). *(B)* Gel-shift assay reveals that Z-VRPR-FMK binds to PmC11. PmC11 was incubated with (+) or without (-) Z-VRPR-FMK and the samples analysed 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-paracacaspase (MALT1-P) superimposed. A 3D 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 MALTI_P structures and indicates the presumed active site of PmC11. Residues surrounding the inhibitor are labelled and represent potentially important binding-site residues, labelled in black and shown in an atomic representation. Carbon atoms are shown in grey, nitrogen in blue, and oxygen in red. *(C)* 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 Ca^{2+}, Mn^{2+}, Zn^{2+}, Co^{2+}, Cu^{2+}, Mg^{2+}, Fe^{3+} with EGTA as a negative control, and 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 require metal ions for proteolytic activity. Furthermore, Cu^{2+}, Fe^{2+}, Zn^{2+} appear to inhibit PmC11.

TABLE 1. **Crystallographic statistics for PDB ID 3UWS.**

Data collection	
Wavelength (Å)	0.9793
Spacegroup	$P2_1$
Unit cell dimensions a, b, c (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
R_{merge} on I (%)	10.2 (49.0)
R_{meas} on I (%)	11.0 (52.7)
R_{pim} on I (%)	4.1 (19.2)
I/ σ_I	15.6 (4.6)
Wilson B ($Å^2$)	15.9
Completeness (%)	99.6 (99.8)
Multiplicity	7.3 (7.5)
Model and Refinement	
Reflections (total/test)	70,883/3,577
R_{cryst}/R_{free} (%)	14.3/17.5
No. protein residues/atoms	700/5612
No. water/EDO molecules	690/7
ESU based on R_{free} (Å)	0.095
B-values ($Å^2$)	
Average isotropic B (overall)	20.0
Protein overall	18.8
All main/side chains	16.7/20.8
Solvent/EDO	29.4/35.6
R.M.S deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.6
Ramachandran analysis (%)	
Favoured regions	97.0
Allowed regions	3.0
Outliers	0.0

Values in parentheses are for the highest resolution shell. $R_{merge}= \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i(hkl)$; $R_{meas} = \Sigma_{hkl}[N/(N-1)]^{1/2}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$; and and $R_{pim.}$ (precision-indicating R_{merge}) = $\Sigma_{hkl}[(1/(N-1)]^{½}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$ (43), where N is the multiplicity of reflection hkl, and $I_i(hkl)$ and $\langle I(hkl)\rangle$ are the intensity of the ith measurement and the average intensity of reflection hkl, respectively (44). R_{cryst} and $R_{free}= \sum \big| |F_{obs}|-|F_{calc}| \big| /\sum|F_{obs}|$ for reflections in the working and test sets respectively, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. R_{free} is the same as R_{cryst} but for 5% of the total reflections chosen at random and omitted from structural refinement. ESU is the estimated standard uncertainties of atoms and the average isotropic B includes TLS and residual B components.

TABLE 2. Summary of PDBeFOLD (45) superposition of structures found to be most similar to PmC11 in the PBD based on DaliLite (46). The results are ordered in terms of structural homology (Q_H), where $\%SSE^{PC\text{-}X}$ 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) $\%SSE^{X\text{-}PC}$ is the percentage of SSEs in X (as above) that can be identified in PmC11 (as above); % Seq. ID is the percentage sequence identity after structural alignment; N_{align} is the number of matched residues and RMSD is the root mean squared deviation on the Cα positions of the matched residues.

Enzyme	Family	PDB ID	Q^H	Z-score	$\%SSE^{PC\text{-}X}$	$\%SSE^{X\text{-}PC}$	% 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

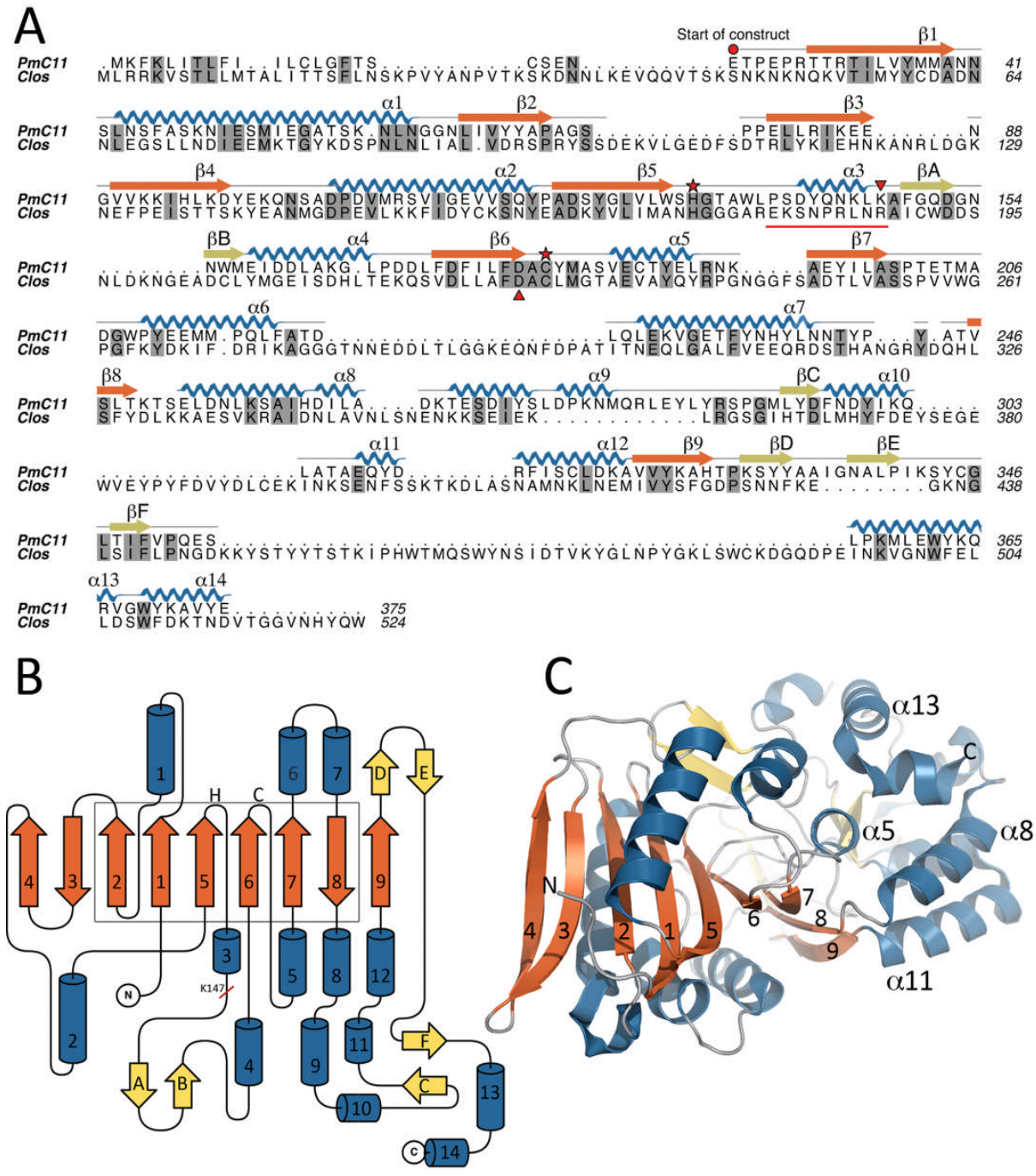
A
Start of construct
β1
PmC11 .MKFKLITLFI..ILCLGFTS.....CSEN.....ETPEPRTTRTILVYMMANN 41
Clos MLRRKVSTLLMTALITTSFLNSKPVYANPVTKSKDNNLKEVQQVTSKSNKNKNQKVTIMYYCDADN 64
α1
β2
β3
PmC11 SLNSFASKNIESMIEGATSK.NLNGGNLIVYYAPAGS.....PPELLRIKEE.....N 88
Clos NLEGSLLNDIEEMKTGYKDSPNLNLIAL.VDRSPRYSSDEKVLGEDFSDTRLYKIEHNKANRLDGK 129
β4
α2
β5
α3
βA
PmC11 GVVKKIHLKDYEKQNSADPDVMRSVIGEVVSQYPADSYGLVLWSHGTAWLPSDYQNKLKAFGQDGN 154
Clos NEFPEISTTSKYEANMGDPEVLKKFIDYCKSNYEADKYVLIMANHGGGARFKSNPRLNRAICWDDS 195
βB
α4
β6
α5
β7
PmC11 .....NWMEIDDLAKG.LPDDLFDFILFDACYMASVECTYELRNK.....AEYILASPTETMA 206
Clos NLDKNGEADCLYMGEISDHLTEKQSVDLLAFDACLMGTAEVAYQYRPGNGGFSADTLVASSPVVWG 261
α6
α7
PmC11 DGWPYEEMM.PQLFATD.....LQLEKVGETFYNHYLNNTYP...Y.ATV 246
Clos PGFKYDKIF.DRIKAGGGTNNEDDLTLGGKEQNFDPATITNEQLGALFVEEQRDSTHANGRYDQHL 326
β8
α8
α9
βC
α10
PmC11 SLTKTSELDNLKSAIHDILA...DKTESDIYSLDPKNMQRLEYLYRSPGMLYDFNDYIKQ..... 303
Clos SFYDLKKAESVKRAIDNLAVNLSNENKKSEIEK.....LRGSGIHTDLMHYFDEYSEGE 380
α11
α12
β9
βD
βE
PmC11 .....LATAEQYD.....RFISCLDKAVVYKAHTPKSYYAAIGNALPIKSYCG 346
Clos WVEYPYFDVYDLCEKINKSENFSSKTKDLASNAMNKLNEMIVYSFGDPSNNFKE.....GKNG 438
βF
PmC11 LTIFVPQES.....LPKMLEWYKQ 365
Clos LSIFLPNGDKKYSTYYTSTKIPHWTMQSWYNSIDTVKYGLNPYGKLSWCKDGQDPEINKVGNWFEL 504
α13
α14
PmC11 RVGWYKAVYE..... 375
Clos LDSWFDKTNDVTGGVNHYQW 524
B
1
6
7
D
E
H
C
4
3
2
1
5
6
7
8
9
3
5
8
12
2
N
K147
4
9
11
F
A
B
C
10
13
C
14
C
α13
C
α5
α8
N
4
3
2
1
5
6
7
8
9
α11

Figure 1

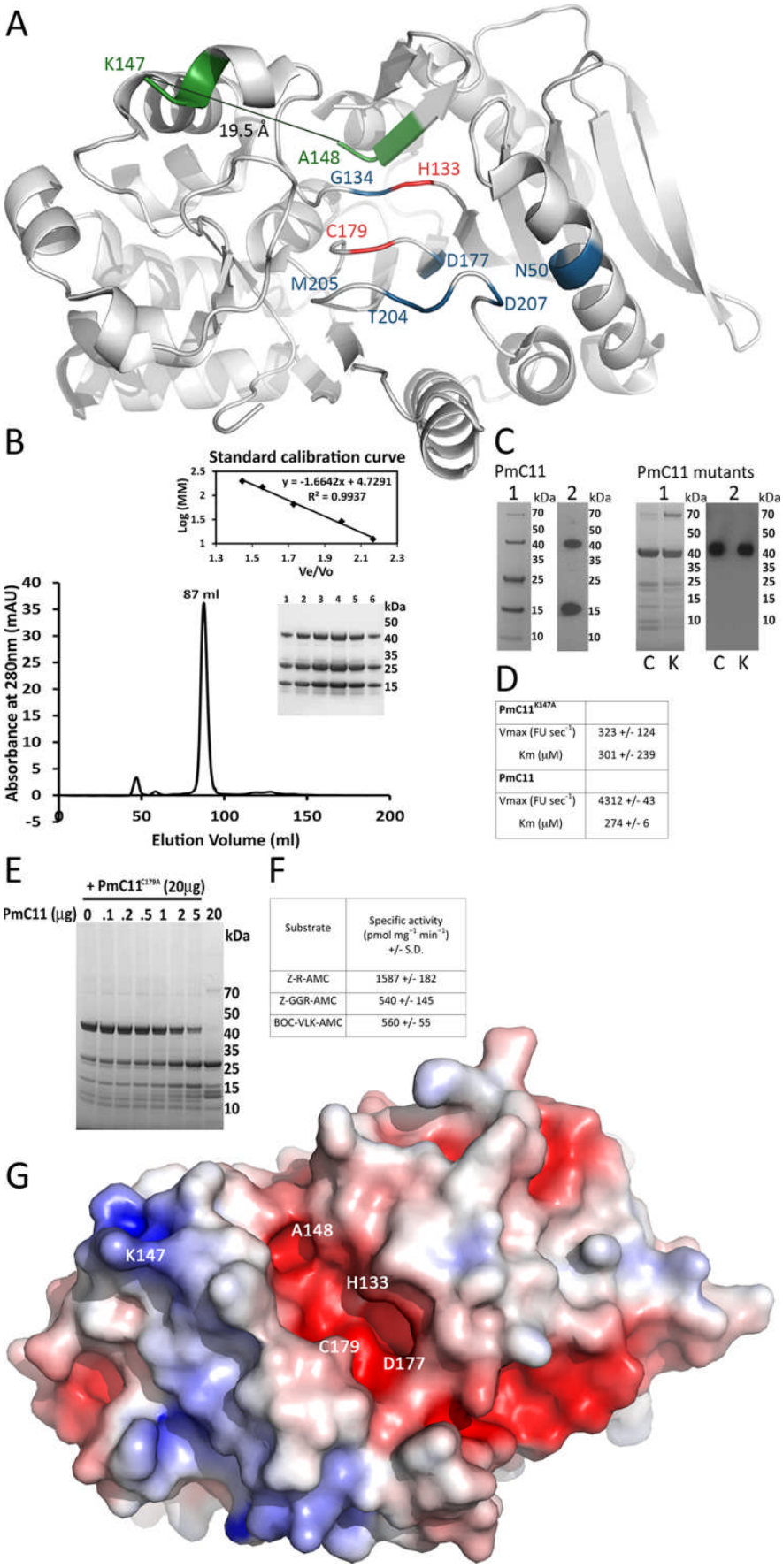
A
K147
19.5 Å
A148
G134
H133
C179
D177
N50
M205
D207
T204
B
Standard calibration curve
y = -1.6642x + 4.7291
R² = 0.9937
Log (MM)
Ve/Vo
87 ml
Absorbance at 280nm (mAU)
Elution Volume (ml)
kDa
C
PmC11
PmC11 mutants
C K C K
D
PmC11K147A
Vmax (FU sec-1) 323 +/- 124
Km (µM) 301 +/- 239
PmC11
Vmax (FU sec-1) 4312 +/- 43
Km (µM) 274 +/- 6
E
+ PmC11C179A (20µg)
PmC11 (µg) 0 .1 .2 .5 1 2 5 20
kDa
F
Substrate
Specific activity (pmol mg-1 min-1) +/- S.D.
Z-R-AMC 1587 +/- 182
Z-GGR-AMC 540 +/- 145
BOC-VLK-AMC 560 +/- 55
G
K147
A148
H133
C179
D177

Figure 2

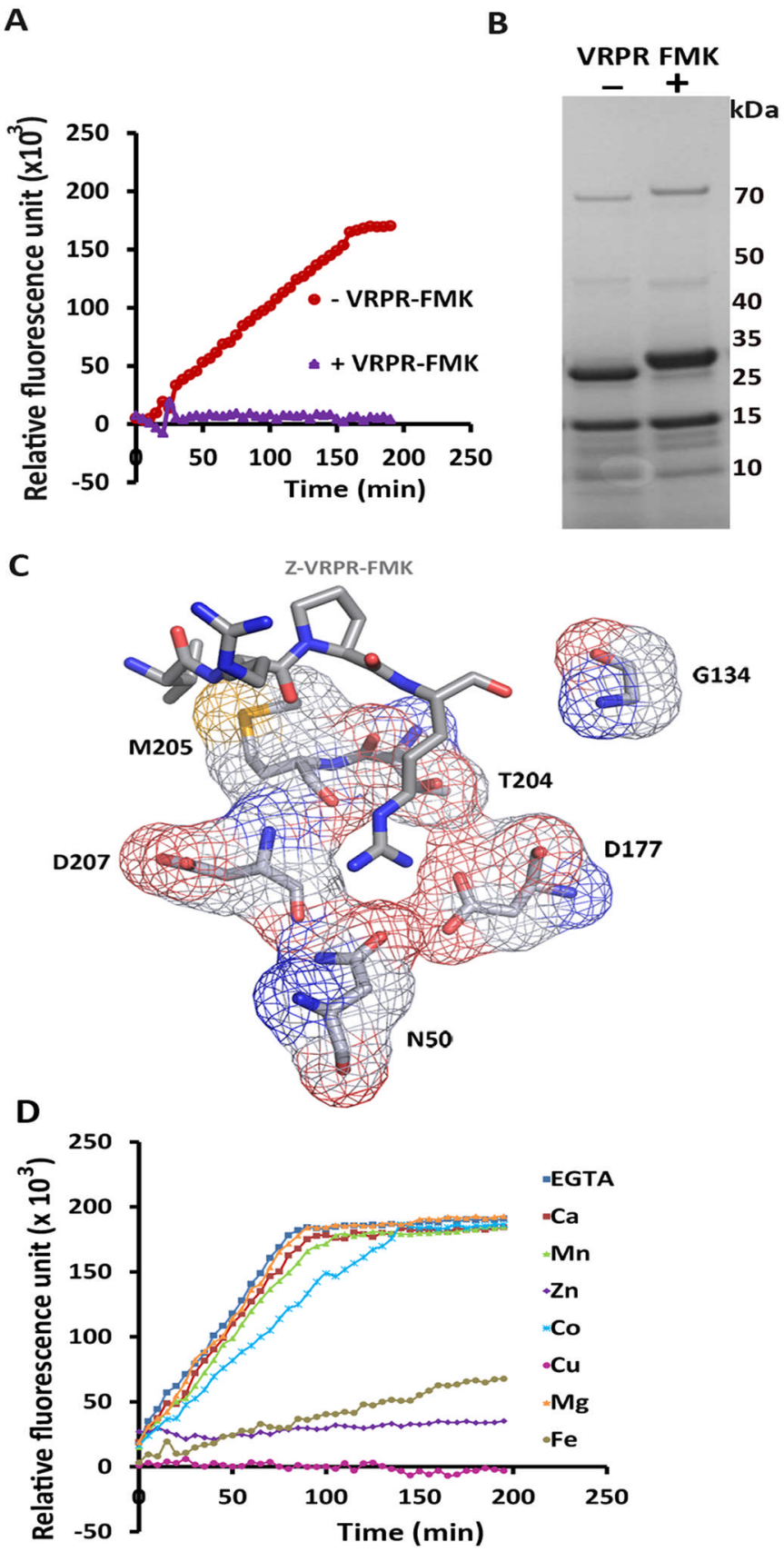
A
Relative fluorescence unit (x10^3)
250
200
150
100
50
0
-50
0
50
100
150
200
250
Time (min)
- VRPR-FMK
+ VRPR-FMK
B
VRPR FMK
– +
kDa
70
50
40
35
25
15
10
C
Z-VRPR-FMK
G134
M205
T204
D207
D177
N50
D
Relative fluorescence unit (x 10^3)
250
200
150
100
50
0
-50
0
50
100
150
200
250
Time (min)
EGTA
Ca
Mn
Zn
Co
Cu
Mg
Fe

Figure 3