

Supplementary Information

Environmental Activity-Based Protein Profiling for Function-Driven Enzyme Discovery from Natural Communities

Sabrina Ninck^{1†*}, Thomas Klaus^{2†}, Tatiana V. Kochetkova³, Sarah P. Esser⁴, Leonard Sewald¹, Farnusch Kaschani¹, Christopher Bräsen², Alexander J. Probst⁴, Ilya V. Kublanov³, Bettina Siebers^{2*}, Markus Kaiser^{1*}

¹Chemical Biology, Centre of Medical Biotechnology (ZMB), Faculty of Biology, University of Duisburg-Essen, Universitätsstr. 2, 45117 Essen, Germany.

²Molecular Enzyme Technology and Biochemistry, Environmental Microbiology and Biotechnology (EMB), Centre for Water and Environmental Research (CWE), Faculty of Chemistry, University of Duisburg-Essen, Universitätsstr. 5, 45117 Essen, Germany.

³Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Prospekt 60-Let Oktyabrya 7-2, 117312 Moscow, Russia.

⁴Group for Aquatic Microbial Ecology, Environmental Microbiology and Biotechnology (EMB), Faculty of Chemistry, University of Duisburg-Essen, Universitätsstr. 5, 45117 Essen, Germany.

*Correspondence

Sabrina Ninck

sabrina.ninck@uni-due.de

Bettina Siebers

bettina.siebers@uni-due.de

Markus Kaiser

markus.kaiser@uni-due.de

†S.N. and T.K. contributed equally to this work

Table of Contents

Supplementary Figures	3
Figure S1. Bioinformatics pipeline.....	3
Figure S2. Biochemical characterization of the heterologously expressed esterase.....	4
Supplementary Tables	5
Table S1. Relative abundance of microorganisms in KAM3811 based on the coverage of scaffolds containing the respective rpS3 gene	5
Table S2. Relative abundance of microorganisms in KAM3808 based on the coverage of scaffolds containing the respective rpS3 gene	6
Table S3. Selected structural homologs of the putative esterase determined with HHpred	8
Experimental section	10
Sample Collection and Chemical Labeling	10
DNA Extraction and Metagenomic Sequencing	10
Genome-Resolved Metagenomics	10
Protein Extraction and Clean-Up.....	11
Click Reaction and Affinity Purification.....	12
On-Bead Digestion of Captured Proteins	12
Sample Clean-Up for LC–MS	13
LC-MS/MS Analysis.....	13
Peptide and Protein Identification Using MaxQuant and Perseus.....	14
Bioinformatic Analyses of Enriched Proteins	16
Heterologous Protein Expression and Purification.....	16
Structural and Bioinformatics Analysis of the Putative Esterase	17
Biochemical Characterization of the Heterologous Protein	17
<i>In vitro</i> labeling	17
Supplementary References	18

Supplementary Figures

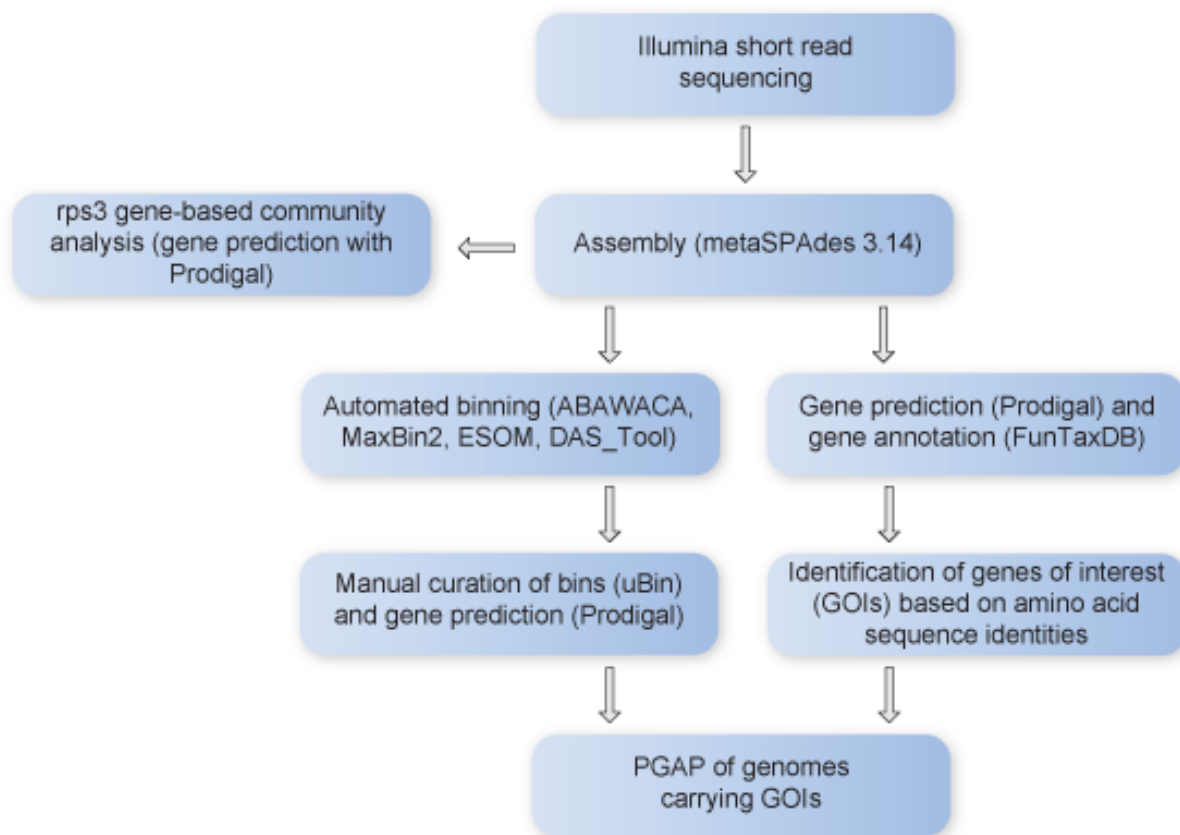


Figure S1. Bioinformatics pipeline.

Visualization of the bioinformatics pipeline described in the Method section. The pipeline starts with sequencing, followed by quality-filtering and assembly, to building the metaproteome database and the identification of genes of interest.

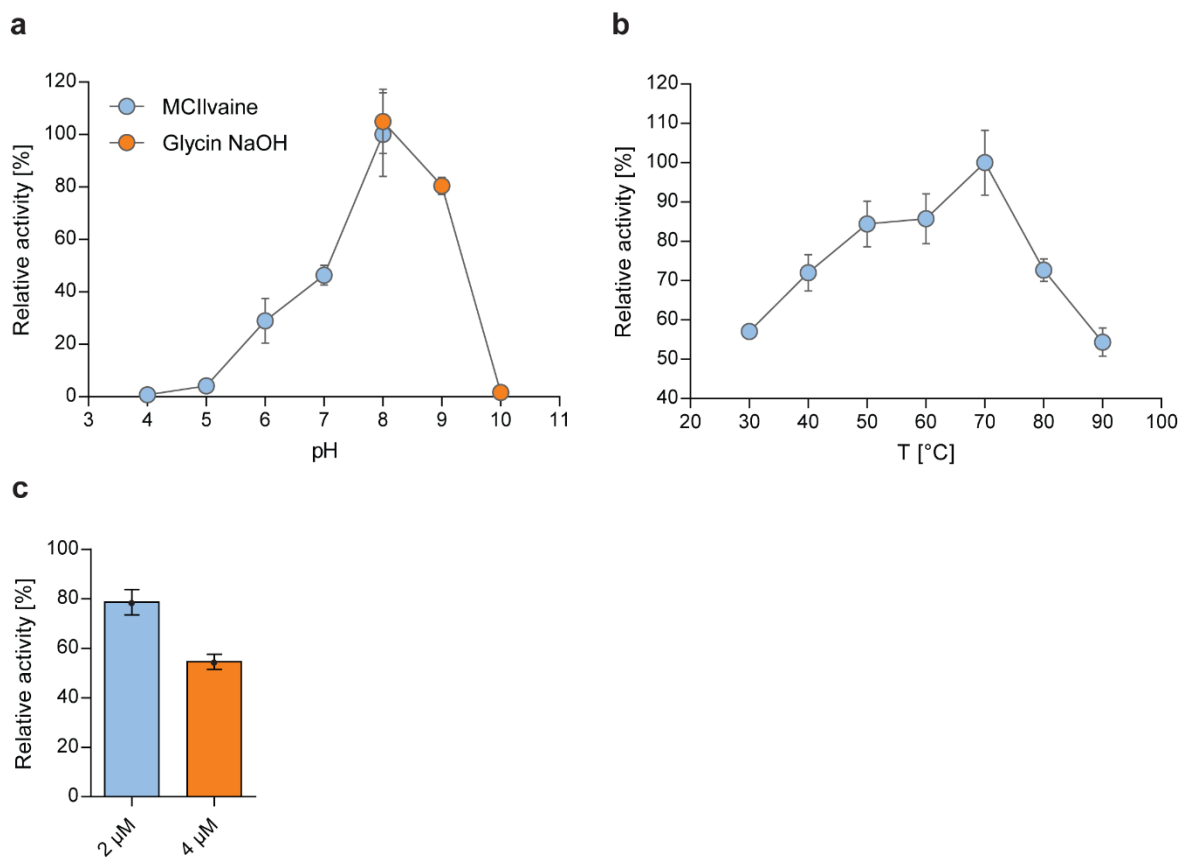


Figure S2. Biochemical characterization of the heterologously expressed esterase.

Effect of pH (A) and temperature (B) on the relative activity of the esterase with pNP-butyrates. C) Activity inhibition of the esterase upon preincubation with 2 μ M or 4 μ M FP-alkyne at pH 7.5 and 70 $^{\circ}$ C for 10 min. The remaining activity was measured using pNP-acetate as the substrate. Error bars represent SD of three technical replicates.

Supplementary Tables

Table S1. Relative abundance of microorganisms in KAM3811 based on the coverage of scaffolds containing the respective rpS3 gene.

Nearest predicted relative based on BLASTp	Phylum	Domain	Abundance (rpS3)
Sulfurihydrogenibium sp.	Aquificota	Bacteria	2329.1
Pyrobaculum ferrireducens	Thermoproteota	Archaea	411.3
Caldisphaera sp.	Thermoproteota	Archaea	389.6
Hydrogenobaculum sp.	Aquificota	Bacteria	281.7
Nitrososphaera sp.	Thermoproteota	Archaea	160.8
Sulfolobales (SCGC AB-777 J03)	Thermoproteota	Archaea	158.5
Aciduliprofundum sp.	Thermoplasmatota	Archaea	139.0
Caldisericum exile	Caldisericota	Bacteria	78.2
Fervidicoccus sp.	Thermoproteota	Archaea	59.7
Caldimicrobium thiodismutans	Desulfobacterota	Bacteria	58.9
Candidatus Nanoarchaeota stetteri	Nanoarchaeota	Archaea	38.5
Thermoproteus sp. (JCHS 4)	Thermoproteota	Archaea	26.9
Aciduliprofundum sp.	Thermoplasmatota	Archaea	19.1
Caldivirga maquilingensis (IC-167)	Thermoproteota	Archaea	17.1
Mesoaciditoga sp.	Thermotogota	Bacteria	13.9
Caldivirga maquilingensis (IC-167)	Thermoproteota	Archaea	11.5
Thiomonas sp. (CB3)	Proteobacteria	Bacteria	11.1
Thiomonas sp. (20-64-5)	Proteobacteria	Bacteria	10.9
Acidilobus saccharovorans (345-15)	Thermoproteota	Archaea	7.7

Table S2. Relative abundance of microorganisms in KAM3808 based on the coverage of scaffolds containing the respective rpS3 gene.

Nearest predicted relative based on BLASTp	Phylum	Domain	Abundance (rpS3)
Aciduliprofundum sp.	Thermoplasmatota	Archaea	488.1
Caldisericum exile	Caldisericota	Bacteria	408.2
Caldimicrobium thiodismutans	Desulfobacterota	Bacteria	139.9
Archaeon, unclassified	not predicted	Archaea	124.6
Sulfurihydrogenibium sp.	Aquificota	Bacteria	119.9
Nitrososphaera sp.	Thermoproteota	Archaea	93.7
Thermus arciformis	Deinococcota	Bacteria	81.6
Candidate division WOR-3 JGI Cruoil (03_51_56)	WOR-3	Bacteria	64.5
Dictyoglomus turgidum (DSM 6724)	Dictyoglomota	Bacteria	54.9
Candidatus Micrarchaeota (CG1_02_51_15)	Micrarchaeota	Archaea	52.9
Thermoflexia	Chloroflexota	Bacteria	46.3
Chlorobi (MS-B_bin-24)	Bacteroidota	Bacteria	37.1
Parcubacteria (ADurb.Bin305)	Patescibacteria	Bacteria	31.5
Candidatus Bathyarchaeota (ex4484_205)	Thermoproteota	Archaea	25.4
Aciduliprofundum sp.	Thermoplasmatota	Archaea	20.1
Thermosulfidibacter takaii (ABI70S6)	Thermosulfidibacterota	Bacteria	18.7
Fervidicoccus sp.	Thermoproteota	Archaea	18.1
Thermodesulfovibrio aggregans	Nitrospirota	Bacteria	12.8
Bacterium, unclassified	not predicted	Bacteria	11.7
Candidatus Cryosericum odellii	Caldisericota	Bacteria	11.7
Archaeon, unclassified	not predicted	Archaea	11.6
Thermogutta terrifontis	Planctomycetota	Bacteria	10.7
Carboxydocella sp.	Firmicutes	Bacteria	10.5
Candidatus Woesearchaeota (ex4484_78)	Nanoarchaeota	Archaea	9.7
Bacterium, unclassified	not predicted	Bacteria	9.5
Candidatus Roizmanbacteria	Patescibacteria	Bacteria	9.1
Elusimicrobia (CG08_land_8_20_14_0_20_51_18)	Elusimicrobiota	Bacteria	8.1
Archaeon, unclassified	not predicted	Archaea	7.8
Verrucomicrobia	Verrucomicrobiota	Bacteria	7.8
Candidate division Zixibacteria (RBG_16_43_9)	Zixibacteria	Bacteria	7.6
Candidatus Omnitrophica	Omnitrophota	Bacteria	7.4
Candidatus Aenigmarchaeota	Aenigmataarchaeota	Archaea	7.2
Geobacter sp.	Desulfobacterota	Bacteria	6.8
Bacterium HR16	Armatimonadota	Bacteria	6.7
Heliorestis acidaminivorans	Firmicutes	Bacteria	6.5
Archaeon, unclassified	not predicted	Archaea	6
Thermofilum uzonense	Thermoproteota	Archaea	5.8
Fervidobacterium nodosum (Rt17-B1)	Thermotogota	Bacteria	5.4
Archaeon, unclassified	not predicted	Archaea	5.3
Thermodesulforhabdus norvegica	Desulfobacterota	Bacteria	5.2
Archaeon, unclassified	not predicted	Archaea	5.1
Candidatus Atribacteria (ADurb.Bin276)	Atribacterota	Bacteria	5.1
Desulfurella sp.	Campylobacterota	Bacteria	5.1
Ignavibacteriales (CG07_land_8_20_14_0_80_59_12)	Bacteroidota	Bacteria	4.3
Parcubacteria (ADurb.Bin305)	Patescibacteria	Bacteria	4
Candidatus Aminicenantes	Acidobacteriota	Bacteria	3.4

Candidatus Aminicenantes	Acidobacteriota	Bacteria	3.4
Lentisphaeria	Verrucomicrobiota	Bacteria	3.4
Ignavibacteria (RBG_13_36_8)	Bacteroidota	Bacteria	2.7

Table S3. Selected structural homologs of the putative esterase determined with HHpred.

Hit	Function	Origin	Probability [%]	E-value	ScoreOrigin
4FLE_A	Ancestral haloalkane dehalogenase AnchLD3;	Synthetic construct	99.96	7.8e-26	140.97
413F_A	serine hydrolase CCSP0084; MCP cleaving	<i>Cycloplasticus</i> sp. P1	99.93	4.3e-24	141.6
3N98_A	Chloroperoxidase F; Haloperoxidase, Oxidoreductase	<i>Pseudomonas fluorescens</i>	99.93	3.5e-24	138.8
3G9X_A	Haloalkane dehalogenase; alpha/beta hydrolase, helical cap domain	<i>Rhodococcus rhodochrous</i>	99.93	1.6e-23	137.84
4UHC_A	Esterase; alpha/beta hydrolase, pnp-ester cleaving	<i>Thermogutta terrifontis</i>	99.93	5.4e-24	138.98
3RM3_A	Thermostable monoacylglycerol lipase; alpha/beta hydrolase fold	<i>Bacillus</i> sp. H257	99.92	1.5e-23	136.84
3PFB_B	Cinnamoyl esterase; alpha/beta hydrolase fold, esterase, hydrolase, cinnamoyl/feruloyl esterase	<i>Lactobacillus johnsonii</i>	99.92	7.7e-23	133.82
1C4X_A	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (BPHD); PCB degradation	<i>Rhodococcus</i> sp. Strain Rha1	99.92	7.9e-24	138.27
5OLU_A	carboxyl esterase, 1,2-O-isopropylidene-glycerol hydrolyzing, lipase, alpha/beta hydrolase	<i>Bacillus coagulans</i>	99.92	1.3e-22	136.85
4LXH_A	MCP Hydrolase; carbon-carbon bond hydrolase, Rossmann Fold, alpha/beta hydrolase fold	<i>Sphingomonas wittichii</i> RW1	99.92	2.2e-23	135.76
5Y6Y_B	Epoxide hydrolase	<i>Vigna radiata</i>	99.92	2.1e-23	138.9
2WTM_C	Promiscuous Feruloyl Esterase (Est1E)	<i>Butyrivibrio proteoclasticus</i>	99.92	1.2e-22	130.56
4C6H_A	Haloalkane dehalogenase	Rhodobacteraceae	99.92	5.1e-23	135.33
2XTO_A	Haloalkane Dehalogenase	<i>Plesiocystis pacifica</i> SIR-I	99.92	3.5e-23	136.21
1Q0R_A	aclacinomycin methylesterase; Anthracycline, methylesterase, polyketide hydrolase	<i>Serratia marcescens</i>	99.92	9.6e-24	139.12
6F9O_A	haloalkane dehalogenase DpcA	<i>Psychrobacter cryohalolentis</i> K5	99.92	5.3e-23	136.58
6Y9G_B	Ancestral haloalkane dehalogenase AnchLD5	Synthetic construct	99.91	1.5e-22	133.85
2E3J_A	epoxide hydrolase B (Rv1938)	<i>Mycobacterium tuberculosis</i>	99.91	1.3e-22	137.43

5XKS_F	Thermostable monoacylglycerol lipase	<i>Geobacillus</i> sp. 12AMOR	99.91	4.6e-22	128.73
2WFL_A	polyneuridine aldehyde esterase (PNAE)	<i>Rauvolfia serpentina</i>	99.91	3.3e-22	128.71
3BF7_B	Esterase YbfF; esterase, thioesterase	<i>Escherichia coli</i>	99.91	2.4e-22	128.91
6THS_A	Serine esterase, cutinase S165A	Uncultured bacterium	99.9	6.5e-22	129
5XWZ_B	Alpha/beta-hydrolase, lactonase, zearalenone hydrolase	<i>Cladophialophora bantiana</i>	99.9	3.8e-22	130.52
1IUP_A	meta-Cleavage product hydrolase; aromatic compounds, cumene, isopropylbenzene, meta-cleavage compound hydrolase	<i>Pseudomonas fluorescens</i> IP01 (CumD)	99.9	1.6e-21	127.29
6BA9_A	yersiniabactin synthesis enzyme, YbtT; Thioesterase, non-ribosomal peptide synthesis	<i>Escherichia coli</i>	99.9	3e-22	129.51
4CCY_A	Carboxylesterase YBFK, naproxenmethylester hydrolase	<i>Bacillus subtilis</i>	99.9	6.4e-23	135.19
2WUE_B	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (BPHD)	<i>Mycobacterium tuberculosis</i>	99.89	2e-21	127.37
2RHW_A	BphD, C-C Bond Hydrolase Involved in Polychlorinated Biphenyls Degradation	<i>Burkholderia xenovorans</i> LB400	99.89	1.1e-21	127.92
3FCY_A	Xylan esterase 1; alpha/beta hydrolase, carbohydrate esterase, CE7	<i>Thermoanaerobacterium</i> sp. JW/SL YS485	99.89	8.9e-22	133.37
6AGQ_A	acetyl xylan esterase	<i>Paenibacillus</i> sp. R4	99.89	1.8e-21	130.01
1MJ5_A	1,3,4,6-tetrachloro-1,4-cyclohexadiene hydrolase; LINB, hydrolase, Haloalkane dehalogenase	<i>Sphingomonas paucimobilis</i> UT26	99.88	1e-20	125.53
5XH2_A	Poly(ethylene terephthalate) hydrolase	<i>Ideonella sakaiensis</i> 201-F6	99.88	5.6e-21	124.45
4CG1_A	Cutinase, PET degrading hydrolase	<i>Thermobifida fusca</i>	99.86	1.4e-19	119.52
7NEI_B	Polyester Hydrolase Leipzig 7 (PHL-7); PETase, Cutinase	unidentified	99.86	7.4e-20	119.78

Experimental section

Sample Collection and Chemical Labeling

Sediments for chemical labeling experiments have been sampled from the two hot springs ‘Arkashin shurf’ (54°30.0016N 160°00.2021E, 65.5-72 °C, pH = 5.01, internal number #3811) and ‘Helicopter spring’ (54°30.003N 160°00.4375E, 58.1 °C, pH = 5.62, internal number #3808), both located at the Uzon volcanic caldera (Kamchatka Peninsula, Russia). FP-alkyne was dissolved in DMSO. A slurry of sediments in spring water was collected and after gentle mixing, 10 ml of the slurry was dispensed in a reaction tube and incubated with 4 μM FP-alkyne for 2 h with occasional shaking while being placed back in the spring. An equal volume of DMSO was added to the negative controls. All samples were prepared in triplicates. For metagenomic sequencing, an unlabeled aliquot of the slurry was prepared. For downstream processing, the samples were transported to the laboratory on dry ice, the slurry was centrifuged to remove the spring water (12,000 × g, room temperature, 10 min) and the sediments were stored at -20° C until further processing.

DNA Extraction and Metagenomic Sequencing

Total DNA was isolated from the sediments using phenol-chloroform extraction as described in Gavrilov *et al* [1]. Prior to isolation, the cells were disrupted using a series of freezing-thawing cycles. Concentration of DNA was measured on a Qubit 2.0 fluorometer (Invitrogen, Waltham, MA, USA). Shotgun metagenome library preparation and sequencing were done at BioSpark Ltd., Moscow, Russia. The KAPA HyperPlus Library Preparation Kit (KAPA Biosystems, Wilmington, MA, USA) was used for library construction according to the manufacturer’s protocol and sequencing was performed on a NovaSeq 6000 platform (Illumina, San Diego, CA, USA) with the NovaSeq 6000 S2 Reagent Kit, which can read 100 nucleotides from each end (200 cycles).

Genome-Resolved Metagenomics

Raw reads were quality filtered and cleaned with BBDuk (<https://jgi.doe.gov/data-and-tools/bbtools>) and Sickle [2], followed by assembly with metaSPAdes (version 3.14) [3]. Assembled sequences below 1 kbp in length were discarded and gene prediction was carried out using Prodigal [4] in meta mode followed by annotation against the FunTaxDB [5], which is based on the UniRef100 database [6]. Community composition of samples was determined based on ribosomal protein S3 (rpS3) and its respective coverage on scaffolds in the individual metagenomes. The coverage was determined via mapping of metagenomic reads and taken as

the relative abundance of rpS3 genes and their respective microbes in the community [7]. Binning of genomes was performed with MaxBin2 [8], ABAWACA [9] and emergent self-organizing maps [10]. High quality genomes were identified using DAS Tool [11] and further curated with uBin [5]. A phylogenetic tree was calculated with GTDB-Tk based on the dereplicated genomes [12]. Genome completeness, contamination, GC content and genome length was calculated via CheckM [13]. The target genome that contained the serine hydrolase chosen for biochemical characterization (see Heterologous protein expression and purification) was analyzed and re-annotated with NCBI's PGAP (Prokaryotic Genomes Annotation Pipeline) to improve start and end of gene prediction [14]. The metagenome redundancy was calculated with Nonpareil 3 (version 3.304) [15].

Protein Extraction and Clean-Up

For protein extraction from labeled sediments, the collected organic matter was thawed on ice and taken up in 5 mL of extraction buffer (100 mM Tris-HCl pH 8.8, 0.1 M DTT, 50 mM EDTA, 1.5% SDS, 30% sucrose). The cells were lysed in a three-step sonication procedure: seven iterations of sonication in an ultrasonic bath (BANDELIN electronic, Berlin, Germany) for 1 min followed by vigorous mixing, 10 cycles with high power in a Bioruptor UCD-200 (Diagenode, Seraing, Belgium) device with the following conditions: 1 min pulse and 30 sec pause and another ten iterations of sonication in an ultrasonic bath as described before. The extracts were then cleared by centrifugation ($100 \times g$, room temperature, 5 min) and the supernatant was collected in a fresh tube. The debris were centrifuged again ($15,000 \times g$, room temperature, 20 min) and the supernatant was combined with the supernatant in the fresh tube. The combined supernatant was subjected to another centrifugation step ($15,000 \times g$, room temperature, 15 min) to separate the soluble protein containing fraction from any remaining debris. Protein clean-up was done by performing a phenol extraction [16] with downstream ammonium acetate precipitation [17] according to the literature with few modifications. In brief, the protein containing fractions were mixed and incubated (15 min, room temperature, shaking) with an equal volume of TE-buffered liquid Phenol (Carl Roth, Karlsruhe, Germany; #0038). To achieve phase separation, the samples were centrifuged ($12,000 \times g$, room temperature, 10 min). The upper phenol phase was collected and re-extracted with extraction buffer thrice. Thereto, the phenol phase was mixed with an equal volume of extraction buffer and the phases were separated by centrifugation ($12,000 \times g$, room temperature, 10 min). The proteins in the phenol phase were precipitated with a five-fold volume of 0.1 M ammonium acetate in methanol ($-20 \text{ }^\circ\text{C}$, overnight). The proteins were collected by centrifugation

(12,000 × g, 4 °C, 30 min) and the pellet was successively washed twice with 0.1 M ammonium acetate in methanol, twice with 80% (v/v) acetone, and once with 70% (v/v) ethanol, respectively. Each washing step included an incubation step at -20 °C for 20-30 min prior to centrifugation (12,000 × g, 4 °C, 30 min). The precipitated proteins were dissolved in 100 μL 8 M urea in 50 mM HNa₂PO₄, pH 8.0 and further diluted with 50 mM HNa₂PO₄ pH 8.0 to a final concentration of 2 M urea. The protein concentration of the resulting protein solutions was determined by a modified Bradford assay with Roti-Nanoquant (Carl Roth, Karlsruhe, Germany).

Click Reaction and Affinity Purification

400-650 μg of total protein were subjected to a click reaction with 10 μM 5/6-TAMRA-biotin-N₃ (Jena Bioscience, Jena, Germany; #CLK-1048), 100 μM TBTA, 2 mM TCEP and 2 mM CuSO₄ (all purchased from Sigma-Aldrich, St. Louis, MO, USA) in a total reaction volume of 500 μL (1 h, room temperature, in the dark). Prior to affinity purification, unbound reporter and salts from the click reaction were removed by methanol-chloroform precipitation [18]. The resulting protein pellet was air-dried and subsequently dissolved (37 °C, ~1 h) in 850 μL 2% (w/v) SDS in 1× PBS (155 mM NaCl, 3 mM Na₂HPO₄, 1.06 mM KH₂PO₄, pH 7.4). Insoluble particles were removed by centrifugation (21,000 × g, 37 °C, 5 min) and the cleared protein solution was diluted with 1× PBS to a final concentration of 0.2% (w/v) SDS. The obtained protein mixture was incubated with 100 μl of pre-equilibrated avidin beads slurry (Thermo Scientific, Waltham, MA, USA; #20219) while gently tumbling (~1h, room temperature, in the dark). Subsequently, the beads were washed five times with 10 mL 1% (w/v) SDS (10 min, room temperature, gently rotating) and collected by centrifugation (400 × g, 5 min). To remove SDS from the samples, the beads were then washed four times with 1 mL of ultrapure water (VWR Chemicals, Radnor, PA, USA; 5 min, room temperature, vigorously shaking) and collected by centrifugation (3,000 × g, 1 min).

On-Bead Digestion of Captured Proteins

After affinity enrichment, the beads were taken up in 100 μL 0.8 M urea in 50 mM ammonium bicarbonate (ABC; Sigma-Aldrich, St. Louis, MO, USA). Disulfide bonds were reduced by adding 10 mM DTT (Sigma-Aldrich, St. Louis, MO, USA) in 50 mM ABC (1 h, room temperature, vigorous shaking) and the generated cysteine mercapto groups were masked by alkylation with 25 mM iodoacetamide (IAM; Sigma-Aldrich, St. Louis, MO, USA) in 50 mM ABC (1 h, room temperature, in the dark, vigorous shaking). Excess IAM was then quenched

by adding DTT (final concentration of 35 mM, 10 min, room temperature, vigorous shaking). Protein on-bead digestion was started by adding 1 μ g Trypsin (Thermo Scientific, Waltham, MA, USA; #90057) dissolved in 50 mM acetic acid (37 °C, ~16 h, vigorous shaking). After digestion, the reaction solution was cleared by centrifugation (3,000 \times g, room temperature, 5 min) and the supernatant (contains the digestion products (peptides)) was transferred to a fresh reaction vessel and the digestion reaction stopped by adding formic acid (FA) to a final concentration of 5% (v/v). Next, the beads were washed with 50 μ L 1% (v/v) FA and the supernatant was combined with the recovered digestion mix. To remove residual beads from the peptide solution, the mix was passed over a home-made two-disc glass microfiber membrane (GE Healthcare, Chicago, IL, USA; pore size 1.2 μ m, thickness 0.26 mm) tip.

Sample Clean-Up for LC–MS

Peptides were desalted on home-made C₁₈ StageTips [19] containing two layers of an octadecyl silica membrane (3M, Saint Paul, MN, USA). All centrifugation steps were carried out at room temperature. The StageTips were first activated and equilibrated by passing 50 μ L of methanol (600 \times g, 2 min), 80% (v/v) acetonitrile (ACN) with 0.5% (v/v) FA (600 \times g, 2 min) and 0.5% (v/v) FA (800 \times g, 3 min) over the tips. Next, the tryptic digests were passed over the tips (800 \times g, 3-4 min). The flow-through was collected and applied a second time (same settings). The immobilized peptides were then washed with 50 μ L and 25 μ L 0.5% (v/v) FA (800 \times g, 3 min). Bound peptides were eluted from the StageTips by application of two rounds of 25 μ L 80% (v/v) ACN with 0.5% (v/v) FA (600 \times g, 2 min). After elution from the StageTips, the peptide samples were dried using a vacuum concentrator (Eppendorf, Hamburg, Germany) and the peptides were dissolved in 15 μ l 0.1% (v/v) FA prior to analysis by MS.

LC-MS/MS Analysis

LC–MS/MS experiments were performed on an Orbitrap Fusion Lumos Tribrid instrument (Thermo Scientific, Waltham, MA, USA) that was coupled to an EASY-nLC 1200 liquid chromatography (LC) system (Thermo Scientific, Waltham, MA, USA). The LC was operated in the one-column mode. The analytical column was a fused silica capillary (75 μ m \times 46 cm) with an integrated PicoFrit emitter (New Objective, Littleton, MA, USA) packed in-house with Reprisil-Pur 120 C18-AQ 1.9 μ m resin (Dr. Maisch, Ammerbuch, Germany). The analytical column was encased by a PRSO-V2 column oven (Sonation, Biberach, Germany) and attached to a nanospray flex ion source (Thermo Scientific, Waltham, MA, USA). The column oven temperature was adjusted to 50 °C during data acquisition. The LC was equipped with two

mobile phases: solvent A (0.1% (v/v) FA in water) and solvent B (0.1% (v/v) FA in 80% (v/v) ACN). All solvents were of UPLC grade (Honeywell, Charlotte, NC, USA). Peptides were directly loaded onto the analytical column with a maximum flow rate that would not exceed the set pressure limit of 980 bar (usually around 0.5–0.8 $\mu\text{l min}^{-1}$). Peptides were subsequently separated on the analytical column by running a 200 min gradient of solvent A and solvent B at a flow rate of 300 nL min^{-1} (gradient: start with 9% solvent B; gradient 9–40% solvent B for 180 min; gradient 40–100% solvent B for 15 min and 100% solvent B for 5 min). The mass spectrometer was operated using Xcalibur software (version 4.3.7.3.11; Thermo Fischer Scientific) and was set in the positive ion mode. The ionization potential (spray voltage) was set to 2.3 kV. A top-speed data-dependent method with a cycle time of 3 seconds was selected for data acquisition. Precursor ion scanning (MS^1) was performed in the Orbitrap analyzer (FTMS; Fourier Transform Mass Spectrometry) at a resolution of 120,000 FWHM (full width at half maximum @ 200 m/z) in the scan range of m/z 375–1,500 with the internal lock mass option turned on (lock mass was m/z 445.12002, polysiloxane) [20]. The automatic gain control (AGC) was set to “standard” and the maximum injection time was machine determined (“auto”). Product ion spectra were recorded in the Orbitrap at a resolution of 15,000 FWHM. The scan range for MS^2 was set to “auto”. Ions for fragmentation were selected in the quadrupole (isolation window of m/z 1.6) based on their intensity (threshold 5×10^4 ions) and charge state (only charge state of 2–7) in the full survey scan. The AGC target was set to “standard” and the maximum injection time to “auto”. Selected precursor ions were fragmented by Higher-energy C-trap dissociation (HCD) with normalized collision energy (NCE) set to 30%. Monoisotopic precursor selection was enabled. During MS^2 data acquisition, dynamic ion exclusion was set to 60 s with a repeat count of 1 and a mass tolerance of ± 10 ppm.

Peptide and Protein Identification Using MaxQuant and Perseus

RAW spectra were submitted to an Andromeda [21] search in MaxQuant (version 1.6.17.0) using the default settings [22]. Label-free quantification was activated [23]. MS/MS spectra data were searched against the self-assembled metaproteome databases of the ‘Arkashin shurf’ (45649 entries) or the ‘Helicopter spring’ (99930 entries), accordingly. All searches included a contaminants database (as implemented in MaxQuant, 246 sequences). The contaminants database contains known MS contaminants and was included to estimate the level of contamination. Andromeda searches allowed for oxidation of methionine residues (16 Da) and acetylation of the protein N-terminus (42 Da) as dynamic modifications while carbamidomethylation of cysteine residues (57 Da, alkylation with IAM) was selected as static

modification. Enzyme specificity was set to “Trypsin/P”. The instrument type in Andromeda searches was set to Orbitrap and the precursor mass tolerance was set to ± 20 ppm (first search) and ± 4.5 ppm (main search). The MS/MS match tolerance was set to ± 20 ppm. The peptide spectrum match FDR and the protein FDR were set to 0.01 (based on target-decoy approach). Minimum peptide length was 7 amino acids. For protein quantification, unique and razor peptides were allowed. In addition to unmodified peptides, modified peptides with dynamic modifications were allowed for quantification. The minimum score for modified peptides was set to 40.

Further data analysis and filtering of the MaxQuant output was done in Perseus [24] (version 1.6.14.0). Label-free quantification (LFQ) intensities were loaded into the matrix from the proteinGroups.txt file and potential contaminants as well as reverse hits from the reverse database and hits only identified based on peptides with a modification site were removed.

Biological replicates of the unlabeled controls and the FP-alkyne labeled samples were combined into separate categorical groups to allow comparison of the different treatments. The data were transformed to the log₂-scale and only protein groups with a minimum of 2 identified unique peptides were kept in the matrix. Furthermore, only hits with a valid LFQ intensity for at least one of the probe-labeled sample replicates were selected for further analysis to exclude hits that were solely identified in the controls. To allow protein quantification, missing values were imputed from a normal distribution (width 0.3, down shift 1.8). Comparison of normalized protein group quantities (relative quantification) between different MS runs was solely based on the LFQ intensities as calculated by MaxQuant (MaxLFQ algorithm).^[22] Briefly, label-free protein quantification was switched on and unique and razor peptides were considered for quantification with a minimum ratio count of 2. Retention times were recalibrated based on the built-in nonlinear time-rescaling algorithm. MS/MS identifications were transferred between LC-MS/MS runs with the “Match between runs” option in which the match time window was set to 0.7 min and the alignment time window to 20 min. The quantification was based on the “value at maximum” of the extracted ion current. At least two quantitation events were required for a quantifiable protein. The log₂-fold enrichment of protein groups with FP-alkyne was calculated based on the mean LFQ intensity compared to the DMSO control. Protein groups with a negative fold enrichment were excluded from further analysis. The remaining protein groups were reported in the respective figure (Fig. 4).

Bioinformatic Analyses of Enriched Proteins

In order to confidently predict potential serine hydrolases among the group of proteins that was enriched with FP-alkyne, the protein sequences of the respective proteins were analyzed using various tools and databases. Sequence similarity to deposited serine hydrolases including the presence of characteristic serine hydrolase domains was analyzed using UniRef100 [6], PFAM [25], NCBI CDD [26] and InterProScan [27]. Structural homology to known serine hydrolases was assessed using the SWISS-MODEL template library [28] and HHpred [29].

Heterologous Protein Expression and Purification

The putative esterase selected for biochemical characterization (identifier: ExploCarb_3811S_S4_483_length_13114_cov_941_5) was heterologously expressed in *E. coli*. There to, *E. coli* Rosetta cells (Novagen, Madison, WI, USA) were transformed with the commercially obtained construct (BioCat, Heidelberg, Germany) of the codon-optimized gene cloned into a pET-28b(+) vector encoding a C-terminal 6× His-tag. For recombinant expression of the UPF0227 gene, a freshly inoculated 1 L culture in LB medium supplemented with 50 µg mL⁻¹ kanamycin and 50 µg mL⁻¹ chloramphenicol was grown to an OD₆₀₀ of 0.4 at 37 °C with constant shaking (180 rpm) until subsequent induction of the protein expression with 500 µM isopropyl-β-D-thiogalactopyranoside (IPTG). Upon further incubation at 18 °C for 16 h, the cells were harvested by centrifugation (8,000 × g, 4 °C, 20 min) and the resuspended in 5 mL 50 mM Tris-HCl pH 7.0 per gram wet weight of the pellet. Cell lysis was performed by sonication in three cycles for 5 min (cycle 0.5, amplitude 50) with a UP 200S sonicator (Hielscher Ultrasonics, Teltow, Germany). The crude extract was cleared by centrifugation (12,000 × g, 45 min, 4 °C) and the lysate was passed through a 0.45 µm filter. Protein affinity purification was done using a Protino™ Ni-TED 1000-packed column (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. Prior to further clean-up of the recombinant protein by size exclusion chromatography (SEC), the elution buffer was exchanged with size exclusion buffer (50 mM Tris-HCl pH 7.5, 20 mM NaCl) by centrifugation (6,000 × g, room temperature, 40 min) using Amicon® centrifugal filter devices (10 kDa cutoff, Merck, Darmstadt, Germany) and the solution was concentrated to 1 mL. SEC was performed on a HiLoad® 16/600 Superdex® 200 pg column (GE Healthcare, Chicago, IL, USA) connected to an ÄKTA™ FPLC system (GE Healthcare, Chicago, IL, USA) at a flow rate of 1 mL min⁻¹. Fractions containing the recombinant protein (detection at 280 nm) were pooled and concentrated as described above. For long-time storage at -80 °C, 50% (v/v) glycerol was added to the protein solution that was flash-frozen in liquid nitrogen.

Structural and Bioinformatics Analysis of the Putative Esterase

Retrieval of homologous sequences, structures and domains was conducted with BLAST (blastp tool) [30], HHpred [29] and HMMER (phmmer tool) [31], respectively. For further structural analysis and structure comparison, a model of the UPF0227 protein was constructed with AlphaFold (version 2.0) using default settings [32]. The resulting PDB file was used for visualization and processing of the protein structure with UCSF Chimera (version 1.14) [33].

Biochemical characterization of the heterologous protein

The activity of the heterologously expressed UPF0227 protein was determined using a continuous assay with the chromogenic *para*-nitrophenyl (*p*NP) substrates *p*NP-acetate, *p*NP-butyrate, *p*NP-octanoate, *p*NP-decanoate and *p*NP-dodecanoate (all obtained from Megazyme, Wicklow, Ireland). The increase in absorbance at 384 nm was measured using a Specord 210[®] photometer (Analytik Jena, Jena, Germany) and the enzymatic activity was calculated from a calibration curve with *p*NP. The pH and temperature optimum of the enzyme was determined with *p*NP-butyrate prior to assessing the enzyme kinetics with the different *p*NP esters at the respective pH and temperature using substrate concentrations up to 0.7 mM. To test the inhibition effect of the FP-alkyne probe used for ABPP, 6 µg/ml protein in 50 mM Tris-HCl pH 7.5, 20 mM NaCl were incubated with 2 or 4 µM of the probe for 10 min at 70 °C. The protein solutions were then submitted to activity measurement with *p*NP-acetate to determine the residual esterase activity compared to a DMSO-treated control.

***In vitro* labeling**

In vitro ABPP with FP-alkyne was performed by incubation of indicated amounts of the enzyme with 4 µM of the probe (1 h, 70 °C) in a final reaction volume of 50 µL. An equal volume of DMSO was added to the negative controls. Preincubation with paraoxon-ethyl (Sigma-Aldrich, St. Louis, MO, USA), if applicable, was done at a final concentration of 100 µM (15 min, 70 °C). The subsequent click reaction was performed as described above (Click reaction and affinity purification) using Cy3-N₃ (synthesized in house) as click tag. For gel-based analysis, the samples were mixed with 1 equivalent 4× LDS gel loading dye (423 mM Tris HCl, 563 mM Tris base, 8% (w/v) lithium dodecyl sulfate (LDS), 40% (w/v) glycerol, 2 mM EDTA, 0.075% (w/v) SERVA Blue G250; supplemented with 100 mM DTT) and incubated at 70 °C for 15 min. Separation of proteins (1/5 of the initial amount) by gel electrophoresis was done on a 11% Bis-Tris resolving gel, followed by visualization of labeled proteins using a Typhoon FLA 9000 laser scanner (GE Healthcare, Chicago, IL, USA).

Supplementary References

1. Gavrilov SN, Stracke C, Jensen K, Menzel P, Kallnik V, Slesarev A, et al. Isolation and Characterization of the First Xylanolytic Hyperthermophilic Euryarchaeon *Thermococcus* sp. Strain 2319x1 and Its Unusual Multidomain Glycosidase. *Front Microbiol.* 2016;7:552.
2. Joshi N, Fass J. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). [Software]. 2011.
3. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 2017;27:824-34.
4. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinformatics.* 2010;11:119.
5. Bornemann TLV, Esser SP, Stach TL, Burg T, Probst AJ. uBin – a manual refining tool for metagenomic bins designed for educational purposes. *bioRxiv.* 2020;2020:07.
6. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt C. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics.* 2015;31:926-32.
7. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357-9.
8. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics.* 2016;32:605-7.
9. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature.* 2015;523:208-U173.
10. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Yelton AP, Banfield JF. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* 2009;10:1-16.
11. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol.* 2018;3:836-43.
12. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2020;36:1925-7.
13. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043-55.
14. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016;44:6614-24.
15. Rodriguez RL, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. *mSystems.* 2018;3:e00039-18.
16. Wang W, Scali M, Vignani R, Spadafora A, Sensi E, Mazzuca S, et al. Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis.* 2003;24:2369-75.
17. Benndorf D, Vogt C, Jehmlich N, Schmidt Y, Thomas H, Woffendin G, et al. Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. *Biodegradation.* 2009;20:737-50.
18. Wessel D, Flügge UI. A Method for the Quantitative Recovery of Protein in Dilute-Solution in the Presence of Detergents and Lipids. *Anal Biochem.* 1984;138:141-3.

19. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc.* 2007;2:1896-906.
20. Olsen JV, de Godoy LMF, Li GQ, Macek B, Mortensen P, Pesch R, et al. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics.* 2005;4:2010-21.
21. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J Proteome Res.* 2011;10:1794-805.
22. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 2008;26:1367-72.
23. Cox J, Hein MY, Lubner CA, Paron I, Nagaraj N, Mann M. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol Cell Proteomics.* 2014;13:2513-26.
24. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods.* 2016;13:731-40.
25. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49:D412-9.
26. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 2020;48:D265-8.
27. Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30:1236-40.
28. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014;42:W252-8.
29. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol.* 2018;430:2237-43.
30. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 2004;32:W20-5.
31. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46:W200-4.
32. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583-9.
33. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25:1605-12.