Check for updates

# Activation mechanism and activity of globupain, a thermostable C11 protease from the Arctic Mid-Ocean Ridge hydrothermal system

Victoria Røyseth[1]*, Brianna M. Hurysz[2],
Anna-Karina Kaczorowska[3], Sebastian Dorawa[4],
Anita-Elin Fedøy[1], Hasan Arsın[1], Mateus Sá M. Serafim[2,5],
Samuel A. Myers[6], Olesia Werbowy[4], Tadeusz Kaczorowski[4],
Runar Stokke[1], Anthony J. O'Donoghue[2] and Ida Helene Steen[1]*

[1]Department of Biological Sciences, Center for Deep Sea Research, University of Bergen, Bergen,
Norway, [2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego,
San Diego, CA, United States, [3]Collection of Plasmids and Microorganisms | KPD, Faculty of Biology,
University of Gdańsk, Gdańsk, Poland, [4]Laboratory of Extremophiles Biology, Department of
Microbiology, Faculty of Biology, University of Gdańsk, Gdańsk, Poland, [5]Department of Microbiology,
Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil,
[6]La Jolla Institute for Immunology, La Jolla, CA, United States

Deep-sea hydrothermal vents offer unique habitats for heat tolerant enzymes
with potential new enzymatic properties. Here, we present the novel C11 protease
*globupain*, which was prospected from a metagenome-assembled genome of
uncultivated *Archaeoglobales* sampled from the Soria Moria hydrothermal vent
system located on the Arctic Mid-Ocean Ridge. Sequence comparisons against
the MEROPS-MPRO database showed that globupain has the highest sequence
identity to C11-like proteases present in human gut and intestinal bacteria.
Successful recombinant expression in *Escherichia coli* of the wild-type zymogen
and 13 mutant substitution variants allowed assessment of residues involved in
maturation and activity of the enzyme. For activation, globupain required the
addition of DTT and $Ca^{2+}$. When activated, the 52kDa proenzyme was processed
at $K_{137}$ and $K_{144}$ into a 12kDa light- and 32kDa heavy chain heterodimer. A structurally
conserved $H_{132}/C_{185}$ catalytic dyad was responsible for the proteolytic activity, and
the enzyme demonstrated the ability to activate *in-trans*. Globupain exhibited
caseinolytic activity and showed a strong preference for arginine in the P1 position,
with Boc-QAR-aminomethylcoumarin (AMC) as the best substrate out of a total
of 17 fluorogenic AMC substrates tested. Globupain was thermostable ($T_{m\ activated\ enzyme}$=94.51°C $\pm$0.09°C) with optimal activity at 75°C and pH 7.1. Characterization
of globupain has expanded our knowledge of the catalytic properties and
activation mechanisms of temperature tolerant marine C11 proteases. The unique
combination of features such as elevated thermostability, activity at relatively low
pH values, and ability to operate under high reducing conditions makes globupain
a potential intriguing candidate for use in diverse industrial and biotechnology
sectors.

KEYWORDS

cysteine peptidase, clostripain, extracellular enzyme, metagenome bioprospecting,
hydrothermal vent

# 1. Introduction

Proteases catalyze the hydrolysis of peptide bonds in proteins and are important in industrial applications (Gimenes et al., 2021). They are used in food and leather processing, as additives to detergents, as pharmaceuticals, and in biorefineries (Barzkar et al., 2018; García-Moyano et al., 2021). Proteases are among the most widely used enzymes globally, accounting for over 60 percent of all enzyme sales (Ward, 2011). Temperature-tolerant proteases offer the possibility for industrial processing at high temperatures by improving reaction rates, enhancing nongaseous reactant solubility, and reducing contamination by mesophiles (Kumar et al., 2000; Barzkar et al., 2018). Deep-sea hydrothermal vents sustain microorganisms at high temperatures (Kuwabara et al., 2007; Pikuta et al., 2007; Nunoura et al., 2008), making them an interesting starting point for the discovery of new thermostable proteases (Barzkar et al., 2018). Moreover, the increasing sequence diversity of encoded proteases revealed in hydrothermal vent microorganisms (Li et al., 2015; Dombrowski et al., 2018; Cheng et al., 2021) offers considerable potential for discovering new and novel proteases with optimized catalytic properties that may support future innovations.

Proteases are remarkably diverse in terms of activity and the nucleophilic residues that participate in hydrolysis (Rawlings and Bateman, 2019). Clostripain is a well-characterized endopeptidase originating from the bacterium *Clostridium histolyticum* (accession MER0000831) and is a member of enzyme family C11. Peptidases in this family are characterized by the presence of a catalytic cysteine-histidine dyad with a preference for hydrolyzing arginine and lysine bonds in the P1 position (Ogle and Tytell, 1953; Barrett and Rawlings, 1996; Labrou and Rigden, 2004). Clostripain-like proteases are synthesized as inactive zymogens that have various requirements for activation (Kembhavi et al., 1991; McLuskey et al., 2016). Some require divalent cations such as $Ca^{2+}$ and/or reducing agents such as dithiothreitol (DTT) for activation and catalysis. Variance in the number of cleavage site(s) for activation is also observed, and in some cases, an amino acid linker peptide is removed (Gilles et al., 1979; Dargatz et al., 1993). Nevertheless, the resulting active peptidase will comprise of a light-and heavy chain making up a macromolecular active heterodimer. *In-trans* activation has been demonstrated in some, while others activate *in-cis*, reflecting the accessibility of cleavage sites to neighboring peptidase activity (Herrou et al., 2016; Roncase et al., 2017; González-Páez et al., 2019; Roncase et al., 2019).

This report presents a C11 protease called globupain, with "*globu*" representing its unclassified *Archaeoglobus* species origin and "*pain*" depicting it as a clostripain homolog. The type species, *Archaeoglobus fulgidus* (Stetter, 1988), of genus *Archaeoglobus*, was one of the first archaea to have its genome sequenced 25 years ago (Klenk et al., 1997). It has since served as a model thermophilic archaeon and has provided important information about archaeal DNA replication (Maisnier-Patin et al., 2002), DNA repair (Birkeland et al., 2002; Knævelsrud et al., 2010), thermostable enzymes (Madern et al., 2001; Steen et al., 2001) and enzymes of biotechnological relevance (Isupov et al., 2019; Palombarini et al., 2020). With globupain, we have discovered a novel archaeal clostripain-like protease with a complex activation mechanism. Its unique catalytic properties and high thermal stability makes globupain a promising candidate for industrial applications.

# 2. Materials and methods

## 2.1. Environmental sampling, DNA extraction, and sequencing

The Soria Moria vent field is part of the Jan Mayen vent fields (JMVFs), located at the southern part of the Mohns Ridge (Pedersen et al., 2005, 2010) in the Norwegian-Greenland Sea (71.2°N, 5.5°W). The end-member fluids of white smokers in the Soria Moria vent field have a pH of 4.1 and a concentration of hydrogen sulfide of $4.1\,mmol\,kg^{-1}$ (Dahle et al., 2015). In June of 2011, an *in-situ* titanium incubator (Stokke et al., 2020) consisting of one chamber filled with 2 g of dried krill shells (Nofima, Bergen, Norway), mixed with grained flange rock material (Dahle et al., 2015), was deployed at ~30–35 cm below seafloor (blsf) in sediments at 716 m depth. The temperature was measured to be ~40°C and ~70°C at 20 and 30 cm blsf, respectively, indicating diffuse hydrothermal venting. The sample was recovered in July 2012, and the incubated material was immediately snap-frozen in liquid nitrogen and stored at −80°C. DNA was extracted with FastDNA™ SPIN Kit for Soil (MP Biomedicals, CA, United States) and sequenced at the Norwegian Sequencing Center in Oslo, NSC.[1]

## 2.2. Metagenomic assembly, binning, and annotation

For the primary metagenome, one plate of 454 GS FLX Titanium shotgun reads (average read length; 730 bp) was sequenced and assembled using the Newbler assembler v.2.8 (Roche, Basel, Switzerland) with a minimum identity of 96% over a minimum of 35 bases. In total, 0.9 million (75%) of the 454 raw reads were assembled into contigs resulting in 7448 contigs >500 bp and an N50 contig size of 10,887 bp. Open reading frame (ORF) predictions were made using Prodigal v2.60 (Hyatt et al., 2010) and screened against MEROPS (Rawlings et al., 2014; Release 9.13). Putative signal peptides were identified by SignalP v4.1 (Petersen et al., 2011). For the secondary metagenome, Illumina NovaSeq 150 bp paired-end reads were filtered and assembled using fastp v0.23.2 and MEGAHIT v1.2.9, respectively. Of the 290 million filtered reads, 90.8% mapped to the assembly using the bwa-mem aligner v.0.7.17 (Vasimuddin et al., 2019). Metagenome-assembled genomes (MAGs) were binned and refined using MetaWrap v1.3.2 (Uritskiy et al., 2018), which included the binning tools, MetaBat2 v2.12.1 (Kang et al., 2015, 2019), MaxBin2 v2.2.6 (Wu et al., 2016) and CONCOCT v1.0.0 (Alneberg et al., 2014). Contamination and completeness of the MAGs were assessed with CheckM v1.0.7 (Parks et al., 2015). Furthermore, the taxonomic classification of the globupain-associated MAG (INS_M23_B45) was performed using the GTDB toolkit v2.1.0 (Chaumeil et al., 2019, 2022) with the GTDB release 207_v2 (Parks et al., 2018, 2020, 2022). ORF predictions were made with Prodigal v2.6.0 (Hyatt et al., 2010) as part of the annotation workflow designed by Dombrowski et al., 2020. Cross-referencing the cloned globupain from the primary metagenome with the assembly from the secondary metagenome

---

1 www.sequencing.uio.no

identified an ORF sharing 100% identity over 481 amino acids and 4 additional amino acids at the C-terminus (CFVD).

## 2.3. Sequence alignment and three-dimensional modeling

Sequence alignment was made using the ESPript 3.0 utility (Robert and Gouet, 2014). The amino acid sequences of distapain (MER0095672), clostripain (MER0095672), thetapain (MER0028004), and PmC11 (MER0199417) were retrieved from the MEROPS database (Rawlings et al., 2018). The alignment was based on NCBI BLAST+ sequence similarity search results using the blastp program (Madeira et al., 2022) with MEROPS-MPRO sequences.

The translated globupain DNA sequence (Supplementary Work Sheet 1) was submitted to AlphaFold (software version 2.1.1), available at NMRbox (Maciejewski et al., 2017),[2] for prediction of a three-dimensional (3D) protein structure with atomic accuracy. AlphaFold (Jumper et al., 2021) modeled globupain structure was downloaded (Varadi et al., 2022), and the model's alignment and analysis was assessed with PyMOL software (v0.99c; DeLano, 2006). Lastly, the modeled structure was compared to a previously published crystallized C11 structure (PDB ID: 4YEC; Roncase et al., 2017) template for comparison purposes.

## 2.4. Gene synthesis

Based on primary metagenome data, the globupain gene (GenBank accession OQ718499) was synthesized by GenScript (GenScript, NJ, United States) and codon-optimized for *Escherichia coli* expression (Supplementary Work Sheet 1). The gene was cloned (cloning site *Nde*I and *Xho*I) into pET-21A by GenScript, omitting the predicted 21 amino acid signal peptide (SignalP v6.0, Teufel et al., 2022). The resulting signal-free protein was extended with Met at the N-terminus, whereas the C-terminus was extended before the C-terminal hexahistidine tag (His tag) with Leu and Glu (LEHHHHHH). For identification of amino acids in the catalytic dyad and in maturation (Figure 1), targeted amino acid residues (Table 1) were substituted with Ala, and the respective coding genes were synthesized and cloned by GenScript as described for the wild-type (WT) globupain.

## 2.5. Protein production and purification

Expression plasmid of globupain and substitution variants were transformed into BL21-Gold (DE3) chemically competent *E. coli* cells (Agilent, TX, United States) using a heat-pulse manual supplied by the manufacturer. Cells were spread onto LB-agar plates supplemented with 100 μg/mL ampicillin and incubated at 37°C overnight. Pre-cultures were inoculated by picking one single colony and incubating at 37°C in LB media containing 100 μg/mL ampicillin with 190 rpm shaking overnight (Innova 44, New Brunswick Scientific, St

Albans, United Kingdom). Expression cultures were inoculated with 5% of pre-culture in LB media with 100 μg/mL ampicillin at 37°C and 190 rpm. At $OD_{600}$ of 0.6, the temperature was set to 20°C, and the culture was equilibrated for 30 min. Heterologous expression was induced by IPTG brought to 0.1 mM IPTG, followed by overnight incubation (20°C). Cells were harvested by centrifugation at 5,500 rpm for 15 min at 4°C (Allegra™ 21R Centrifuge, Beckman Coulter, CA, United States). Pellets were stored at −20°C.

For purification of globupain and substitution variants, cells were resuspended in lysis buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 0.25 mg/mL lysozyme, 10% glycerol), placed on ice for 30 min, and lysed by ultra-sonication (5 times with 30% amplitude, in intervals of 20 s on ice using the Vibra Cell with probe model CV188, Sonics and Materials INC, LT, United States). The lysate was clarified by centrifugation at 5,500 rpm for 20 min at 4°C (Allegra™ 21R Centrifuge, Beckman Coulter, CA, United States). The sample was then loaded into a HisTrap HP 5 mL column (Cytiva, Uppsala, Sweden) equilibrated with 20 mM HEPES, 500 mM NaCl, 25 mM imidazole, pH 7.5 with a flow rate of 1 mL/min. After elution with 20 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 7.5. fractions with the highest amount of enzyme were pooled and concentrated. The buffer was changed (20 mM HEPES, 150 mM NaCl, 0.1% CHAPS, pH 7.5) using Amicon® Ultra-15 centrifugal filter unit (Merck KGaA, Darmstadt, Germany) with a 30 K molecular weight cut-off. Approximately 1 mL of the concentrated enzyme preparations were purified by gel filtration using a GE 16/600 Superdex 200 pg. column (Cytiva, Uppsala, Sweden). Purified globupain and substitution variants were stored in 20 mM HEPES, 150 mM NaCl, 0.1% CHAPS, pH 7.5 at 4°C.

## 2.6. Maturation/activation

For activation of globupain and substitution variants (Table 1), purified enzyme (< 5 mg/mL) was incubated at 75°C for up to 4.5 h in 20 mM tri-sodium citrate dihydrate, 150 mM NaCl, pH 5.5 (at RT) with 2.5 mM DTT and 1 mM $CaCl_2$, respectively (activation buffer). To investigate if globupain could *in-trans* activate, 10 μg of activated WT globupain was mixed with 10 μg of inactive $C_{185}A$ variant. The number and size of cleavage products were assessed by visualization of protein bands on 8%–16% SurePAGE precast gels (GenScript) using MES SDS running buffer (GenScript) in a Bio-Rad Mini-PROTEAN Tetra Cell (BioRad, Hercules, CA, United States). For sample preparation, 4′ lithium dodecyl sulfate (LDS) sample buffer (GenScript) with 2-mercaptoethanol was mixed with the protein sample, followed by denaturing at 95°C for 10 min. Gels were stained with InstantBlue™ ultrafast protein stain (Abcam, Cambridge, United Kingdom), and the size of bands was indicated by broad multi-color pre-stained protein standard (GenScript). Edman sequencing was performed on a Shimadzu PPSQ-53A at the Iowa State University Protein Facility, United States.

## 2.7. Mass spectrometry sample preparation

Following staining with InstantBlue™ ultrafast protein stain (Abcam, Cambridge, United Kingdom), the gel bands (Supplementary Figure 10C) were excised and washed three times
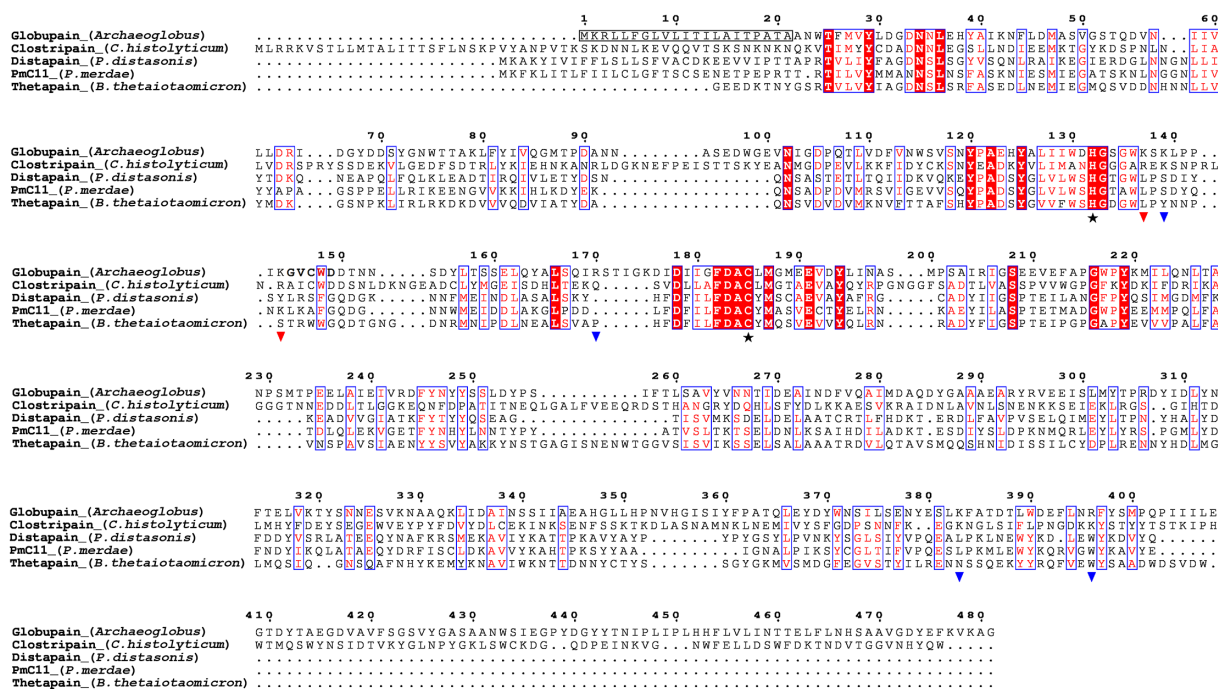
---

2   https://nmrbox.nmrhub.org/

**FIGURE 1**
Sequence alignment of the C11 proteases globupain (*Archaeoglobus*), clostripain (*Clostridium histolyticum*), distapain (*Parabacteroides distasonis*), PmC11 (*Parabacteroides merdae*), and thetapain (*Bacteroides thetaiotaomicron*) by ESPript 3.0. Symbols depict results from site-directed mutagenesis of the globupain coding sequence; ★, His/Cys catalytic dyad; ▼ (red), sites showing resistance against cleavage when the amino acid was mutated into alanine; ▼ (blue), sites able to cleave when mutated into alanine. The detected N-terminal residues following activation are shown in bold.

**TABLE 1** Overview of mutation variants of globupain and their targeted function.

| Position in globupain | Putative function |
|---|---|
| $C_{185}A$ | Catalytic dyad |
| $H_{132}A$ | Catalytic dyad |
| $K_{137}A$ | Self-activation |
| $K_{139}A$ | Self-activation |
| $K_{144}A$ | Self-activation |
| $K_{137}A/K_{139}A$ | Self-activation |
| $K_{137}A/K_{144}A$ | Self-activation |
| $K_{139}A/K_{144}A$ | Self-activation |
| $K_{137}A/K_{139}A/K_{144}A$ | Self-activation |
| $R_{170}A$ | Self-activation |
| $K_{383}A$ | Cleavage C-terminal peptide |
| $R_{396}A$ | Cleavage C-terminal peptide |
| $K_{383}A/R_{396}A$ | Cleavage C-terminal peptide |

with 25 mM $NH_4HCO_3$ in 50% acetonitrile for 10 min each time. The gels were then dried completely in a Savant Speed Vac Plus AR (Thermo Fisher Scientific, MA, USA). A mixture of 10 mM TCEP and 25 mM iodoacetamide in 25 mM $NH_4HCO_3$ was added to cover the gel pieces and this reaction proceeded in the dark for 1 h. Gels were then washed with 25 mM $NH_4HCO_3$ and dehydrated with 25 mM $NH_4HCO_3$ in 50% acetonitrile. Samples were then dried in a Savant Speed Vac Plus AR (Thermo Fisher Scientific) before addition of

12.5 ng/uL trypsin in 25 mM $NH_4HCO_3$ for peptide digestion. Following a 10 min incubation at 4°C, the samples were covered in 25 mM $NH_4HCO_3$ and the digestion proceeded at 37°C for 20 h. The supernatant was then transferred to a clean tube and the remaining peptides were extracted from the gel by addition of 50% acetonitrile, 5% formic acid. The extracted digests were then dried and resuspended in 0.1% formic acid to prepare for C18 (CPI International) ZipTip desalting.C18 columns were washed with methanol and spun for 45 s at 3,500 × g. Columns were then cleaned and equilibrated with 50% acetonitrile, 0.1% formic acid and 0.1% formic acid in water, respectively. Samples were then loaded onto columns and spun for 2 min at 2,000 × g. Samples were washed with 0.1% formic acid and spun at 3,500 × g for 45 s. Peptides were eluted from C18 with 50% acetonitrile, 0.1% formic acid by spinning at 3,500 × g for 45 s. Samples were dried in a Savant Speed Vac Plus AR (Thermo Fisher Scientific) and stored at −80°C until they were prepared for mass spectrometry.

## 2.8. LC-MS/MS

Samples were redissolved in 0.1% formic acid prior to LC-MS/MS injection. Chromatography was performed as previously described on an Easy-nLC 1200 (Thermo Fisher Scientific; Myers et al., 2019). Mass spectrometry was performed on an Orbitrap Eclipse with ETD and PTCR (Thermo Fisher Scientific). The scan range was 350–1,800 m/z at a resolution of 60,000 with a 50 ms maximum injection time. The top 8 scans were selected for MS2. MS/MS spectra were analyzed in PEAKS Studio (v 8.5) software (Bioinformatics Solutions Inc.). MS2 data were searched against the combined *E. coli* (GCA_000022665.2)

and globupain proteome (Supplementary Work Sheet 1). A precursor tolerance of 20 ppm and 0.01 Da was defined. Trypsin digestion was specified. The number of identified peptides were adjusted such that the false discovery rate was <1%. The data can be accessed on ProteomeXChange: PXD042411 or at ftp://massive.ucsd.edu/MSV000092007/.

## 2.9. Size-exclusion chromatography

Size-exclusion chromatography (SEC) analysis was performed using a Superdex 75 10/300 GL prepacked column connected to ÄKTA pure 25 chromatography system (GE Healthcare). The column was equilibrated with a 50 mM potassium phosphate buffer (pH 7.0), 150 mM NaCl and then loaded with a 500 μL sample of globupain protein (1 mg/mL). The flow rate of the run was adjusted to 0.5 mL/min, and the absorbance was measured at 280 nm (mAU, milli-absorbance units). For the experiment, the column was calibrated with proteins of known molecular weight: alcohol dehydrogenase (tetramer), 146,800; bovine serum albumin, 66,000; ovalbumin, 43,000; trypsin inhibitor, 22,000; and cytochrome C, 12,400 (Sigma-Aldrich, St. Louis, MO, United States). Dextran blue 2000 (Cytiva) was used to determine the column void volume.

## 2.10. Analytical ultracentrifugation

Sedimentation velocity experiments were performed in a Beckman-Coulter ProteomeLab XL-I analytical ultracentrifuge (Indianapolis, IN, United States), equipped with AN 60Ti 4-hole rotor and 12 mm path length, double-sector charcoal-Epon cells, loaded with 400 μL of samples and 410 μL of buffer (50 mM potassium phosphate buffer pH 7.0, 150 mM NaCl, and 1 mM EDTA). The experiments were conducted at 20°C and 50,000 rpm, using continuous scan mode and radial spacing of 0.003 cm. Scans were collected in absorbance, in 4 min intervals at 280 nm. Data were analyzed using the "Continuous c(s) distribution" model of the SEDFIT program (Schuck, 2000), with a confidence level (F-ratio) specified to 0.6. Biophysical parameters of the buffer: density (1,01395 g/cm³), and viscosity (1,030 mPa s), were measured at 20°C using Anton Paar DMA 5000 density meter and Lovis 2000 ME viscometer. Protein partial specific volume (V-bars) was estimated at 0.7309 mL/g using SEDNTERP software (version 1.10, Informer Technologies Inc., Dallas, TX, United States). The results were plotted using GUSSI graphical program (Brautigam, 2015).

## 2.11. Thermal stability analysis

Thermostability of the inactive globupain and its activated form were assayed by nanoscale differential scanning fluorimetry (nanoDSF). Measurements were performed with Prometheus NT.48 instrument (NanoTemper Technologies, München, Germany) and PR.ThermControl software using standard grade capillaries. Before measurement, the capillaries were sealed with a sealing paste according to the manufacturer recommendations. The results were further analyzed with PR.StabilityAnalysis software. Thermostability of globupain zymogen at 0.4 mg/mL concentration and its activated

form were assayed in 20 mM tri-sodium citrate (pH 5.5) buffer with 150 mM NaCl and 20 mM HEPES buffer (pH 7.5), 500 mM NaCl with 25 mM imidazole, respectively. Melting temperature ($T_m$) of proteins was determined by thermal unfolding with a temperature gradient between 20°C and 110°C at a ramp rate of 1°C/min. Thermal unfolding was measured by tryptophan and tyrosine fluorescence change at 330 and 350 nm emission wavelengths. All measurements were performed in triplicates.

## 2.12. Casein activity assays

The proteolytic activity of activated globupain and its substitution variants was assessed using the casein gelzan™ CM plate assay and the EnzChek™ Protease Assay Kit (Thermo Fisher Scientific, MA, USA), respectively. The gelzan™ CM plate assay was prepared by autoclaving 1.5% gelzan™ CM (Sigma-Aldrich) dissolved in 20 mM tri-sodium citrate (pH 5.5) buffer with 150 mM NaCl. The casein powder (Sigma-Aldrich) was dissolved in 20 mM tri-sodium citrate dihydrate (pH 5.5), 150 mM NaCl. NaOH was added until the casein was fully dissolved in the solution and then autoclaved at 115°C for 10 min. Casein was added to the gelzan™ CM solution at a final concentration of 1.0%. The casein gelzan™ CM solution was poured into sterile glass Petri dishes and set to harden. Wells were made by punching holes in the plates using an inverted sterile 1 mL pipet tip. To test for proteolytic activity, 60 μL of activated globupain at 0.7–1.0 mg/mL was added to wells and incubated overnight at 75°C. Clearance zones would indicate caseinolytic activity.

When the proteolytic activity was assessed using the EnzChek™ Protease Assay Kit (Thermo Fisher Scientific), 20 mM tri-sodium citrate (pH 5.5) buffer with 150 mM NaCl was used to dilute the 1.0 mg/mL stock solution of BODIPY FL casein to 10 μg/mL. An aliquot of the activated enzyme (0.15 μg) was then added to the reaction mixture (100 μL of total volume) comprising 12.5 μL of 10 μg/mL BODIPY FL casein working solution and 77.5 μL of 20 mM tri-sodium citrate (pH 5.5) buffer with 150 mM NaCl, 10 mM DTT, and 1 mM $CaCl_2$. The caseinolytic activity was measured by running a time-resolved fluorescence read at 60°C, measuring fluorescence intensity every 20 s for 100 cycles. Fluorescence was measured with excitation wavelength 485 nm and emission wavelength 530 nm using an EnSpire™ 2300 Multilabel Reader (PerkinElmer, Turku, Finland). All measurements were done in triplicates and baseline corrected using GraphPad Prism 9.1.0.

## 2.13. Substrate screening

A total of 17 fluorogenic substrates containing a C-terminal 7-amino-4-methylcoumarin (AMC) reporter group were screened for hydrolysis by globupain. These substrates consisted of Ac-VLTK-AMC, Ac-VLGK-AMC, Ac-VLVK-AMC, Ac-IK-AMC, Ac-YK-AMC, Ac-LK-AMC, Ac-LETK-AMC, Ac-IETK-AMC, Ac-AEIK-AMC, Ac-AIK-AMC, Boc-LRR-AMC (R&D Systems S-300), N-Benzoyl-FVR-AMC (Bachem I-1080), z-RR-AMC (Sigma C5429), Ac-RLR-AMC (AdipoGen AG-CP3-0013), Boc-QAR-AMC (Peptide International 3,135-v), z-VVR-AMC (Peptide International 3,211-v), Pyr-RTKR-AMC (Peptide International 3,159-v) where the N-terminal blocking groups Ac, Boc, z, and Pyr, correspond to acetyl,

tert-butyloxycarbonyl, benzyloxycarbonyl, and pyroglutamyl, respectively. Substrates containing K-AMC were synthesized by Dennis Wolan, The Scripps Research Institute, La Jolla, California and purified to >95%. All substrates were stored at −20°C as 10 mM stocks in DMSO. Substrates were diluted to 100 μM in 20 mM tri-sodium citrate dihydrate, 150 mM NaCl, 2.5 mM DTT, and 1 mM CaCl$_2$, pH 5.5, and mixed 1:1 with globupain such that the final concentration in the assay was of 2.9 μg/mL enzyme and 50 μM substrate. Assays were performed in triplicate wells of a black 384-well plate (Thermo Fisher Scientific). Fluorescence was measured at 50°C over 1 h at excitation 360 nm and emission 460 nm on a BioTek Synergy HTX Multimode Reader (BioTek, Agilent, Tx, United States). The reaction rate was calculated as the maximum velocity over 12 sequential readings and means with standard errors were calculated. A Welch's ANOVA and Brown Forsythe ANOVA were performed to calculate significances in GraphPad Prism 9.1.0. With Boc-QAR-AMC substrate, the Michaelis Menten kinetics was assessed at final concentrations ranging from 0 to 400 μM, and a Michaelis–Menten curve was fitted in GraphPad Prism 9.1.0 (Supplementary Figure 1).

## 2.14. Determining the pH-and temperature optimum

The pH optimum was determined using 5 nM globupain assayed against 50 μM Boc-QAR-AMC substrate in citrate phosphate buffer at various pH values. Buffers were made by mixing 0.2 M NaHPO$_4$ and 0.1 M citric acid following the McIlvaine's buffer system (McIlvaine, 1921) and pH was verified using a pH-meter. Samples were preincubated at 50°C for 10 min before fluorescence was measured. The optimum temperature for activity was assessed using Boc-QAR-AMC by incubating the enzyme and substrate at temperatures; 30°C, 40°C, 50°C, 60°C, 70°C, 80°C, 85°C, 90°C, 95°C, 100°C, 105°C, 110°C, 120°C, and 130°C in triplicate tubes in a final volume of 50 μL. The reaction temperatures were controlled using a digital dry bath (Thermo Fisher Scientific) with max temperature of 130°C set to the respective temperatures. After 10 min, the enzyme was inactivated by mixing 1:5 in 8 M urea. Samples were plated on a black 384-well plate (Thermo Fisher Scientific), and the total fluorescence was measured at excitation 360 nm and emission 460 nm. The data reported is the average RFU for each temperature with standard error. Gaussian distribution was fitted in GraphPad Prism 9.1.0. For the time-dependent loss of enzyme activity at pH 5.5 and pH 7.1, the enzyme activity (0.1 mg/mL) was measured at 60°C with the EnzChek™ Protease Assay Kit. All readings were done in triplicate reactions with the exception of duplicates for pH 7.1 at 120 min. Measurements were baseline corrected in GraphPad Prism 9.1.0. Statistical analyses were performed using RStudio 2022.07.1 + 554 "Spotted Wakerobin" release (RStudio, 2023).

## 2.15. Data availability

The native C11 globupain protease has been submitted to GenBank under the accession number OQ718499. The sequence is archived under BioProject PRJNA296938 and derived from the primary metagenomic assembly, BioSample SAMN04111445. The reconstructed *Archaeoglobus* genome, INS_M23_B45, has been archived under the BioSample SAMN33944460 derived from the secondary metagenome, BioSample SAMN33925184. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JARQZL000000000. The version described in this paper is version JARQZL010000000.
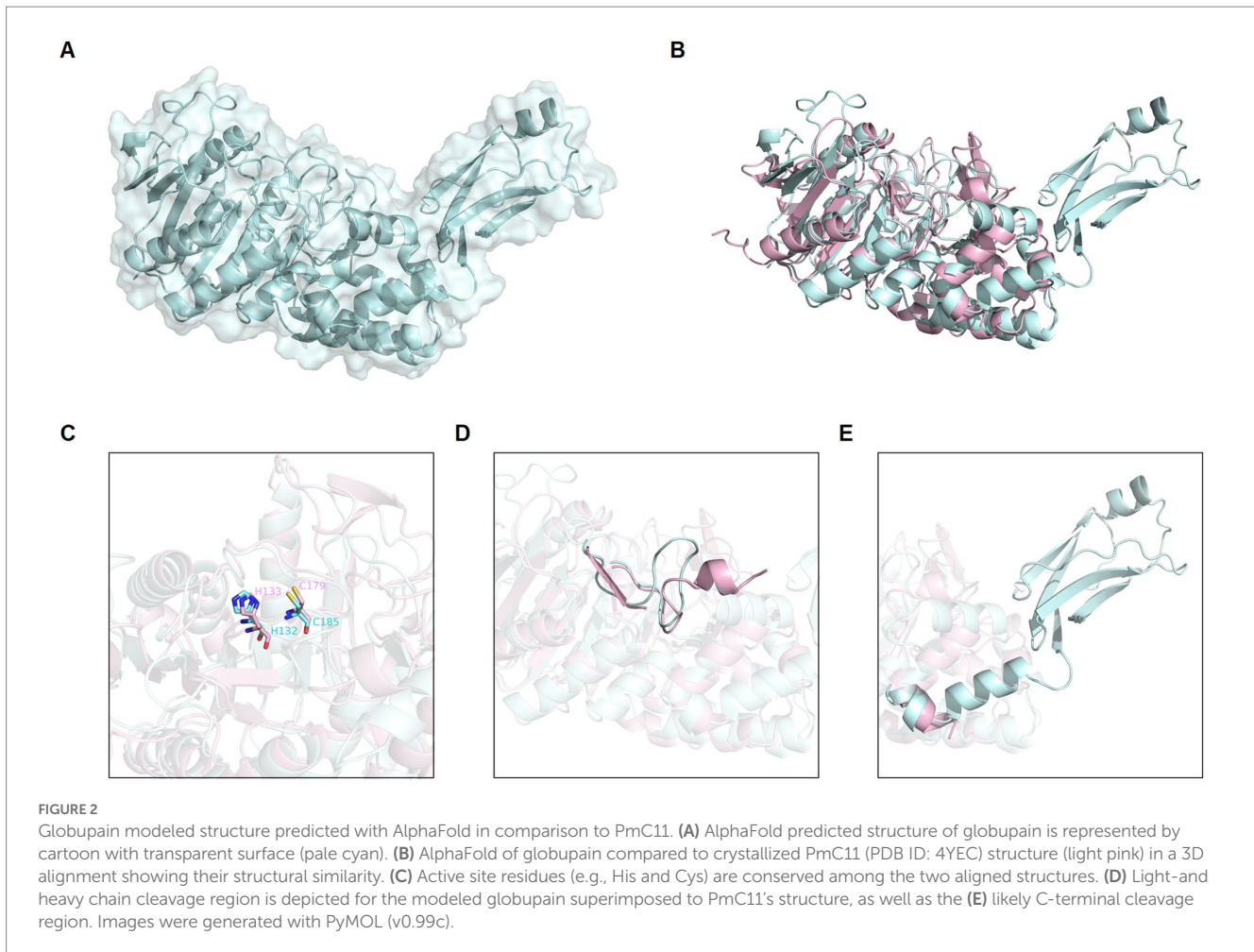
# 3. Results

## 3.1. Metagenomic globupain discovery

By conducting deep-sea hydrothermal *in situ* enrichments emended with targeted biomass, we have previously shown the induced shifts in community structure toward higher fractions of heterotrophic microorganisms (Stokke et al., 2020). Furthermore, *in silico* screening from derived metagenomes has shown a high potential for discovering novel enzymes (Fredriksen et al., 2019; Stepnov et al., 2019; Vuoristo et al., 2019; Arntzen et al., 2021). In the current study, a novel protease named globupain was identified from an *in situ* enriched metagenome and targeted for expression and characterization. The selected gene encoded a C11 protease and originated from a metagenome-assembled genome (MAG) classified as an uncharacterized genome within the genus *Archaeoglobus* (INS_M23_B45; SAMN33944460 INS_M23_B45; SAMN33944460). The putative polypeptide comprised 481 amino acids with a 21 amino acid N-terminal signal peptide (Figure 1). The estimated molecular mass after signal peptide removal was 52.0 kDa, and the pI was 4.2, as determined by the ProtParam tool (Gasteiger et al., 2005). The highest sequence identity scores against the MEROPS-MPRO database (Rawlings et al., 2016) were of the human gut and intestinal C11 members; clostripain of 23.5% (*C. histolyticum*), distapain of 27.3% (*Parabacteroides distasonis*), PmC11 of 24.2% (*Parabacteroides merdae*) and thetapain of 26.9% (*Bacteroides thetaiotaomicron*). Sequence alignments (Figure 1) indicated a conserved catalytic His/Cys dyad in globupain at positions 132 and 185, respectively. Moreover, in the globupain model obtained with AlphaFold (Figure 2A) deep learning-based algorithm (Jumper et al., 2021), structural similarities were observed between the predicted structure and the available PmC11 crystal structure (Figure 2B; PDB ID: 4YEC). The active site residues in PmC11 (e.g., $D_{177}$), including catalytic $H_{133}$ and $C_{179}$, were conserved between both structures (Figure 2C). The residues which did not overlay well with PmC11 include an area between the known heavy-and light chain of PmC11 (Figure 2D) and a long C-terminal region (Figure 2E).

## 3.2. Globupain activation

Globupain and substitution variants were expressed as soluble proteins in *E. coli* BL21-Gold (DE3) cells, with almost 100% of the total recombinant protein as soluble enzyme (Supplementary Figures 2, 3). The purified WT enzyme (Supplementary Figure 4) was produced as an inactive zymogen. However, incubation at 75°C for 4.5 h (Supplementary Figure 5) in the activation buffer resulted in an active form of the C11 globupain. SDS-PAGE imaging showed that the 52 kDa zymogen was cleaved into a 32 kDa heavy chain and a 12 kDa light chain (Figures 3A,B), forming a heterodimer stabilized by noncovalent bonding. Oligomeric structure analysis performed by

FIGURE 2
Globupain modeled structure predicted with AlphaFold in comparison to PmC11. **(A)** AlphaFold predicted structure of globupain is represented by cartoon with transparent surface (pale cyan). **(B)** AlphaFold of globupain compared to crystallized PmC11 (PDB ID: 4YEC) structure (light pink) in a 3D alignment showing their structural similarity. **(C)** Active site residues (e.g., His and Cys) are conserved among the two aligned structures. **(D)** Light-and heavy chain cleavage region is depicted for the modeled globupain superimposed to PmC11's structure, as well as the **(E)** likely C-terminal cleavage region. Images were generated with PyMOL (v0.99c).
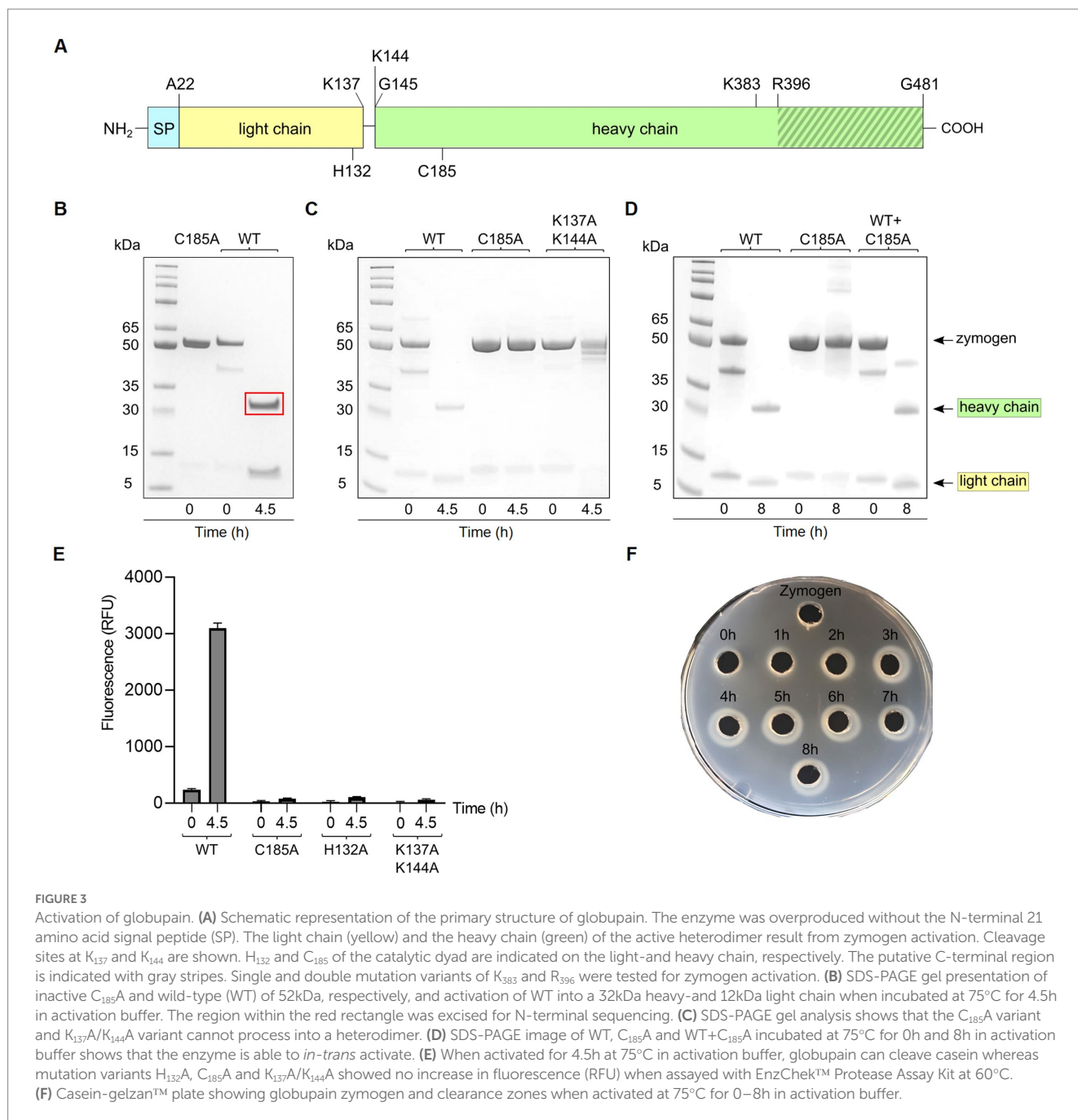
size-exclusion chromatography and analytical ultracentrifugation (AUC) revealed that globupain in zymogen form exists in solution as a homodimer (Supplementary Figures 6, 7). While the zymogen is inactive, the enzyme, after activation, can hydrolyze casein (Figures 3E,F). Globupain activation and its enzymatic activity, as in the case of other C11 proteases (Labrou and Rigden, 2004), depends on the presence of a His/Cys catalytic dyad in the primary protein sequence ($H_{132}/C_{185}$). $H_{132}$ of the light chain is responsible for deprotonating the neighboring $C_{185}$ in the heavy chain, which then promotes its nucleophilic attack on the substrate. For globupain, we found that substitution variant $C_{185}A$ generated by site-directed mutagenesis cannot be activated (Figure 3C). Also, no caseinolytic activity was observed for either $H_{132}A$ or $C_{185}A$ variants (Figure 3E).

Edman sequencing on a Shimadzu PPSQ-53A at the Iowa State University Protein Facility of the heavy chain revealed that the N-terminus consisted of $G_{145}$VCWD; hence cleavage (*) occurred between $K_{144}$ and $G_{145}$ within the sequence LPPIK*GVCWD (Figure 1). To further evaluate globupain autoprocessing at this cleavage site, a $K_{144}A$ variant was constructed. Notably, processing of the zymogen into this variant's heavy-and light chain still occurred with similar size of cleaved products as WT globupain, as visualized by SDS-PAGE (Supplementary Figure 8). To further assess if this result could be explained by cleavage after nearby Lys residues, 7 new variants were synthesized (Table 1; Supplementary Figure 8). Only the double ($K_{137}A/K_{144}A$) and triple ($K_{137}A/K_{139}A/K_{144}A$) mutants failed to activate into the processed form

(Figure 3C; Supplementary Figure 8) and remained catalytically inactive (Figure 3E), which altogether suggests that globupain can self-activate by cleavage after both $K_{137}$ and $K_{144}$ (Figure 1), respectively.

When mixing the WT zymogen with inactive $C_{185}A$ and performing the standard activation protocol, both the WT and $C_{185}A$ proteins were processed into the light-and heavy chain (Figure 3D). This finding demonstrates that globupain can activate *in-trans* and indicates that the sites for activation are exposed for cleavage by nearby proteases. Interestingly, the activation sites result in the removal of the unique region that poorly overlays with PmC11 (Figure 2D).

The combined molecular mass of the heavy-and light chain of activated globupain was determined to be 44 kDa, which, when compared to the 52.0 kDa zymogen (Figures 3A,B), indicates that additional autoprocessing occurs during activation. This discrepancy in molecular weight points to a likely cleavage in the C-terminal region, which the model supports (Figure 2E). Activated globupain (Supplementary Figure 9) failed to bind to the Ni²⁺ affinity column, and the C-terminal His tag of globupain was not detected on either light-or heavy chain (Supplementary Figure 10), altogether, revealing that a C-terminal fragment that contains the His tag was removed during autoprocessing. In-gel digest and subsequent proteomics of the 44 kDa protein band without C-terminal His tag upon activation (Supplementary Figure 10) revealed the most N-terminal and C-terminal tryptic peptides to be $I_{65}$DGYDDSYGNWTTAK$_{79}$L and $F_{384}$ATDTLWDEFLNR$_{396}$ (data can be found at ProteomeXChange:

FIGURE 3
Activation of globupain. **(A)** Schematic representation of the primary structure of globupain. The enzyme was overproduced without the N-terminal 21 amino acid signal peptide (SP). The light chain (yellow) and the heavy chain (green) of the active heterodimer result from zymogen activation. Cleavage sites at $K_{137}$ and $K_{144}$ are shown. $H_{132}$ and $C_{185}$ of the catalytic dyad are indicated on the light-and heavy chain, respectively. The putative C-terminal region is indicated with gray stripes. Single and double mutation variants of $K_{383}$ and $R_{396}$ were tested for zymogen activation. **(B)** SDS-PAGE gel presentation of inactive $C_{185}A$ and wild-type (WT) of 52kDa, respectively, and activation of WT into a 32kDa heavy-and 12kDa light chain when incubated at 75°C for 4.5h in activation buffer. The region within the red rectangle was excised for N-terminal sequencing. **(C)** SDS-PAGE gel analysis shows that the $C_{185}A$ variant and $K_{137}A/K_{144}A$ variant cannot process into a heterodimer. **(D)** SDS-PAGE image of WT, $C_{185}A$ and WT+$C_{185}A$ incubated at 75°C for 0h and 8h in activation buffer shows that the enzyme is able to *in-trans* activate. **(E)** When activated for 4.5h at 75°C in activation buffer, globupain can cleave casein whereas mutation variants $H_{132}A$, $C_{185}A$ and $K_{137}A/K_{144}A$ showed no increase in fluorescence (RFU) when assayed with EnzChek™ Protease Assay Kit at 60°C. **(F)** Casein-gelzan™ plate showing globupain zymogen and clearance zones when activated at 75°C for 0−8h in activation buffer.

PXD042411 or ftp://massive.ucsd.edu/MSV000092007/). This corresponds to a 332 amino acid protein fragment with estimated Mw of 37.67 kDa using the ProtParam tool (Gasteiger et al., 2005). Although this proteomics study cannot clarify the exact cleavage location, it suggests that the site for C-terminal processing occurs at $R_{396}$ or C-terminal of this site. From the sequence alignments (Figure 1) of globupain and several family C11 members, it is clear that $R_{396}$ does not represent a conserved Arg cleavage site for the respective enzymes. Further, overlay of the PmC11 crystal structure (Figures 2B,E) and the modeled globupain structure suggest that the processing occurs in the non-conserved structural region of the two enzymes. Moreover, manual inspection of the primary sequence suggested that $K_{383}$ and $R_{396}$ might be the putative cleavage sites. However, each of the enzyme variants $K_{383}A$, $R_{396}A$, and $K_{383}A/R_{396}A$

were still processed into the active form, and their C-terminal portion was removed (Supplementary Figure 11).

## 3.3. Substrate specificity (AMC) determination

To quantify globupain activity in a microwell plate assay, the enzyme was incubated with three substrates that were previously developed for another clostripain-like C11 family member known as PmC11 (Roncase et al., 2017). This enzyme was encoded in the *P. merdae* genome. The substrates consisted of tetrapeptides (VLXK) with an N-terminal acetyl group (Ac) and a C-terminal AMC reporter group. These substrates were chosen as the P1 residue corresponds to

the N-terminal auto-activation site of globupain. For PmC11, Ac-VLTK-AMC was most efficiently cleaved, followed by Ac-VLGK-AMC. However, the substitution of the P2 residue for a hydrophobic Val side chain ablated substrate turnover by PmC11. All three substrates were tested against globupain and found them to be cleaved at a similar rate (Figure 4A). This finding revealed that globupain has broader substrate specificity to PmC11 at the P2 position. Subsequently, 7 other substrates with Lys at P1 available in the laboratory were tested (Figure 4B). Globupain was able to cleave all substrates and cleaved Ac-AIK-AMC with highest efficiency. Interestingly, globupain cleaved each of the 7 new substrates more efficiently than the set of three initial substrates, which were based on the optimal substrate of PmC11 from *P. merdae*, indicating a distinct specificity to PmC11. The most statistically significant cleavage differences ($p < 0.01$) among the new 7 substrates occurred between Ac-AIK-AMC (the most efficient) and Ac-YK-AMC (least efficient). Activity with Ac-AIK-AMC was also statistically significantly increased ($p < 0.01$) compared to Ac-AEIK-AMC. We hypothesized that the broad enzymatic activity of globupain for degrading casein may be partially due to cleavage following a structurally similar amino acid, arginine. Therefore, fluorogenic substrates were examined with Arg as the P1 residue (Figure 4C). This screen of 7 additional substrates revealed that globupain cleaved substrates with even higher efficiency than the previous best with Lys at P1. However, some substrates, such as z-RR-AMC and Pyr-RTKR-AMC, showed minimal cleavage, indicating that globupain may favor non-polar amino acids at the P2 position. Out of the 17 AMC substrates that were tested in this study, it was clear that globupain had a strong preference for Arg in the P1 position, with Boc-QAR-AMC as the best substrate.
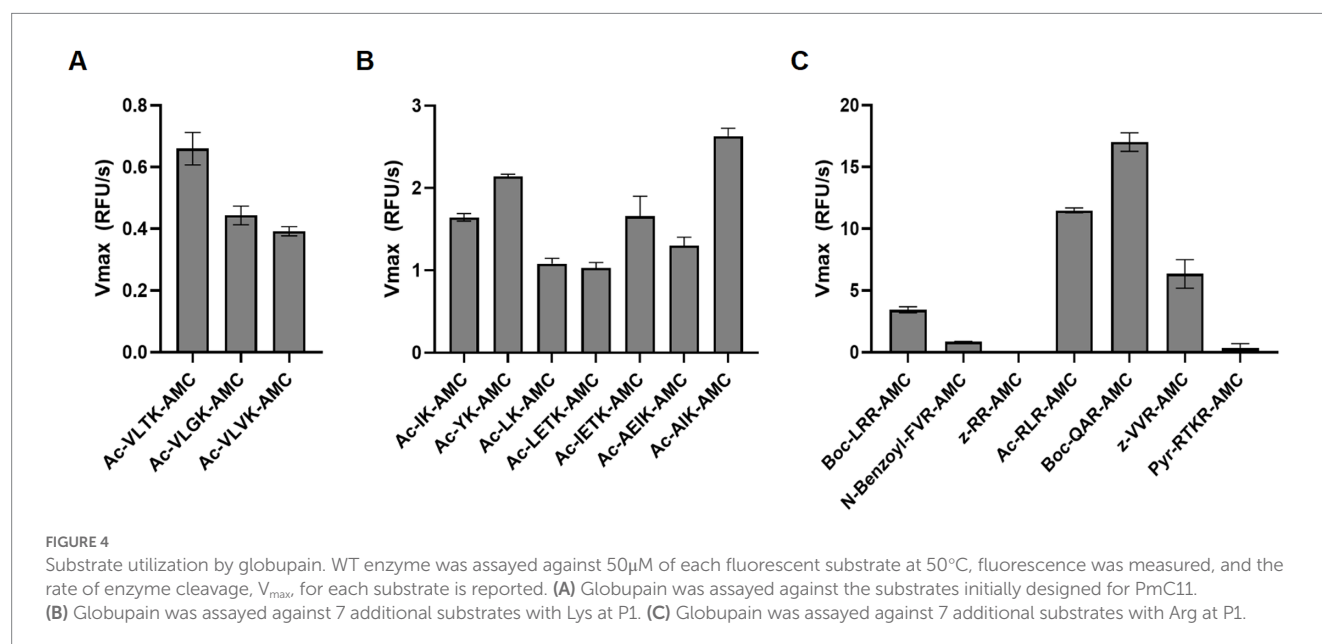
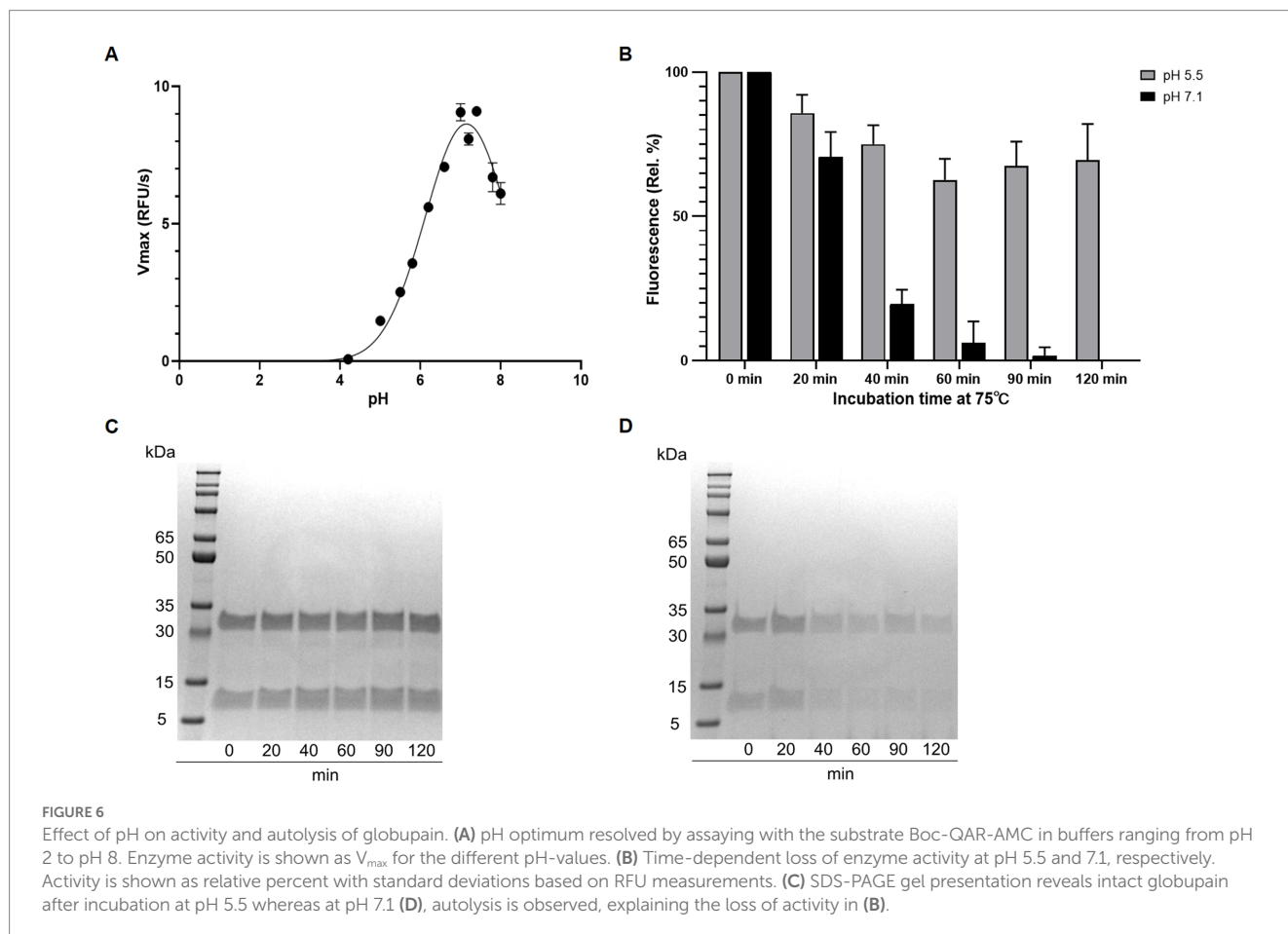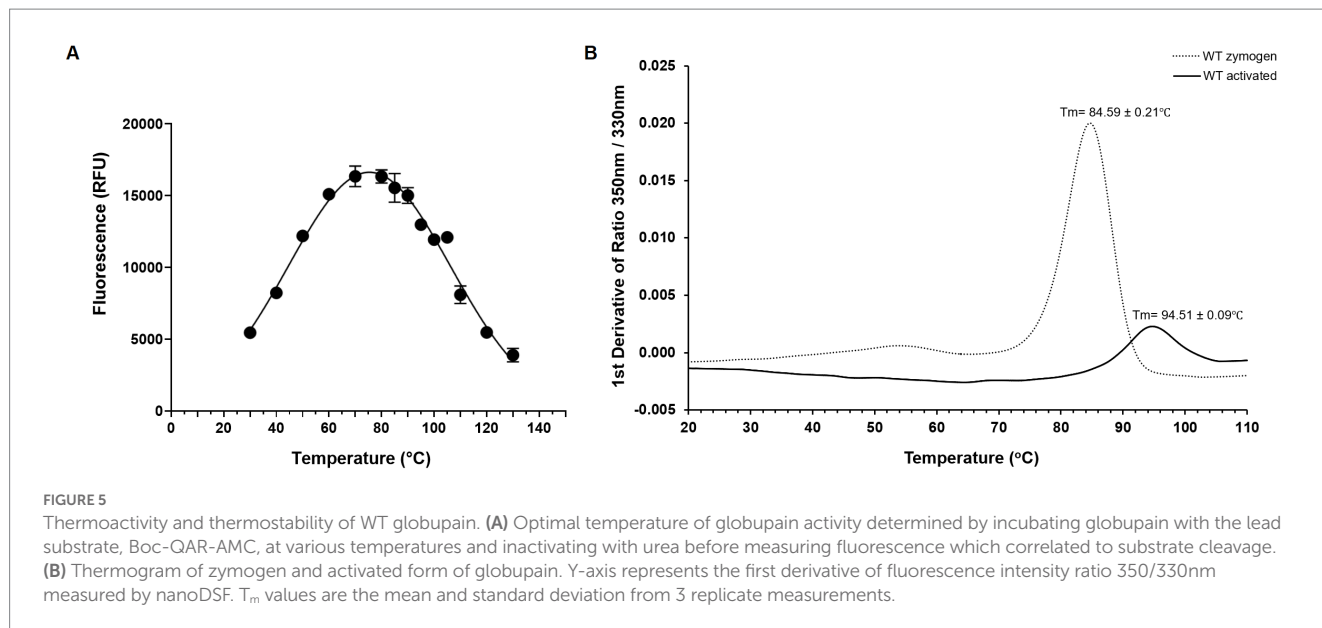## 3.4. Thermal stability, optimal temperature, and pH

Using the Boc-QAR-AMC substrate, the temperature optimum of globupain was determined to be $75.4°C \pm 0.56°C$ and remained 90% active at 60°C and 90°C (Figure 5A). Thermal stability of inactive globupain as characterized by melting temperature, which indicates the point at which half the protein is unfolded was $84.59°C \pm 0.21°C$. The activated heterodimer's melting temperature was $94.51°C \pm 0.09°C$ (Figure 5B). Finally, the optimum pH of globupain using the Boc-QAR-AMC substrate was evaluated. The optimum pH for catalytic activity was calculated to be pH 7.1 (Figure 6A). Initial Analysis of Covariance (ANCOVA) indicated that there was a significant effect of pH on fluorescence after controlling for time, $F(3,26) = 42.85$, $p < 0.05$, $R^2 = 83.18\%$. A follow-up *post hoc* Tukey's Honest Significance Difference Test (HSD) indicated that pH 7.1 had a stronger effect on decreasing RFU over time relative to the effect measured on RFU at pH 5.5, $p < 0.05$ (Figure 6B). Thus, while the optimum pH is higher than the pH used for activation, the enzyme was shown to be more stable against autolysis at pH 5.5 than at pH 7.1 (Figures 6C,D) which supported our use of pH 5.5 buffers for the biochemical characterization of globupain.

## 4. Discussion

In this study, we characterized the novel cysteine protease, globupain belonging to enzyme family C11. Globupain was prospected from metagenomic data assigned to an unclassified *Archaeoglobus* species from the Arctic Mid-Ocean Ridge vent fields. The enzyme was highly soluble, expressing at relatively high concentrations in *E. coli* (Supplementary Figure 4). Two protein bands (52 kDa and 40 kDa) with intact C-terminal His tag were visualized on SDS-PAGE gels after protein purification (Supplementary Figure 10). The zymogen of globupain is processed in the N-terminal region at $K_{137}$ and $K_{144}$ to yield a heavy-and light chain when exposed to activation conditions. Similar to clostripain from *C. histolyticum* (Kembhavi et al., 1991), globupain requires calcium and a reducing environment for activation (Supplementary Figure 5). This condition contrasts the C11 protease, PmC11 from *P. merdae*, which activates independently of calcium (McLuskey et al., 2016). When activated, the globupain enzyme cleaves



FIGURE 4
Substrate utilization by globupain. WT enzyme was assayed against 50 µM of each fluorescent substrate at 50°C, fluorescence was measured, and the rate of enzyme cleavage, $V_{max}$, for each substrate is reported. **(A)** Globupain was assayed against the substrates initially designed for PmC11. **(B)** Globupain was assayed against 7 additional substrates with Lys at P1. **(C)** Globupain was assayed against 7 additional substrates with Arg at P1.

FIGURE 5
Thermoactivity and thermostability of WT globupain. **(A)** Optimal temperature of globupain activity determined by incubating globupain with the lead substrate, Boc-QAR-AMC, at various temperatures and inactivating with urea before measuring fluorescence which correlated to substrate cleavage. **(B)** Thermogram of zymogen and activated form of globupain. Y-axis represents the first derivative of fluorescence intensity ratio 350/330nm measured by nanoDSF. $T_m$ values are the mean and standard deviation from 3 replicate measurements.



FIGURE 6
Effect of pH on activity and autolysis of globupain. **(A)** pH optimum resolved by assaying with the substrate Boc-QAR-AMC in buffers ranging from pH 2 to pH 8. Enzyme activity is shown as $V_{max}$ for the different pH-values. **(B)** Time-dependent loss of enzyme activity at pH 5.5 and 7.1, respectively. Activity is shown as relative percent with standard deviations based on RFU measurements. **(C)** SDS-PAGE gel presentation reveals intact globupain after incubation at pH 5.5 whereas at pH 7.1 **(D)**, autolysis is observed, explaining the loss of activity in **(B)**.

off a C-terminal region which our proteomic analysis indicated that it was at $R_{396}$ or C-terminal to this site. This kind of autoprocessing is not uncommon for C11 proteases; for example, activation of clostripain starts with a 23 amino acid pro-peptide removal (Dargatz et al., 1993). The two cut sites of globupain at $K_{137}$ and $K_{144}$, leads to the removal of a

7-amino acid linker sequence and the formation of a heterodimer consisting of a heavy-and light chain. For clostripain, a linker peptide is removed by cleavage at two Arg sites (Gilles et al., 1979; Dargatz et al., 1993). When activated, globupain showed the ability to *in-trans* activate and implied that the cut sites ($K_{137}$ and $K_{144}$) are exposed to proteolytic

cleavage by neighboring proteases. This kind of activation is known to occur for several C11 enzymes such as thetapain (Roncase et al., 2019), fragipain (Herrou et al., 2016), and distapain (González-Páez et al., 2019) and contrasts PmC11 which activates only *in-cis* (Roncase et al., 2017). Globupain showed maximum activity at pH 7.1. This value is in the same range as known pH optima of PmC11 (pH 8.0), clostripain (pH 7.4–7.8), and thetapain (pH 7.4), respectively (Ogle and Tytell, 1953; Mitchell and Harrington, 1968; Roncase et al., 2017, 2019). However, globupain showed an optimum temperature of 75°C and matures into a heat tolerant enzyme, which allows it to function in its thermal environment (Dahle et al., 2015). The observed thermal properties are in line with the growth characteristics of cultivated species within the genus *Archaeoglobus* (Stetter, 1988; Burggraf et al., 1990; Huber et al., 1997; Mori et al., 2008; Steinsbu et al., 2010; Slobodkina et al., 2021) and enzymes characterized previously (Steen et al., 2001). Moreover, in comparison to well-characterized industrially relevant marine thermostable proteases, the thermal tolerance of globupain is superior to proteases sourced from marine *Bacillus* species and in the same range as of proteases from (hyper)thermophilic archaea (Barzkar et al., 2018).

Active clostripain-like proteases have been identified in marine sediment archaea (Lloyd et al., 2013). However, the highest sequence similarity scores of globupain using the MEROPS-MPRO database (Rawlings et al., 2016) were C11 proteases that originate from bacteria such as *C. histolyticum, P. distasonis, P. merdae, B. thetaiotaomicron* that have been found in the human intestinal microbiota (Salyers, 1984; Johnson et al., 1986; Franks et al., 1998). Some of these bacteria have been reported to cause disease and/or affect human health and have been studied to a greater extent (Salyers, 1984; Parracho et al., 2005; McLuskey et al., 2016; Roncase et al., 2017, 2019; Ezeji et al., 2021). This finding highlights the significance of acquiring greater knowledge of marine C11 proteases. Notably, all C11 proteases, including globupain, show a conserved His/Cys catalytic dyad by sequence alignment. Moreover, the catalytic residues were also conserved in the globupain model obtained with AlphaFold (Jumper et al., 2021). Finally, it was shown experimentally using site-directed mutagenesis that in globupain, $H_{132}$ and $C_{185}$ were critical for activation and activity. When assayed against several AMC substrates, the enzyme showed a clear preference for the substrate Boc-QAR-AMC. Preference for hydrolyzing Arg bonds in the P1 position is a known trait for C11 members (Ogle and Tytell, 1953; Labrou and Rigden, 2004). Globupain showed much lower activity against the Ac-VLTK-AMC substrate, which both PmC11 and thetapain hydrolyze efficiently (Roncase et al., 2017, 2019). This observation indicates that the substrate specificity may vary substantially between different C11 proteases despite having sequence and structural similarities around the active site. In conclusion, the revealed temperature tolerance and catalytic properties of globupain render it as a promising protease in diverse industrial and biotechnology sectors. Further studies focused on in-depth knowledge of the substrate specificity (O'Donoghue et al., 2012; Rohweder et al., 2023), effects of protease inhibitors, resistance to organic solvents and chemical denaturants may provide a deeper understanding of the applicability of globupain.

## Data availability statement

The datasets presented in this study are deposited in the NCBI online repository, under accession numbers PRJNA296938 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA296938), OQ718499 (https://www.ncbi.nlm.nih.gov/nuccore/OQ718499.1/), SAMN04111445 (https://www.ncbi.nlm.nih.gov/biosample/SAMN04111445/), and JARQZL000000000 (https://www.ncbi.nlm.nih.gov/nuccore/JARQZL000000000.1/).

## Author contributions

VR, AO'D, and IHS conceived the study. VR, BH, and IHS wrote the manuscript. VR, BH, A-KK, SD, A-EF, HA, MSMS, SM, OW, TK, and RS performed the experiments. All authors contributed to the article and approved the submitted version.

## Funding

## Acknowledgments

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1199085/full#supplementary-material

# References

Alneberg, J., Bjarnason, B., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103

Arntzen, M. Ø., Pedersen, B., Klau, L. J., Stokke, R., Oftebro, M., Antonsen, S. G., et al. (2021). Alginate degradation: insights obtained through characterization of a thermophilic exolytic alginate lyase. *Appl. Environ. Microbiol.* 87:6. doi: 10.1128/AEM.02399-20

Barrett, A. J., and Rawlings, N. D. (1996). Families and clans of cysteine peptidases. *Perspect. Drug discov. Des.* 6, 1–11. doi: 10.1007/BF02174042

Barzkar, N., Homaei, A., Hemmati, R., and Patel, S. (2018). Thermostable marine microbial proteases for industrial applications: scopes and risks. *Extremophiles* 22, 335–346. doi: 10.1007/s00792-018-1009-8

Birkeland, N., Ånensen, H., Knævelsrud, I., Kristoffersen, W., Bjørås, M., Robb, F. T., et al. (2002). Methylpurine DNA glycosylase of the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *Biochemistry* 41, 12697–12705. doi: 10.1021/bi020334w

Brautigam, C. A. (2015). Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol.* 562, 109–133. doi: 10.1016/bs.mie.2015.05.001

Burggraf, S., Jannasch, H. W., Nicolaus, B., and Stetter, K. O. (1990). *Archaeoglobus profundus* sp. nov., represents a new species within the sulfate-reducing Archaebacteria. *Syst. Appl. Microbiol.* 13, 24–28. doi: 10.1016/S0723-2020(11)80176-1

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38, 5315–5316. doi: 10.1093/bioinformatics/btac672

Cheng, J. H., Wang, Y., Zhang, X. Y., Sun, M. L., Zhang, X., Song, X. Y., et al. (2021). Characterization and diversity analysis of the extracellular proteases of thermophilic Anoxybacillus caldiproteolyticus 1A02591 from deep-sea hydrothermal vent sediment. *Front. Microbiol.* 12:643508. doi: 10.3389/fmicb.2021.643508

Dahle, H., Økland, I., Thorseth, I. H., Pederesen, R. B., and Steen, I. H. (2015). Energy landscapes shape microbial communities in hydrothermal systems on the Arctic Mid-Ocean ridge. *ISME J.* 9, 1593–1606. doi: 10.1038/ismej.2014.247

Dargatz, H., Diefenthal, T., Witte, V., Reipen, G., and von Wettstein, D. (1993). The heterodimeric protease clostripain from *Clostridium histolyticum* is encoded by a single gene. *Mol. Gen. Genet.* 240, 140–145. doi: 10.1007/BF00276893

DeLano, W. L. (2006). *The PyMOL molecular graphics system*. San Carlos, CA, USA: DeLano Scientific LLC.

Dombrowski, N., Teske, A. P., and Baker, B. J. (2018). Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* 9:4999. doi: 10.1038/s41467-018-07418-0

Dombrowski, N., Williams, T. A., Sun, J., Woodcroft, B. J., Lee, J. H., Minh, B. Q., et al. (2020). Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* 11:3939. doi: 10.1038/s41467-020-17408-w

Ezeji, J. C., Sarikonda, D. K., Hopperton, A., Erkkila, H. L., Cohen, D. E., Martinez, S. P., et al. (2021). *Parabacteroides distasonis*: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes* 13:1. doi: 10.1080/19490976.2021.1922241

Franks, A. H., Harmsen, H. J. M., Raangs, G. C., Jansen, G. J., Schut, F., and Welling, G. W. (1998). Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Appl. Environ. Microbiol.* 64, 3336–3345. doi: 10.1128/AEM.64.9.3336-3345.1998

Fredriksen, L., Stokke, R., Jensen, M. S., Westereng, B., Jameson, J. K., Steen, I. H., et al. (2019). Discovery of a thermostable GH10 xylanase with broad substrate specificity from the Arctic Mid-Ocean ridge vent system. *Appl. Environ. Microbiol.* 85:6. doi: 10.1128/AEM.02970-18

García-Moyano, A., Diaz, Y., Navarro, J., Almendral, D., Puntervoll, P., Ferrer, M., et al. (2021). Two-step functional screen on multiple proteinaceous substrates reveals temperature-robust proteases with a broad-substrate range. *Appl. Microbiol. Biotechnol.* 105, 3195–3209. doi: 10.1007/s00253-021-11235-9

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). Protein Identification and Analysis Tools on the ExPASy Server., in Walker, J.M. (ed.) (Totowa, NJ, USA: The Proteomics Protocols Handbook. Humana Press Inc.), 571–607.

Gilles, A.-M., Imhoff, J.-M., and Keil, B. (1979). α-Clostripain. Chemical characterization, activity, and thiol content of the highly active form of clostripain*. *J. Biol. Chem.* 254, 1462–1468.

Gimenes, N. C., Silveira, E., and Tambourgi, E. B. (2021). An overview of proteases: production, downstream processes and industrial applications. *Sep. Purif. Rev.* 50, 223–243. doi: 10.1080/15422119.2019.1677249

González-Páez, G. E., Roncase, E. J., and Wolan, D. W. (2019). X-ray structure of an inactive zymogen clostripain-like protease from *Parabacteroides distasonis*. *Acta Crystallogr. D Biol. Crystallogr.* 75, 325–332. doi: 10.1107/S2059798319000809

Herrou, J., Choi, V. M., Bubeck Wardenburg, J., and Crosson, S. (2016). Activation mechanism of theBacteroides fragilisCysteine peptidase, Fragipain. *Biochemistry* 55, 4077–4084. doi: 10.1021/acs.biochem.6b00546

Huber, H., Jonnasch, H., Rachel, R., Fuchs, T., and Stetter, K. O. (1997). *Archaeoglobus veneficus* sp. nov. a novel facultative chemolithoautotrophic hyperthermophilic sulfite reducer, isolated from abyssal black smokers. *Syst. Appl. Microbiol.* 20, 374–380.

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Isupov, M. N., Boyko, K. M., Sutter, J.-M., James, P., Sayer, C., Schmidt, M., et al. (2019). Thermostable branched-chain amino acid transaminases from the archaea Geoglobus acetivorans and *Archaeoglobus fulgidus*: biochemical and structural characterization. *Front. Bioeng. Biotechnol.* 7:7. doi: 10.3389/fbioe.2019.00007

Johnson, N. L., Moore, W. E. C., and Moore, L. V. H. (1986). *Bacteroides caccae* sp. nov., Bacteroides merdae sp. nov., and *Bacteroides stercoris* sp. nov. isolated from human feces. *Int. J. Bacteriol.* 36, 499–501. doi: 10.1099/00207713-36-4-499

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with alpha fold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *Peer J.* 3:e1165. doi: 10.7717/peerj.1165

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 7:e7359. doi: 10.7717/peerj.7359

Kembhavi, A. A., Buttle, D. J., Rauber, P., and Barrett, A. J. (1991). Clostripain: characterization of the active site. *FEBS Lett.* 283, 277–280. doi: 10.1016/0014-5793(91)80607-5

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., et al. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390, 364–370. doi: 10.1038/37052

Knævelsrud, I., Moen, M., Grøsvik, K., Haugland, G. T., Birkeland, N. K., Klungland, A., et al. (2010). The Hyperthermophilic euryarchaeon *Archaeoglobus fulgidus* repairs uracil by single-nucleotide replacement. *J. Bacteriol.* 192, 5755–5766. doi: 10.1128/JB.00135-10

Kumar, S., Tsai, C.-J., and Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Eng.* 13, 179–191. doi: 10.1093/protein/13.3.179

Kuwabara, T., Minaba, M., Ogi, N., and Kamekura, M. (2007). *Thermococcus celericrescens* sp. nov., a fast-growing and cell-fusing hyperthermophilic archaeon from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 57, 437–443. doi: 10.1099/ijs.0.64597-0

Labrou, N. E., and Rigden, D. J. (2004). The structure-function relationship in the clostripain family of peptidases. *Eur. J. Biochem.* 271, 983–992. doi: 10.1111/j.1432-1033.2004.04000.x

Li, M., Baker, B. J., Anantharaman, K., Jain, S., Breier, J. A., and Dick, G. J. (2015). Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat. Commun.* 6:8933. doi: 10.1038/ncomms9933

Lloyd, K. G., Schreiber, L., Petersen, D. G., Kjeldsen, K. U., Lever, M. A., Steen, A. D., et al. (2013). Predominant archaea in marine sediments degrade detrital proteins. *Nature* 496, 215–218. doi: 10.1038/nature12033

Maciejewski, M. W., Schuyler, A. D., Gryk, M. R., Moraru, I. I., Romero, P. R., Ulrich, E. L., et al. (2017). NMRbox: a resource for biomolecular NMR computation. *Biophysics* 112, 1529–1534. doi: 10.1016/j.bpj.2017.03.011

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., et al. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50, W276–W279. doi: 10.1093/nar/gkac240

Madern, D., Ebel, C., Dale, H. A., Lien, T., Steen, I. H., Birkeland, N. K., et al. (2001). Differences in the Oligomeric states of the LDH-likel-MalDH from the Hyperthermophilic ArchaeaMethanococcus Jannaschiiandarchaeoglobus fulgidus. *Biochemistry* 40, 10310–10316. doi: 10.1021/bi010168c

Maisnier-Patin, S., Malandrin, L., Birkeland, N. K., and Bernander, R. (2002). Chromosome replication patterns in the hyperthermophilic euryarchaeal Archaeoglobus fulgidus and Methanocaldococcus (Methanococcus) jannaschii. *Mol. Microbiol.* 45, 1443–1450. doi: 10.1046/j.1365-2958.2002.03111.x

McIlvaine, T. C. (1921). A buffer solution for colorimetric comparison. *J. Biol. Chem.* 49, 183–186. doi: 10.1016/s0021-9258(18)86000-8

McLuskey, K., Grewal, J. S., das, D., Godzik, A., Lesley, S. A., Deacon, A. M., et al. (2016). Crystal structure and activity studies of the C11 cysteine peptidase from

*Parabacteroides merdae* in the human gut microbiome. *J. Biol. Chem.* 291, 9482–9491. doi: 10.1074/jbc.M115.706143

Mitchell, W. M., and Harrington, W. F. (1968). Purification and properties of Clostridiopeptidase B (Clostripain)*. *J. Biol. Chem.* 243, 4683–4692. doi: 10.1016/S0021-9258(18)93173-X

Mori, K., Maruyama, A., Urabe, T., Suzuki, K. I., and Hanada, S. (2008). *Archaeoglobus infectus* sp. nov., a novel thermophilic, chemolithoheterotrophic achaeon isolated from a deep-sea rock collected at Suikyo seamount, Izu-Bonin arc, western Pacific Ocean. *Int. J. Syst. Evol. Microbiol.* 58, 810–816. doi: 10.1099/ijs.0.65422-0

Myers, S. A., Rhoads, A., Cocco, A. R., Peckner, R., Haber, A. L., Schweitzer, L. D., et al. (2019). Streamlined protocol for deep proteomic profiling of FAC-sorted cells and its application to freshly isolated murine immune cells. *Mol. Cell. Proteomics* 18, 995–1009. doi: 10.1074/mcp.RA118.001259

Nunoura, T., Oida, H., Miyazaki, M., and Suzuki, Y. (2008). *Thermosulfidibacter takaii* gen. Nov., sp. nov., a thermophilic, hydrogen-oxidizing, sulfur-reducing chemolithoautotroph isolated from a deep-sea hydrothermal field in the southern Okinawa trough. *Int. J. Syst. Evol. Microbiol.* 58, 659–665. doi: 10.1099/ijs.0.65349-0

O'Donoghue, A. J., Eroy-Reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M., Statnekov, J. B., et al. (2012). Global identification of peptidase specificity by multiplex substrate profiling. *Nat. Methods* 9, 1095–1100. doi: 10.1038/nmeth.2182

Ogle, J. D., and Tytell, A. A. (1953). The activity of *Clostridium histolyticum* proteinase on synthetic substrates. *Arch. Biochem. Biophys.* 42, 327–336. doi: 10.1016/0003-9861(53)90362-2

Palombarini, F., di Fabio, E., Boffi, A., Macone, A., and Bonamore, A. (2020). Ferritin nanocages for protein delivery to tumor cells. *Molecules* 25:825. doi: 10.3390/molecules25040825

Parks, D. H., Chuvochina, M., Chaumeil, P. A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* 38, 1079–1086. doi: 10.1038/s41587-020-0501-8

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P. A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794. doi: 10.1093/nar/gkab776

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). Check M: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Parracho, H. M. R. T., Bingham, M. O., Gibson, G. R., and McCartney, A. L. (2005). Differences between the gut microflora of children with autistic spectrum disorders and that of healthy children. *J. Med. Microbiol.* 54, 987–991. doi: 10.1099/jmm.0.46101-0

Pedersen, R. B., Thorseth, I. H., Hellevang, B., Schultz, A., Taylor, P., Knudsen, H. P., et al. (2005). Two vent fields discovered at the ultraslow spreading Arctic ridge system. *Eos. Trans. AGU* 86:52. Fall Meet. Suppl., Abstract OS21C-01.

Pedersen, R. B., Thorseth, I. H., Nygård, T. E., Lilley, M. D., and Kelly, D. S. (2010). "Hydrothermal activity at the Arctic Mid-Ocean ridges," in P. A. Rona, C. W. Devey, J. Dyment, B. J. Murton (eds) (Washington, DC, USA: Diversity of hydrothermal systems on slow spreading ocean ridges. American Geophysical Union), 67–89.

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. doi: 10.1038/nmeth.1701

Pikuta, E. V., Marsic, D., Itoh, T., Bej, A. K., Tang, J., Whitman, W. B., et al. (2007). *Thermococcus thioreducens* sp. nov., a novel hyperthermophilic, obligately sulfur-reducing archaeon from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 57, 1612–1618. doi: 10.1099/ijs.0.65057-0

Rawlings, N. D., Barrett, A. J., and Finn, R. D. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 44, D343–D350. doi: 10.1093/nar/gkv1118

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632. doi: 10.1093/nar/gkx1134

Rawlings, N. D., and Bateman, A. (2019). Origins of peptidases. *Biochimie* 166, 4–18. doi: 10.1016/j.biochi.2019.07.026

Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42, D503–D509. doi: 10.1093/nar/gkt953

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316

Rohweder, P. J., Jiang, Z., Hurysz, B. M., O'Donoghue, A. J., and Craik, C. S. (2023). Multiplex substrate profiling by mass spectrometry for proteases. *Methods Enzymol.* 682, 375–411. doi: 10.1016/bs.mie.2022.09.009

Roncase, E. J., González-Páez, G. E., and Wolan, D. W. (2019). X-ray structures of two *Bacteroides thetaiotaomicron* C11 proteases in complex with peptide-based inhibitors. *Biochem.* 58, 1728–1737. doi: 10.1021/acs.biochem.9b00098

Roncase, E. J., Moon, C., Chatterjee, S., González-Páez, G. E., Craik, C. S., O'Donoghue, A. J., et al. (2017). Substrate profiling and high resolution co-complex crystal structure of a secreted C11 protease conserved across commensal bacteria. *ACS Chem. Biol.* 12, 1556–1565. doi: 10.1021/acschembio.7b00143

RStudio (2023). Integrated development for R. RStudio. Available at: http://www.rstudio.com/

Salyers, A. A. (1984). Bacteroides of the human lower intestinal tract. *Annu. Rev. Microbiol.* 38, 293–313. doi: 10.1146/annurev.mi.38.100184.001453

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* 78, 1606–1619. doi: 10.1016/S0006-3495(00)76713-0

Slobodkina, G., Allioux, M., Merkel, A., Cambon-Bonavita, M. A., Alain, K., Jebbar, M., et al. (2021). Physiological and genomic characterization of a hyperthermophilic archaeon Archaeoglobus neptunius sp. nov. isolated from a deep-sea hydrothermal vent warrants the reclassification of the genus Archaeoglobus. *Front. Microbiol.* 12:679245. doi: 10.3389/fmicb.2021.679245

Steen, I. H., Madern, D., Karlström, M., Lien, T., Ladenstein, R., and Birkeland, N. K. (2001). Comparison of isocitrate dehydrogenase from three hyperthermophiles reveals differences in thermostability, cofactor specificity, oligomeric state, and phylogenetic affiliation. *J. Biol. Chem.* 276, 43924–43931. doi: 10.1074/jbc.M105999200

Steinsbu, B. O., Thorseth, I. H., Nakagawa, S., Inagaki, F., Lever, M. A., Engelen, B., et al. (2010). *Archaeoglobus sulfaticallidus* sp. nov., a thermophilic and facultatively lithoautotrophic sulfate-reducer isolated from black rust exposed to hot ridge flank crustal fluids. *Int. J. Syst. Evol. Microbiol.* 60, 2745–2752. doi: 10.1099/ijs.0.016105-0

Stepnov, A. A., Fredriksen, L., Steen, I. H., Stokke, R., and Eijsink, V. G. H. (2019). Identification and characterization of a hyperthermophilic GH9 cellulase from the Arctic Mid-Ocean ridge vent field. *PLoS One* 14:e0222216. doi: 10.1371/journal.pone.0222216

Stetter, K. O. (1988). *Archaeoglobus fulgidus* gen. Nov., sp.nov.: a new taxon of extremely thermophilic archaebacteria. *Syst. Appl. Microbiol.* 10, 172–173. doi: 10.1016/S0723-2020(88)80032-8

Stokke, R., Reeves, E. P., Dahle, H., Fedøy, A. E., Viflot, T., Lie Onstad, S., et al. (2020). Tailoring hydrothermal vent biodiversity towards improved biodiscovery using a novel in-situ enrichment strategy. *Front. Microbiol.* 11:249. doi: 10.3389/fmicb.2020.00249

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025. doi: 10.1038/s41587-021-01156-3

Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018). Meta WRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome.* 6:158. doi: 10.1186/s40168-018-0541-1

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). Alpha fold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. doi: 10.1093/nar/gkab1061

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. (2019). *Efficient architecture-aware acceleration of BWA-MEM for multicore systems.* IEEE (IPDPS), Rio de Janeiro, Brazil.

Vuoristo, K. S., Fredriksen, L., Oftebro, M., Arntzen, M. Ø., Aarstad, O. A., Stokke, R., et al. (2019). Production, characterization, and application of an alginate lyase, AMOR_PL7A, from hot vents in the Arctic Mid-Ocean ridge. *J. Agric. Food Chem.* 67, 2936–2945. doi: 10.1021/acs.jafc.8b07190

Ward, O. P. (2011). Proteases. *Comprehens Biotechnol.* 2011, 604–615. doi: 10.1016/B978-0-08-088504-9.00222-1

Wu, Y. W., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. doi: 10.1093/bioinformatics/btv638