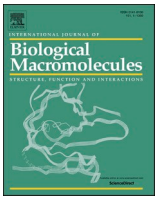
International Journal of Biological Macromolecules 275 (2024) 133449

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac



Evolutionary history and activity towards oligosaccharides and polysaccharides of GH3 glycosidases from an Antarctic marine bacterium

Alessandro Marchetti [a], Marco Orlando [a], Luca Bombardi [b], Salvatore Fusco [b], Marco Mangiagalli [a,*], Marina Lotti [a]

[a] Department of Biotechnology and Biosciences, University of Milano Bicocca, Piazza della Scienza 2, Milano 20126, Italy
[b] Biochemistry and Industrial Biotechnology (BIB) Laboratory, Department of Biotechnology, University of Verona, Verona, Italy

ARTICLE INFO

Keywords:
Glycoside hydrolases (GH)
Marinomonas sp. ef1
Cold-active enzymes
Marine bacteria
Oligosaccharides and polysaccharides degradation

ABSTRACT

Glycoside hydrolases (GHs) are pivotal in the hydrolysis of the glycosidic bonds of sugars, which are the main carbon and energy sources. The genome of *Marinomonas* sp. ef1, an Antarctic bacterium, contains three GHs belonging to family 3. These enzymes have distinct architectures and low sequence identity, suggesting that they originated from separate horizontal gene transfer events.

M-GH3_A and M-GH3_B, were found to differ in cold adaptation and substrate specificity. M-GH3_A is a *bona fide* cold-active enzyme since it retains 20 % activity at 10 °C and exhibits poor long-term thermal stability. On the other hand, M-GH3_B shows mesophilic traits with very low activity at 10 °C (< 5 %) and higher long-term thermal stability. Substrate specificity assays highlight that M-GH3_A is a promiscuous β-glucosidase mainly active on cellobiose and cellotetraose, whereas M-GH3_B is a β-xylosidase active on xylan and arabinoxylan. Structural analysis suggests that such functional differences are due to their differently shaped active sites. The active site of M-GH3_A is wider but has a narrower entrance compared to that of M-GH3_B.

Genome-based prediction of metabolic pathways suggests that *Marinomonas* sp. ef1 can use monosaccharides derived from the GH3-catalyzed hydrolysis of oligosaccharides either as a carbon source or for producing osmolytes.

1. Introduction

Polar marine bacteria must face various stressful conditions [1]. Indeed, they are constantly exposed to low temperatures, which reduce membrane fluidity and hinder macromolecular interactions and enzyme kinetics. Moreover, they have to cope with additional stresses, such as high oxidative stress, high osmotic pressure and low nutrient availability [2]. Therefore, psychrophiles have evolved a multitude of adaptive strategies to counteract all these stressors [3]. Among these, a wide range of cold-active enzymes allow psychrophilic organisms to maintain high metabolic activity at low temperatures [4–6], and to survive in nutrient-poor environments such as Arctic and Antarctic marine sediments [7–9]. The organic matter in marine sediments of the polar regions is largely composed of high molecular weight polymers such as proteins and glycans, and its composition depends on several factors including the activity of bacterial communities or river runoff induced by the warming of soil's permafrost [10]. Glycans are polysaccharides produced by photosynthetic organisms, i.e. terrestrial plants, and marine algae, thus representing the largest carbon reservoirs for marine environments. Recently, xylans from terrestrial plants have been found in Baltic Sea sediments suggesting a transport of plant matter from land to sea [11].

Abbreviations: AF, AlphaFold 2; ΔG_{bind}, binding free energy; CD, circular dichroism; GH, glycoside hydrolase; GH3, glycoside hydrolase belonging to family 3; HPAEC, high-performance anion-exchange chromatography; M-GH1, GH1 identified in the genome of *Marinomonas* sp. ef1; M-GH3, GH3 identified in the genome of *Marinomonas* sp. ef1; M-GH42, GH42 identified in the genome of *Marinomonas* sp. ef1; MCMC, Bayesian Monte Carlo Markov Chain; ML, Maximum Likelihood; Mya, million years ago; oNPGal, *ortho*-Nitrophenyl β-D-galactopyranoside; *p*NPClb, *para*-nitrophenyl β-D-cellobioside; *p*NPFuc, *para*-nitrophenyl β-D-fucopyranoside; *p*NPGlc, *para*-nitrophenyl β-D-glucopyranoside; *p*NPGluA, *para*-nitrophenyl β-D-glucopyranosiduronic acid; *p*NPMan, *para*-nitrophenyl β-D-mannopyranoside; *p*NPXyl, *para*-nitrophenyl β-D-xylopyranoside; RMSD, root-mean-square deviation; rpoB, β-subunit of bacterial RNA polymerase; SEC, size-exclusion chromatography; T_m, unfolding transition midpoint temperature; T_{opt}, optimum temperature of the catalysis.

* Corresponding author.
E-mail address: marco.mangiagalli@unimib.it (M. Mangiagalli).

https://doi.org/10.1016/j.ijbiomac.2024.133449
Received 22 January 2024; Received in revised form 30 May 2024; Accepted 24 June 2024
Available online 27 June 2024

Among hydrolytic enzymes, glycoside hydrolases (GHs) play a key role in the degradation of glycans and in their metabolism [12,13]. GHs catalyze the hydrolysis of glycosidic bonds and are classified by the CAZy database based on their sequence identity in 183 different families [14]. Among GHs, family 3 (GH3) groups promiscuous enzymes with β-glucosidase, β-xylanase, β-glucuronidase, β-N-acteyl-hexoaminidase and α-L-arabinofuranosidase activities. These enzymes are widespread in plants, fungi and bacteria, where they perform diverse functions including carbohydrate degradation, cell wall remodeling and defense against pathogens [15,16]. Most GH3s have a catalytic *core* formed by two domains with an $(\alpha/\beta)_8$ (TIM) barrel structure and an $(\alpha/\beta)_6$ sandwich structure, respectively. The $(\alpha/\beta)_8$ (TIM) barrel domain contains the Asp residue acting as the catalytic nucleophile, while the $(\alpha/\beta)_6$ sandwich domain with a Glu acting as the catalytic acid/base residue in a double displacement mechanism. In addition to the catalytic *core*, the GH3 architecture sometimes displays accessory domains such as the fibronectin-like and the PA14 domains, whose functions are still unknown [17–21].

In this work, we investigated the structural and functional features of two GH3s identified in the genome of *Marinomonas* sp. ef1 (M-GH3s), an Antarctic bacterium [22,23]. Although these enzymes belong to the same glycoside hydrolase family, they show different evolutionary origins and activity towards natural oligosaccharides and polysaccharides.

2. Material and methods

2.1. Bioinformatic analysis

2.1.1. Search for GH3s in the genome of Marinomonas sp. ef1

Genes coding for GHs were searched in the genome of *Marinomonas* sp. ef1 (NCBI: GCA_002806845) with *hhmscan* from HHMER v3.3.2 [24], using the family/subfamily profile hidden Markov models from dbCAN2 [25], by using a restrictive *e*-value of e^{-30}. The predicted molecular weight was determined with Expasy ProtParam [26].

2.1.2. Distribution of GH3s in Marinomonas species

The genomes of *Marinomonas* species available in the NCBI BioSample database (https://www.ncbi.nlm.nih.gov/genbank/, accessed on 17/05/2023) were annotated with dbCAN2 to identify the sequence of putative GH3s. The sequence identity between M-GH3s and GH3s from other *Marinomonas* species was determined using an identity matrix calculated with pseqsid script (https://github.com/amaurypm/pseqsid).

The evolutionary lineage of *Marinomonas* species was inferred through phylogenetic analysis of the nucleotide sequences of 16S rDNA and of the gene encoding the β-subunit of bacterial RNA polymerase (rpoB). For each genome, multiple copies of the 16S rDNA genes were clustered at the 98 % sequence identity threshold and the resulting centroid was used for phylogenetic analysis performed by a Bayesian Monte Carlo Markov Chain method using BEAST v.1.10.4 [27]. The 16S rDNA and the rpoB genes of *Oceanispirillum sanctuary* and *Pseudospirillum japonicum* served as outgroups.

Divergence times of the two outgroup species were calibrated based on the data obtained from https://timetree.org/. According to these data, *O. sanctuary* and *P. japonicum* diverged 338 Mya and separated from the *Marinomonas* lineage 425 Mya. A single joint tree was built based on the 16S rDNA and the rpoB genes. A general time reversible matrix, specific to each gene, was used to model nucleotide substitution patterns. This matrix incorporated a (per-gene) proportion of invariant sites and a (per-gene) gamma-distributed rate variation with four categories. A strict molecular clock was employed, considering distinct evolution rates for each gene. The tree prior was set up as a calibrated Yule model, with time calibrations established based on the average of a normal distribution with standard deviation of 10 %. Default values were used for all other parameters. A Monte Carlo Markov Chain analysis was conducted, implementing three chains of 20 million steps each. The initial 50 % was discarded as "burn-in", and sampling was performed every 2000 steps. Tracer v.1.7.2 [28] was used for verifying the convergence of parameters within each chain.

2.1.3. Evolutionary history of GH3 enzymes

A phylogenetic analysis of the GH3 family was performed to predict the functional properties of the M-GH3s and to explore their evolution within the family. The sequences of characterized GH3s were retrieved from the CAZy database (http://www.cazy.org/GH3.html, accessed 15/06/2023), while information regarding substrate specificity was collected from literature. The 3D structure of every GH3 was acquired from either the PDB or AlphaFold (AF) databases (https://alphafold.ebi.ac.uk). For each GH3, the catalytic and accessory domains were manually trimmed and annotated (if average plDDT >0.75 for AF models). To exclude GH3 enzymes with high levels of similarity, the catalytic domains were clustered at 90 % sequence identity threshold with cd-hit 4.8.1 (https://github.com/weizhongli/cdhit). The sequences obtained listed in Table S1 were aligned with mafft v.7.471 [29]. The DASH option was employed to use the structural information to guide multiple sequence alignment [30]. The alignment obtained by trimming insertions shared by fewer than 25 % of sequences was used to estimate a rooted maximum likelihood tree with IQ-Tree v.2.2.2.7 software [31]. The phylogenetic analysis used the non-time reversible protein substitution matrix NQ.pfam (estimated from Pfam version 31 database [32]) and included a gamma distributed rate variation with four categories. Branch supports were obtained by using 1000 ultrafast bootstrap replicates [33] and transfer bootstrap expectation [34].

2.1.4. Identification of hypothetical metabolic pathways

The hypothetical metabolic pathways (https://biocyc.org/) were inferred from the *Marinomonas* sp. ef1 genome using GapSeq v1.2 [35]. The default parameters were retained and only "Good Blast" annotations of protein coding genes were considered. Only pathways 100 % complete are reported. OperonMapper [36] was used to identify operons of multiple genes involved in the same metabolic pathway of interest.

2.1.5. 3D structure prediction

The 3D molecular models of M-GH3_A and M-GH3_B were predicted using AlphaFold v.2.3.2 [37], v3 model and ColabFold v.1.5.2 (https://github.com/sokrypton/ColabFold) [38], in the oligomeric state determined by size exclusion chromatography (SEC, see below). The default ColabFold parameters were retained, allowing the use of structural templates from PDB and a more thorough sampling by activating the dropout option and using 3 different seeds. Only monomeric models with plDDT $\geq$0.90 or multimeric models with iPTM score $\geq$ 0.75 were retained and the best according to both metrics was selected. The side-chain torsion angles were refined using DiffPack [39], followed by an energy minimization and a short NVT classical molecular dynamics simulation (8000 steps) in TIP3P water molecules with 10 Å padding. This step was performed using OpenMM 7.7.0 [40] under the amber ff14SB force field [41], with a time step of 2.0 fs and at 30 °C.

2.1.6. Per-residue substrate binding affinity estimate

In silico models of D-cellobiose and D-xylobiose were prepared using the Avogadro 1.2.0 software [42] and minimized by a steepest-descent algorithm under the general Amber force field [43]. The AM1-BCC charges were assigned to the ligand by using Antechamber with the semi-empirical quantum mechanics method within the Ambertools21 package. Cellobiose and xylobiose were docked with the Gnina v.1.0.3 software [44] to the refined 3D models of M-GH3_A and M-GH3_B. The docking box (a square with 20 Å per side), representing the active site, was centered at catalytic Asp (D232 in M-GH3_A and D288 in M-GH3_B). The exhaustiveness of 32 was set to sample 10 docking poses with default root-mean-square deviation (RMSD) for clustering. The CNN score was used to rank docking poses. According to the retaining catalytic mechanism [45], a docking pose was considered catalytically

competent when the distances between: i) the O_γ of catalytic Asp and the C_1 of the non-reducing end sugar monomer and ii) the O_δ of catalytic Glu (E419 in M-GH3_A and E529 in M-GH3_B) and the glycosidic oxygen (facing the catalytic Glu) are <4.5 Å. The best catalytically competent docking pose was selected for subsequent steps of refinement through AdaptivePELE v1.7.2 [46]. Five independent replicas were performed and the result averaged. The system was prepared for each replica by hydrogenating residues with pdb2pqr v.3.2.0 [47] at the optimum pH of each enzyme. An MD engine (OpenMM v.7.7.0 [40]) was employed for propagation, with a "production length" of 4 ns, reporting every 200 ps, and performing 5 iterations. The ff14SB force field and explicit water solvent were added with a 10 Å padding. "minimization iterations" was set to "8000". All other parameters were at default.

The calculation of binding free energy (ΔG_{bind}) for the interaction between the AdaptivePELE samples of each GH3-substrate complex was estimated by applying molecular mechanics energies combined with the generalized Born and surface area continuum solvation, using mmpbsa.py [48] from the ambertools21 package. The prEFED protocol was used to decompose the binding free energy at residue level as described in [49]. The ΔG_{bind} values were averaged over the replica means. Residues were considered hot spots of interaction if their average energy contribution was ≤ -1.0 kcal·mol^{-1}.

2.2. M-GH3s expression and purification

Sequence coding for M-GH3s were optimized for expression in *Escherichia coli* cells, chemically synthesized (Genscript, Piscataway, NJ, USA) and cloned *in frame* with a C-terminal 6× His-Tag into the pET21 plasmid (EMD, Millipore, Billerica, MA, USA) between the *Nde*I and *Xho*I sites. These plasmids were used to transform *E. coli* BL21(DE3) cells (EMD, Millipore, Billerica, MA, USA). Recombinant M-GH3s were produced in Zym 5052 medium [50] with 100 μg/L of ampicillin (Merck, Darmstadt, Germany), for 24 h at 25 °C. Cells from 1 L of culture were harvested by centrifugation at 4000*g* for 10 min at 4 °C and the cell pellet was suspended in 15 mL of lysis buffer (50 mm sodium phosphate pH 8.0, 300 mM NaCl and 10 mM imidazole). Crude extracts were prepared by lysing the cells with a cell disruptor (Constant Systems Ltd., Daventry, UK) at $3.67{\cdot}10^5$ atm (25 kpsi) and clarified by centrifugation at 6000*g* for 10 min at 4 °C. M-GH3_A and M-GH3_B were purified from the soluble fraction of the cell lysate by metal ion affinity chromatography on Nickel-nitrilotriacetic acid agarose resin (Thermo Fisher Scientific, Waltham, MA, USA). M-GH3_C was extracted from the insoluble fraction of *E. coli* cells with 10 mL of extraction buffer (50 mM sodium phosphate pH 11.0, 300 mM NaCl and 8 M urea). Then, the pH of the solution was adjusted to pH 8.0 with HCl (0.1 M) and the solution was clarified by centrifugation at 5000*g* for 10 min at room temperature. The clarified solution was loaded in a column containing 1 mL of Nickel-nitrilotriacetic acid agarose resin and a series of washes at decreasing concentrations of urea (from 8 M to 0 M) were carried out. Samples were eluted with 2 mL of elution buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl and 250 mM imidazole).

Elution fractions containing highest protein concentration were pooled and buffer-exchanged with 100 mM sodium phosphate buffer (PB), pH 7 by pre-packed PD10 columns (GE Healthcare, Little Chalfont, UK). Samples were concentrated with Amicon Ultra centrifugal filters (Merck Millipore, Burlington, US) to a final concentration of 2 mg/mL of protein. Protein concentration was determined by Bradford protein assay (Bio-Rad, California, USA), using bovine serum albumin as a standard.

2.3. Activity assay

Enzymatic assays were carried out using a panel of substrates: *para*-nitrophenyl β-D-glucopyranoside (*p*NPGlc), *para*-nitrophenyl α-D-glucopyranoside (*p*NPαGlc) *para*-Nitrophenyl β-D-galactopyranoside (*p*NPGal), *para*-nitrophenyl β-D-xylopyranoside (*p*NPXyl), *para*-nitrophenyl β-D-fucopyranoside (*p*NPFuc), *para*-nitrophenyl α-L-arabinofuranoside (*p*NPAra), *para*-nitrophenyl β-D-cellobioside (*p*NPClb), *para*-nitrophenyl β-D-mannopyranoside (*p*NPMan) and *para*-nitrophenyl β-D-glucopyranosiduronic acid (*p*NPGlcA). Reactions containing 0.01 IU of enzyme were performed in PB and stopped, after 3 min, by adding an equal volume of 1 M sodium carbonate pH 11. The absorbance was measured at 405 nm (molar extinction coefficient: 18.6 mM^{-1} ·cm^{-1}) using a Jasco V-770 UV/NIR spectrophotometer (JASCO Europe, Lecco, Italy). One unit of enzyme activity was defined as the amount of the enzyme catalyzing the formation of 1 μmol of *para*-nitrophenol per minute under saturating substrate conditions (10 mM) at 25 °C.

The optimal catalysis conditions were determined in PB using 10 mM of *p*NPGlc and *p*NPXyl as substrates, for M-GH3_A and M-GH3_B, respectively. The optimal pH of catalysis was measured in the pH range 3.0–10.0 in Britton–Robinson buffer, at 50 °C for M-GH3_A and at 60 °C for M-GH3_B. The optimal temperature of catalysis (T_{opt}) was recorded in the temperature range 10–90 °C, at pH 7.5 for M-GH3_A and at 6.5 for M-GH3_B.

The kinetic parameters were determined at optimal catalysis conditions on *p*NPGlu, *p*NPXyl from 0.1 mM to 20 mM for M-GH3_A and M-GH3_B, and on *p*NPGal in the range from 2 mM to 30 mM for M-GH3_A only. For each substrate concentration, four time points were obtained by stopping the reactions at 30-s intervals. The angular coefficient of the resulting linear regression was used to calculate V_0. To calculate the kinetic parameters, the V_0 values from three independent measurements were plotted against the substrate concentration and fitted with the ORIGINLAB software (OriginLab Corporation, Northampton, MA, USA), using the Michaelis-Menten equation. The resulting kinetic parameters were reported with their fitting errors.

2.4. M-GH3s thermal stability assays

Thermal denaturation experiments were carried out by monitoring the circular dichroism (CD) signal at 200 nm in the temperature range from 10 °C to 90 °C with a Jasco J815 spectropolarimeter (JASCO Europe, Lecco, Italy). Measurements were performed in a 0.1 cm path-length quartz cuvette and a temperature slope of 1 °C/min.

Long-term thermal stability was assessed by measuring the residual activity at T_{opt}, after incubating M-GH3s (protein concentration: 0.5 mg/mL) in PB, pH 7.5, at 5 °C, 25 °C and 35 °C. Experiments were performed in triplicate and reported as a mean ± standard deviation.

2.5. M-GH3s quaternary structure

The quaternary structure of M-GH3s was determined by SEC analysis using an NGC Quest 10 Plus Chromatography System (Bio-Rad, California, USA) equipped with a Superdex 10/200 column (Cytiva, Marlborough, US) with a cutoff of 10–600 kDa, saline phosphate buffer (25 mM, 150 mM NaCl, pH 7) as the mobile phase, and 1.0 mL/min as the flow rate. The calibration curve, reported in Fig. S1, was carried out with the following molecular weight standards (Cytiva, Marlborough, US): Ribonuclease A (MW: 13.7 kDa), Ovalbumin (MW: 43.0 kDa), Aldolase (MW: 158.0 kDa), Ferritin (MW: 440.0 kDa), and Thyroglobulin (MW: 669.0 kDa).

2.6. Polysaccharides degradation

Hydrolysis of polysaccharides and oligosaccharides was evaluated on carbohydrates purchased from Megazyme (Megazyme International Bray, Ireland), Merck (Merck Darmstadt, Germany) and VWR Chemicals (VWR International, Radnord, USA), including carboxy-methyl-cellulose (VWR code: 22525.296), xylan (Megazyme code: P-XYLNBE), arabinoxylan (Megazyme code: P-WAXYL), xyloglucan (Megazyme code P-XYGLN), mannan (Megazyme code: P-MANIV), κ-carrageenan (Sigma-Aldrich code: 22048), galactomannan (Megazyme code: P-GGMMV), cellobiose (Sigma-Aldrich code: 1.02352) and cellotetraose (Megazyme

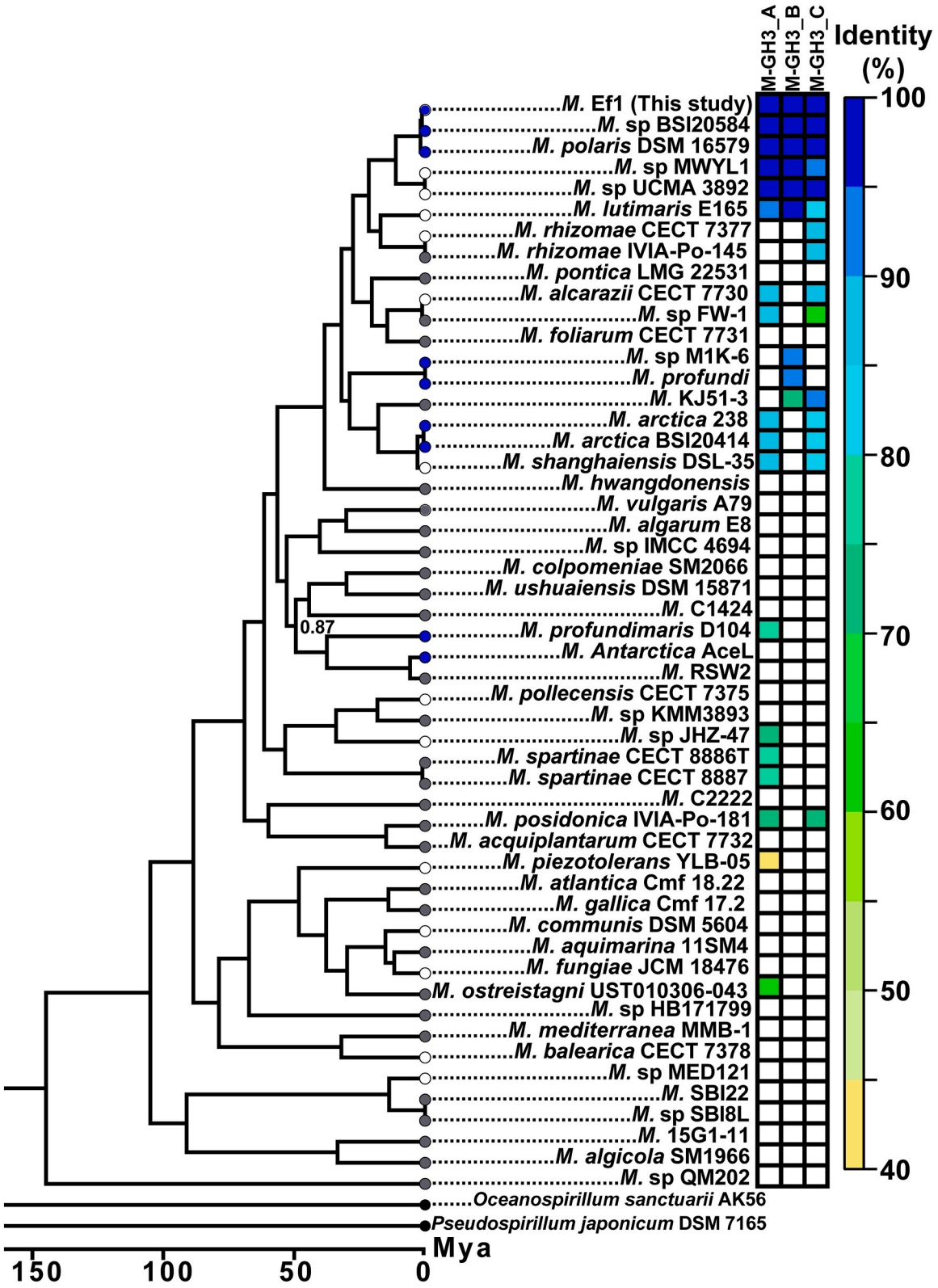


Fig. 1. Distribution of GH3s in *Marinomonas* spp. Sequences encoding putative GH3s were extracted from the genome of *Marinomonas* spp. available in the NCBI BioSample database. The heatmap on the right displays the sequence identity of *Marinomonas* spp. GH3s in comparison to M-GH3s. White cells represent enzymes with <40 % identity with M-GH3s or those missing in the genome. Phylogeny of *Marinomonas* spp. was performed based on 16S rDNA and rpoB genes using BEAST v1.10.4 software [33] and two outgroup time calibrations from Timetree of Life. The branch lengths represent the median of the posterior distribution. The posterior distribution of nodes is reported if <1.0. Psychrophilic and mesophilic species are highlighted with blue and gray dots, respectively; empty dots indicate *Marinomonas* spp. isolated from sites whose environmental conditions are unknown. The phylogenetic tree was generated with FigTree v1.4.4 (https://github.com/rambaut/figtree/releases). Maa: millions of years ago.

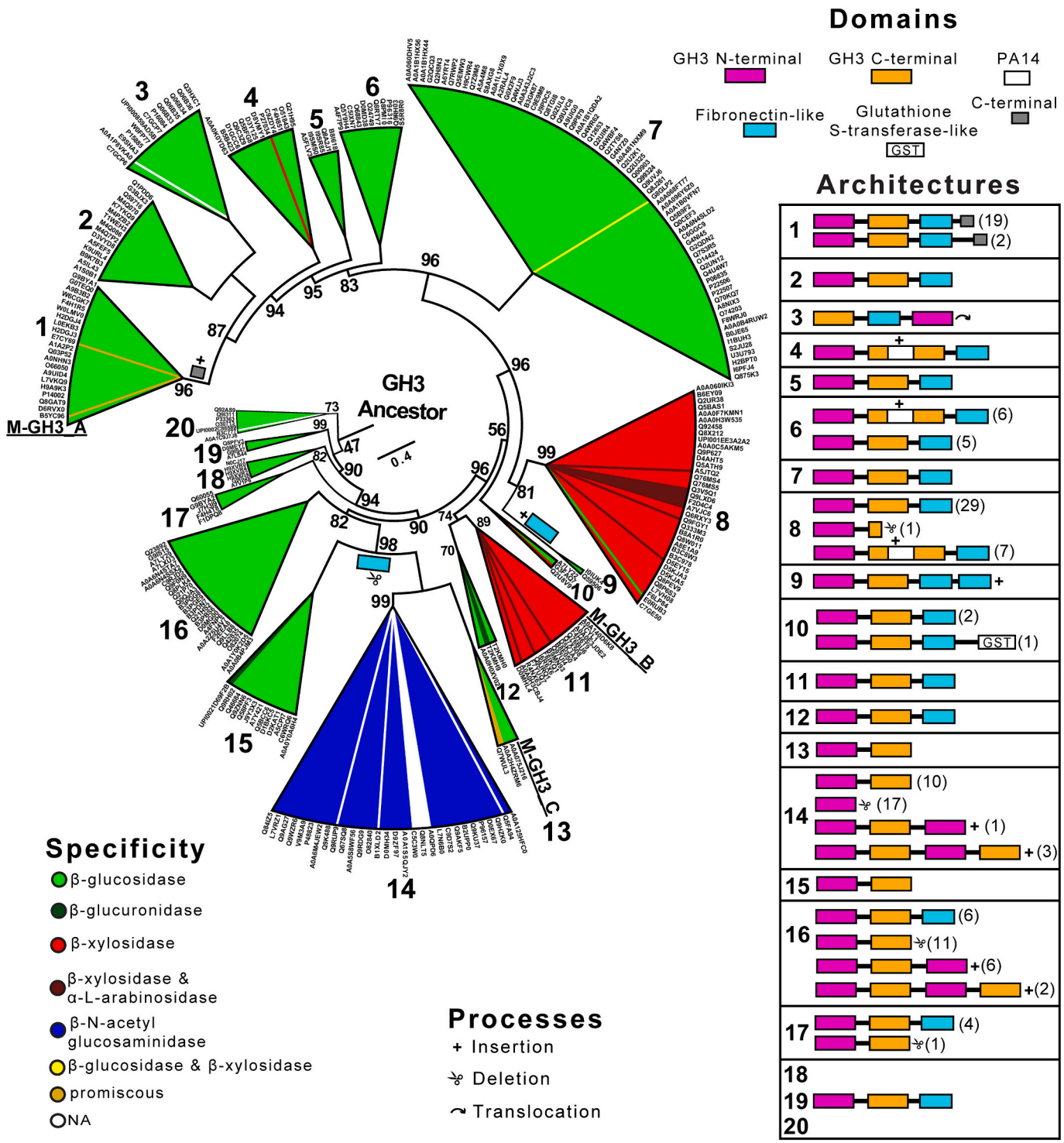


Fig. 2. Molecular phylogeny of GH3s. The rooted maximum likelihood phylogenetic tree was obtained by aligning the catalytic domains of 300 characterized M-GH3s extracted from the CAZy database (http://www.cazy.org/GH3.html). To compute the branch support values, 1000 ultra-fast bootstrap replicates were performed, and support values are reported if <100. The GH3s from *Marinomonas* sp. ef1 are underlined. Insertion, deletion, and translocation of the domains was reported with respect to the inferred evolutionary history of GH3 catalytic domains. The phylogenetic tree was generated with FigTree v1.4.4 (https://github.com/rambaut/figtree/releases) and customized.

The table contains the architectures of M-GH3s manually annotated from the 3D structure available in PDB or AF predicted 3D structures (average domain-wise pLDDT >0.75) available in AlphaFold-DB (https://alphafold.ebi.ac.uk). Numbers in bold indicate the cluster number in the phylogenetic tree, while those in brackets indicate the number of sequences with that specific architecture, in case more than one is present per-cluster.

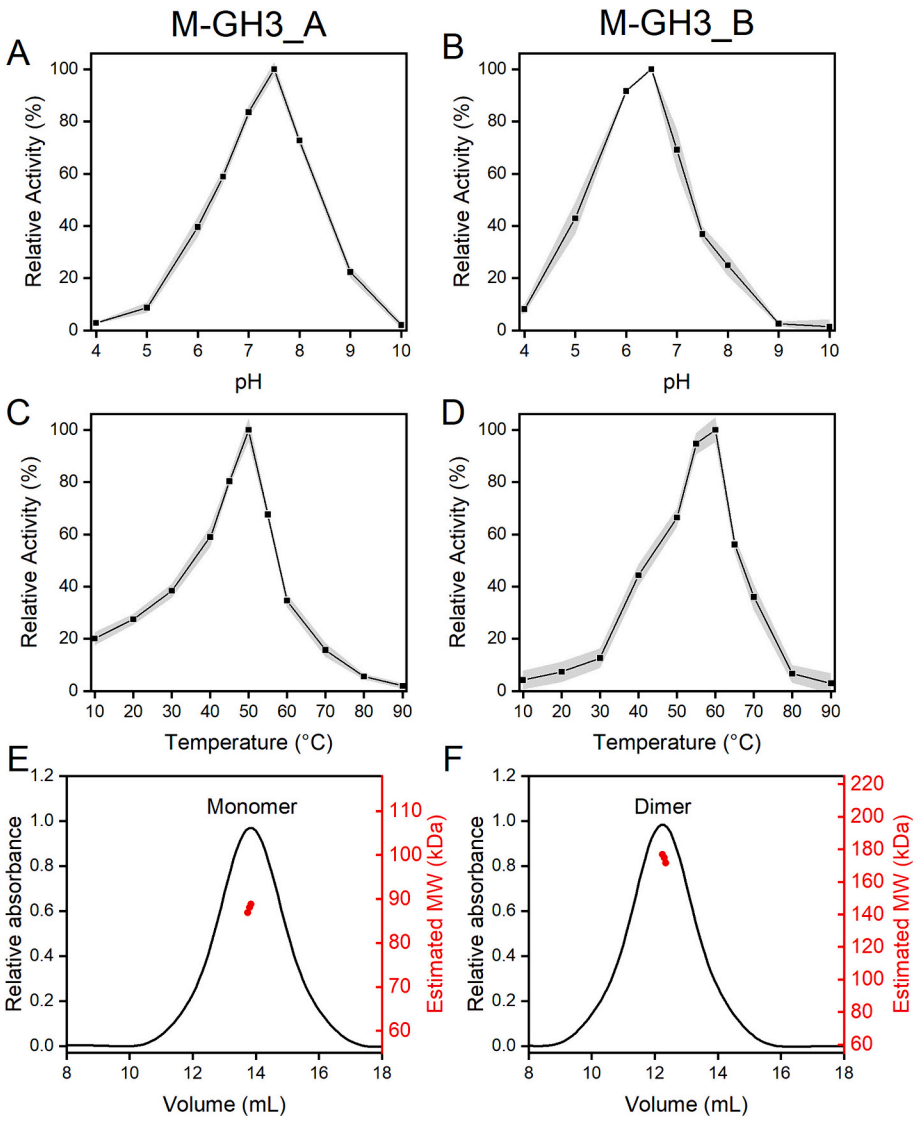


Fig. 3. Biochemical features of M-GH3s. Effects of pH on the activity of M-GH3_A (A) and M-GH3_B (B). Temperature profile of M-GH3_A (C) and M-GH3_B (D). The activity of M-GH3_A and M-GH3_B were monitored using pNGlc and pNXyl as a substrate, respectively. All the experiments were performed in quadruplicate and the shadowed area refers to the standard deviation of the data ($n = 4$). SEC analysis of M-GH3_A (E) and M-GH3_B (F). SEC were performed in PB, one of three independent measurements is shown. The red dots represent the MW estimated from three independent measurements using the calibration curve shown in Fig. S1.

code: O-CTE). Reactions containing 1 % of polysaccharide or oligosaccharide and 1 mg/mL of each purified M-GH3 were performed in PB at 25 °C, at 800 rpm mixing speed in a thermal shaker (Eppendorf, Hamburg, Germany). After 2 h of incubation, hydrolysis products were analyzed by high-performance anion-exchange chromatography (HPAEC) on a Dionex ICS–6000 Ion Chromatography System coupled with pulsed amperometric detection (PAD) and equipped with a CarboPac PA 210 column (Dionex Corporation, CA, USA). Elution was performed in KOH gradient. An initial phase of 6.5 min of isocratic elution at 12 mM was followed by a linear gradient from 12 mM to 100 mM in 5 min. 100 mM KOH was held for 6 min, then the initial condition of 12 mM KOH was reached in 0.5 min. Elution was performed at a flow rate of 0.6 mL/min. Calibration curves were prepared using pure monosaccharides (glucose, xylose, arabinose), dissolved in Milli-Q water. Chromeleon® (6.8) software was utilized for data processing. The HPAEC chromatograms of standard glucose, arabinose and xylose are shown in Fig. S2.

3. Results

3.1. The genome of Marinomonas sp. ef1 contains three different GH3s

The genome of *Marinomonas* sp. ef1 contains 34 genes coding for GHs belonging to 19 different families. Among these, three are classified in the GH3 family, and are hereafter reported as M-GH3_A (NCBI: WP_100636789.1, Length: 800 amino acids, theoretical MW: 87.5 kDa), M-GH3_B (NCBI: WP_100636060.1, Length: 788 amino acids, theoretical MW: 86.1 kDa) and M-GH3_C (NCBI: WP_198515313.1, Length: 553 amino acids, theoretical MW: 61.9 kDa) and collectively referred to as M-GH3s. M-GH3s are conserved among closely related psychrophilic *Marinomonas* spp., but absent in genetically distant species, suggesting that they have been acquired through horizontal gene transfer events (Fig. 1). Overall, M-GH3s share low sequence identity (< 25 %) and have three different architectures (Fig. 2).

To study the evolutionary history of M-GH3s and infer their substrate specificity, a phylogenetic analysis was performed using a dataset containing all the characterized GH3s available in the CAZy database. The resulting rooted tree shows two main lineages (Fig. 2), one containing

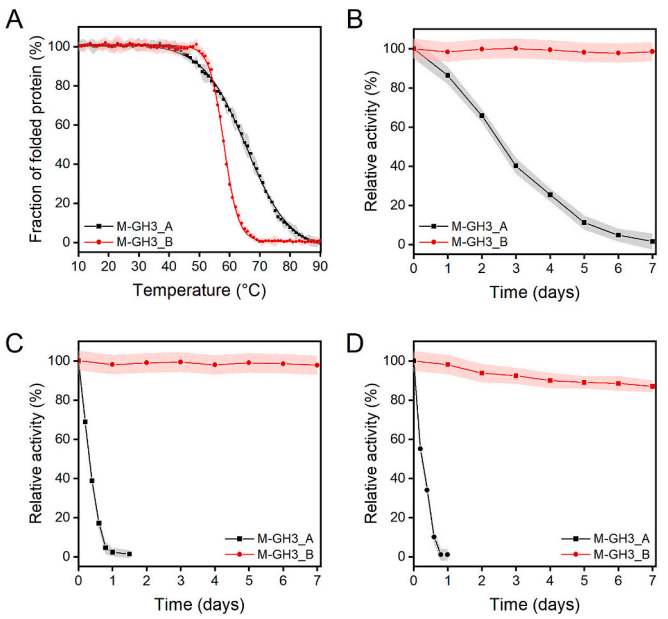


Fig. 4. Thermal stability of M-GH3s. A) Thermal stability of M-GH3s determined by CD spectroscopy. Ellipticity values were recorded at 205 nm during heating from 10 to 90 °C. The initial CD signal was taken as 100 % for normalization. Long-term thermal stability was measured by incubating enzymes at 5 °C (B), 25 °C (C) and 35 °C (D). M-GH3_A (black line), M-GH3_B (red line). All the experiments were performed in quadruplicate and the shadowed area refers to the standard deviation of the data (n = 4).

exclusively enzymes with β-glucosidase activity (clusters 1–7), and other, more heterogeneous, grouping mainly β-glucosidases (clusters 9, 10, 12, 13 and 15–20), β-xylosidases (clusters 8 and 11) and β-*N*-acetylhexosaminidase (cluster 14). M-GH3_A, M-GH3_B and M-GH3C are nested in clusters 1, 10 and 13, suggesting they have β-glucosidase, β-xylosidase and β-glucosidase activity, respectively. The sequence identity between the M-GH3s and the members of each cluster ranges from 31 % to 52 % (Table S2).

The architecture of known GH3s (Fig. 2) includes a two-domain catalytic *core* consisting of an N-terminal catalytic domain (in magenta in Fig. 2), a C-terminal catalytic domain (in orange in Fig. 2), and a fibronectin-like domain (in cyan in Fig. 2). Notably, the fibronectin-like domain is lacking in clusters 13, 14 and 15, whereas it is duplicated in cluster 9. Additional domains observed in the GH3s architectures include the C-terminal domain (inserted in cluster 1), the PA14 domain (clusters 4, 6 and 8), and the glutathione S-transferase-like domain (in cluster 10). The architecture of each M-GH3 reflects that of the cluster to which it belongs.

3.2. M-GH3s have different biochemical features

M-GH3s were recombinantly produced in *Escherichia coli* cells. M-GH3_A and M-GH3_B were obtained as soluble proteins and purified by affinity chromatography with a yield of 5 mg and 15 mg per liter of culture, respectively. Recombinant M-GH3_C was insoluble, and the refolding of solubilized aggregates resulted in a partially folded and inactive protein (data not shown). For this reason, we will describe the structural and functional characterization of M-GH3_A and M-GH3_B.

Activity assays point out that M-GH3_A and M-GH3_B exhibit the highest activity at pH 7.5 and 6.5, respectively (Fig. 3A and B). M-GH3_A shows highest activity at 50 °C and retains 20 % activity at 10 °C, whereas M-GH3_B has a T_{opt} of 60 °C and maintains only 5 % activity at 10 °C (Fig. 3C and D).

The thermal stability of M-GH3s was investigated by combining thermal denaturation experiments with long-term thermal stability assays. Thermal denaturation experiments, performed by CD spectroscopy, show that M-GH3_A (T_m: 66.8 ± 0.9 °C, Fig. 4A) has a higher unfolding transition midpoint temperature than M-GH3_B (T_m: 59.5 ± 1.3 °C, Fig. 4B). Long-term thermal stability assays, carried out at 5 °C, 25 °C and 35 °C, suggest that M-GH3_A is more thermolabile than M-GH3_B. Indeed, M-GH3_A completely loses its activity after 6 days of incubation at 5 °C and after 8 and 4 h at 25 °C and 35 °C, respectively (Fig. 4B–D). On the other hand, M-GH3_B maintains its activity for 7 days at all tested temperatures (Fig. 4B–D). Overall, our results indicate that M-GH3_A is a bona fide cold-active enzyme, while M-GH3_B is endowed with some mesophilic traits such as thermostability and low activity in the cold.

Table 1
Substrate specificity of M-GH3s. Degradation yields were determined by applying a calibration curve with diverse concentrations of standards.

	Specific activity (U/mg)		Degradation yields (mg/L)	
	Colorimetric substrates		Natural substrates	
	M-GH3_A	M-GH3_B	M-GH3_A	M-GH3_B
*p*NPGlc	20.2 ± 2.1	1.1 ± 0.1		
*p*NPαGlc	–	–		
*p*NPGal	2.9 ± 0.8	–		
*p*NPXyl	6.6 ± 0.1	11.0 ± 2.5		
*p*NPAra	0.8 ± 0.2	9.3 ± 2.0		
*p*NPMan	0.1 ± 0.03	–		
*p*NPFuc	2.8 ± 0.1	–		
*p*NPGlcA	0.5 ± 0.1	0.02 ± 0.01		
*p*NPCell	3.6 ± 0.2	0.02 ± 0.01		
Cellulose			–	–
Cellobiose			3204.3 ± 61.3	380.2 ± 5.2
Cellotetrose			Glc: 251.9 ± 6.2	Glc: 10.6 ± 0.4
Laminarin			–	–
Xylan			2.9 ± 1.0	170.8 ± 9.9
Arabinoxylan			Ara: 2.8 ± 0.5 Xyl: 5.2 ± 1.0	Ara: 83.6 ± 1.2 Xyl: 280.8 ± 8.2
Mannan			–	–
Galactomannan			–	–
κ-carrageenan			–	–

3.3. M-GH3s displays different substrate specificity

The hydrolytic activity of M-GH3s was tested on *para* nitrophenyl glycosides, cellobiose, cellotetraose and polysaccharides (cellulose, laminarin, xylan, arabinoxylan, mannan, galactomannan and κ-carrageenan). M-GH3_A exhibits the highest specific activity on *p*NPGlc, *p*NPGal, *p*NPXyl, *p*NPFuc and *p*NPClb, while displaying lower activity on *p*NPMan, *p*NPAra and *p*NPGlcA, and no activity on *p*NPαGlc (Table 1). On the other hand, M-GH3_B exhibits a narrower substrate specificity, displaying high specific activity on *p*NPXyl and *p*NPAra, poor activity on *p*NPGlc and negligible or no activity on other substrates (Table 1). The analysis of the kinetic parameters indicates that M-GH3_A has a higher catalytic efficiency ($kcat/K_M$) towards *p*NPGlc than *p*NPXyl, suggesting that this enzyme is a β-glucosidase rather than a β-xylosidase (Table 2). In contrast, the catalytic efficiency of M-GH3_B towards *p*NPXyl is 2 and 14 times higher than those determined with *p*NPAra and *p*NPGlc as substrates, respectively (Table 2), indicating that this enzyme is likely a β-xylosidase.

Both M-GH3s degrade xylan and arabinoxylan (Fig. 5) whereas they are not active on mannan, galactomannan, laminarin and κ-carrageenan (Table 1). The analysis of xylan and arabinoxylan degradation products indicates that both enzymes release arabinose and/or xylose although with different yields (Fig. 5, Table 1). This suggests that both M-GH3s

Table 2
Kinetics parameters of M-GH3s. The kinetics parameters were determined at optimal catalysis conditions.

	M-GH3_A			M-GH3_B		
	*p*NPGlc	*p*NPGal	*p*NPXyl	*p*NPGlc	*p*NPAra	*p*NPXyl
K_M (mM)	1.0 ± 0.1	12.3 ± 0.6	1.1 ± 0.1	11.5 ± 0.7	9.8 ± 0.8	6.0 ± 0.5
k_{cat}^{App} (s^{-1})	162.6 ± 3.1	23.3 ± 2.3	40.5 ± 1.2	14.0 ± 0.7	81.9 ± 3.7	102.9 ± 2.2
k_{eff} (mM^{-1} s^{-1})	162.6 ± 3.1	1.8 ± 0.1	36.8 ± 0.2	1.2 ± 0.1	8.3 ± 0.2	17.2 ± 0.5

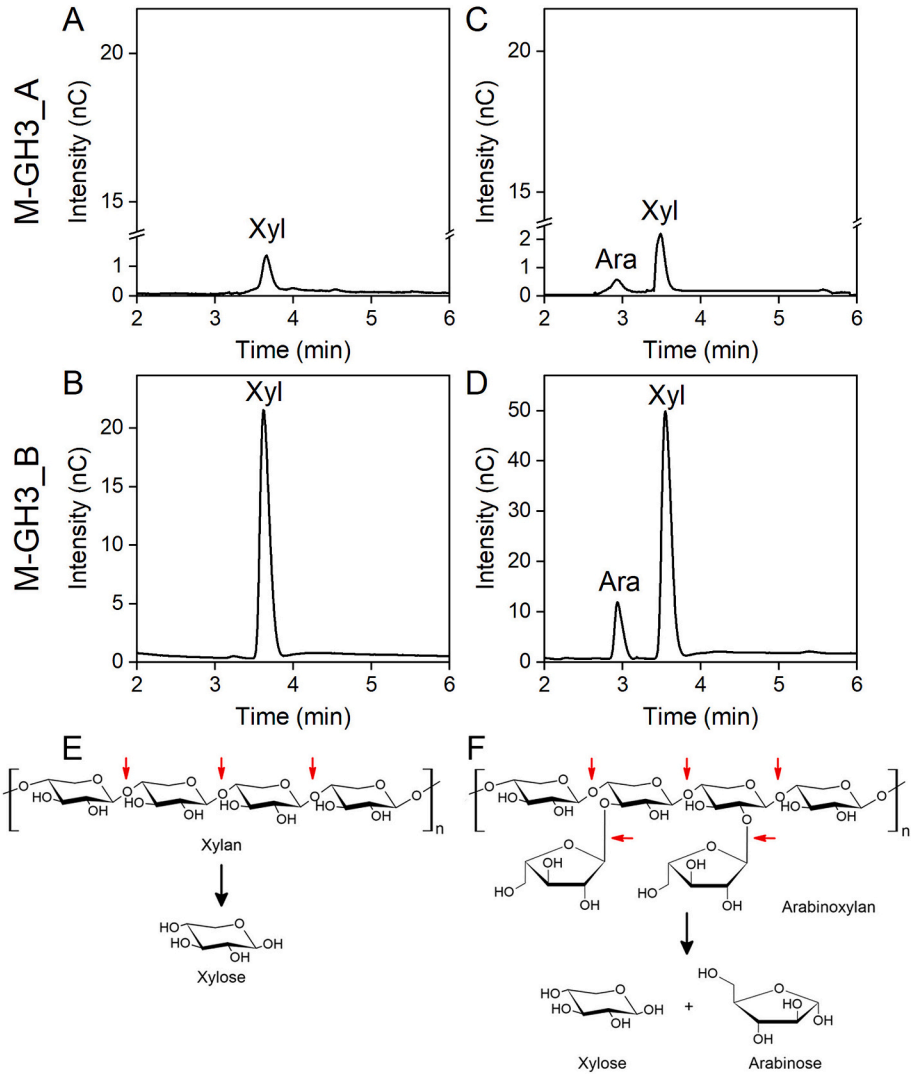


Fig. 5. Polysaccharides degradation. Degradation of xylan in the presence of M-GH3_A (A) and M-GH3_B (B). Hydrolysis of arabinoxylan in the presence of M-GH3_A (C) and M-GH3_B (D). Reactions were carried out in triplicate at 25 °C under shaking for 2 h and analyzed with HPAEC. The pattern of xylan and arabinoxylan degradation was reported in panel E and F, respectively. Ara = arabinose; Xyl = xylose. The chromatograms of the standards are reported in Fig. S2.

act as exoglycosidases and are also active on α1–3 or α1–2 L-arabinofuranosidic bonds (Fig. 5E and F).

The hydrolysis of cellulose and its derivatives, namely cellobiose and cellotetraose is more complex (Fig. 6 and Table 1). While M-GH3_A is catalytically active towards both cellobiose (glucose yield: 3.2 ± 0.6 g/L) and cellotetraose (glucose yield: 251.9 ± 6.2 mg/L), M-GH3_B does not show significant activity towards these compounds. Both enzymes are inactive towards cellulose.

The hypothetical metabolic pathways of monosaccharides derived from GH3-catalyzed hydrolysis (i.e. glucose, xylose and arabinose) were predicted by genome analysis using GapSeq v1.2 [35]. While glucose can enter the glycolytic pathway (Fig. S3), two main pathways can be hypothesized for xylose and arabinose metabolism. The D-xylose metabolic pathway (BioCyc ID: XYLCAT-PWY, Fig. S3) includes the *xylA* gene coding for a xylose isomerase and an operon that contains genes responsible for xylose metabolism and transport (Fig. S3). The xylulose 5-P produced at the end of this pathway can then enter other metabolic pathways, such as the pentose phosphate pathway. The L-arabinose metabolic pathway (BioCyc ID: PWY-5515, Fig. S3) is more elusive since the genes putatively involved are distributed across three different operons. It was hypothesized that the enzymes involved in arabinose metabolism collectively convert L-arabinose to xylitol, an osmolyte typically associated with cold stress resistance [51,52].

Overall, the results demonstrate that both M-GH3s have *exo*-activity with distinct specificities towards colorimetric and natural substrates. M-GH3_A is a β-glucosidase with a broad substrate specificity, while M-

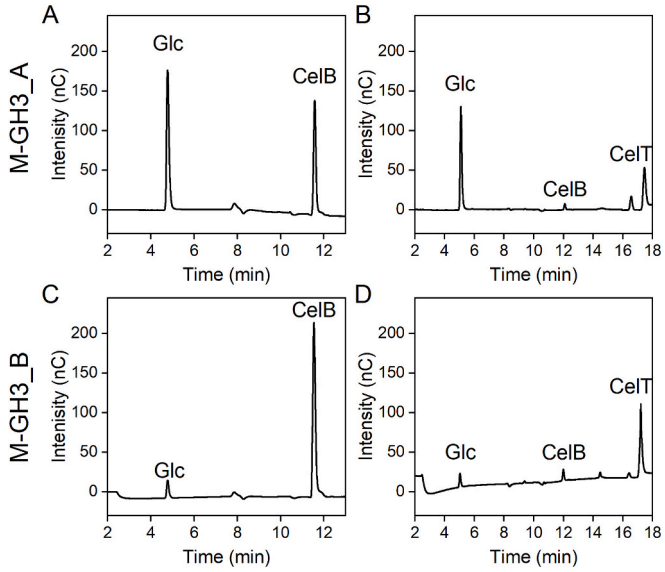


Fig. 6. Cellobiose and cellotetraose degradation. Degradation of cellobiose in the presence of M-GH3_A (A) and M-GH3_B (C). Hydrolysis of cellotetraose in the presence of M-GH3_A (B) and M-GH3_B (D). Reactions were carried out in triplicate at 25 °C under shaking for 2 h and analyzed with HPAEC. Glc = glucose; CelB = cellobiose; CelT = cellotetraose.

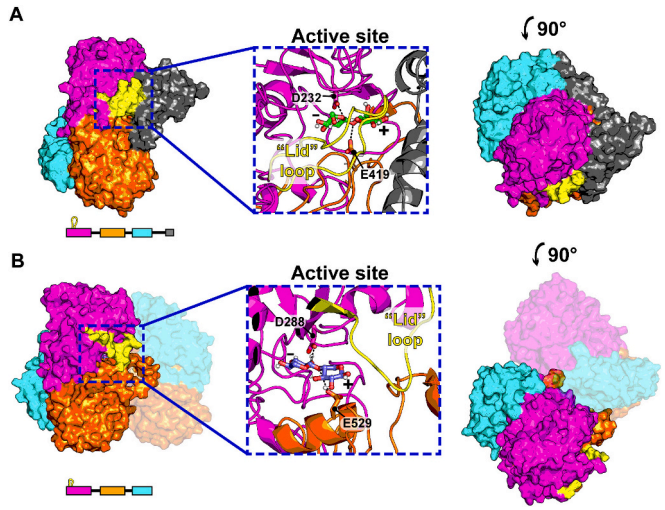


Fig. 7. 3D models of M-GH3s. 3D models of M-GH3_A (A) and M-GH3_B (B) predicted with AF2 (see main text) represented in surface style. The domains are colored according to the architecture reported in Fig. 2. The active site containing xylobiose and cellobiose is represented in ribbon style. The oligomerization state of M-GH3s was determined by SEC analysis (visualized on the left). Models were rendered using Pymol v.2.5.0 (Schrödinger, LLC, New York, NY).

GH3_B is a β-xylosidase that is also active on α1–3 or α1–2 L-arabinofuranosidic bonds. It is worth noting that the experimental results support the activities predicted by the phylogenetic analysis. In addition, the monosaccharides resulting from GH3-catalyzed hydrolysis can be used by *Marinomonas* sp. ef1 either as a carbon source or for producing osmolytes.

3.4. Different substrates-active site interactions determine the specificity of M-GH3s

To investigate the structural reason for the different substrate specificity of M-GH3_A and M-GH3_B, we modeled their 3D structures with AF2 [37] and performed molecular docking simulations. The enzyme-substrate complex was refined with AdaptivePELE coupled with a per-residue end-state binding free energy estimate. The per-residue accuracy value indicates good quality of both models (plDDT of 0.949 for M-GH3_A and of 0.944 for M-GH3_B; iPTM score of 0.946 for M-GH3_B), making them suitable for further analyses. M-GH3_A is monomeric and consists of four domains (Figs. 3E and 7A): a $(\beta/\alpha)_8$ TIM barrel domain (residues G61 to L286, in magenta), an $(\alpha/\beta)_6$ sandwich domain (residues V319 to Y549, in orange), fibronectin-like domain (residues F552 to A669, in cyan), a long linker (c.a. 40 amino acids) followed by an additional C-terminal domain (residues L670 to I800, in gray). The fibronectin-like and the C-terminal domains are additional domains with unknown structural and functional roles. The GH3s belonging to cluster 1 whose 3D structures are known, namely BglB from *Acetivibrio thermocellus* ATCC 27405 (PDB: 7MSE, sequence identity: 48.4 %) and PstG from *Paenibacillus relictisesami* (PDB: 8J9F, sequence identity: 49.8 %), are dimers in which the quaternary structure is stabilized by the interaction of the C-terminal domain of one protomer with the (β/α)8 TIM barrel domain of the other protomer [53]. The monomeric state of M-GH3_A is probably due to the length of the linker connecting the fibronectin-like and C-terminal domains, which is 30 amino acid residues longer than those of BglB and PstG.

M-GH3_B is a dimer, and each monomer consists of three domains (Figs. 3F and 7B): a $(\beta/\alpha)_8$ TIM barrel domain (residues T23 to V354, in magenta), an $(\alpha/\beta)_6$ sandwich domain (residues V390 to Y648, in orange) and a fibronectin-like domain (residues E686 to A755, in cyan). The $(\beta/\alpha)_8$ TIM barrel domain and $(\alpha/\beta)_6$ sandwich domain form the catalytic *core* of both enzymes. The quaternary structure of M-GH3_B is similar to that of the GH3 from *Thermotoga maritima* (PDB: 7ZB3), which belongs to cluster 11 and shares 44.2 % of sequence identity.

The putative catalytic residues are an Asp residue (D232 in M-GH3_A and D288 in M-GH3_B), which acts as a nucleophile and a Glu residue (E419 in M-GH3_A and E529 in M-GH3_B), which acts as an acid/base residue. In both M-GH3 enzymes, a lid loop, spanning residues Q53 to A64 in M-GH3_A, and R47 to I60 in M-GH3_B (in yellow in Fig. 7A and B), probably controls the access to the active site. In M-GH3_A, this lid loop, along with the additional C-terminal domains completely covers the entrance of the active site. Conversely, the absence of the additional C-terminal domain in M-GH3_B results in a small opening that might help the access of polysaccharides into the active site.

Molecular docking simulations were employed to investigate the interactions between M-GH3s and cellobiose/xylobiose, the substrates that support highest specific activities. Through this analysis, we identified the amino acidic residues that are predicted to interact with both cellobiose and xylobiose. Molecular docking simulations revealed differences in the substrate binding residues of the two enzymes, which are F27, D45, R51, R164, M165, W233, S352, M765 and F777 in M-GH3_A (Fig. 8A), and W35, R73, M117, H222, D288, Y289, Y424, H429, L432, L521, F522 in M-GH3_B (Fig. 8B). The same residues were identified using a complementary approach, namely AF2BIND [54], which makes us confident about our docking analysis. In particular, M-GH3_A residues D45 and W233 have binding free energy (ΔG_{bind}) more negative when interacting with cellobiose than xylobiose and could play a key role in the interaction between M-GH3_A and this sugar, making it favorable compared to that with xylobiose (Fig. 8C). The conservation analysis indicates that these two residues are highly conserved in cluster 1 enzymes (Fig. 8E) and play a key role in the coordination of co-crystallized glycerol and glucose molecules contained in the active site of PstG and of GlyA1, respectively (Fig. S4A) [53]. On the other hand, in M-GH3_B both substrates show similar interaction energies, with the conserved residues R73, E111, Y424 and H429 contributing the most to the positioning of the sugar moieties in the active site (Fig. 8D and E). Slight differences in the binding modes of the two substrates were observed; the distance between the Oγ of the catalytic Asp with C1 of the disaccharide suggested that for M-GH3_B the xylobiose is in a

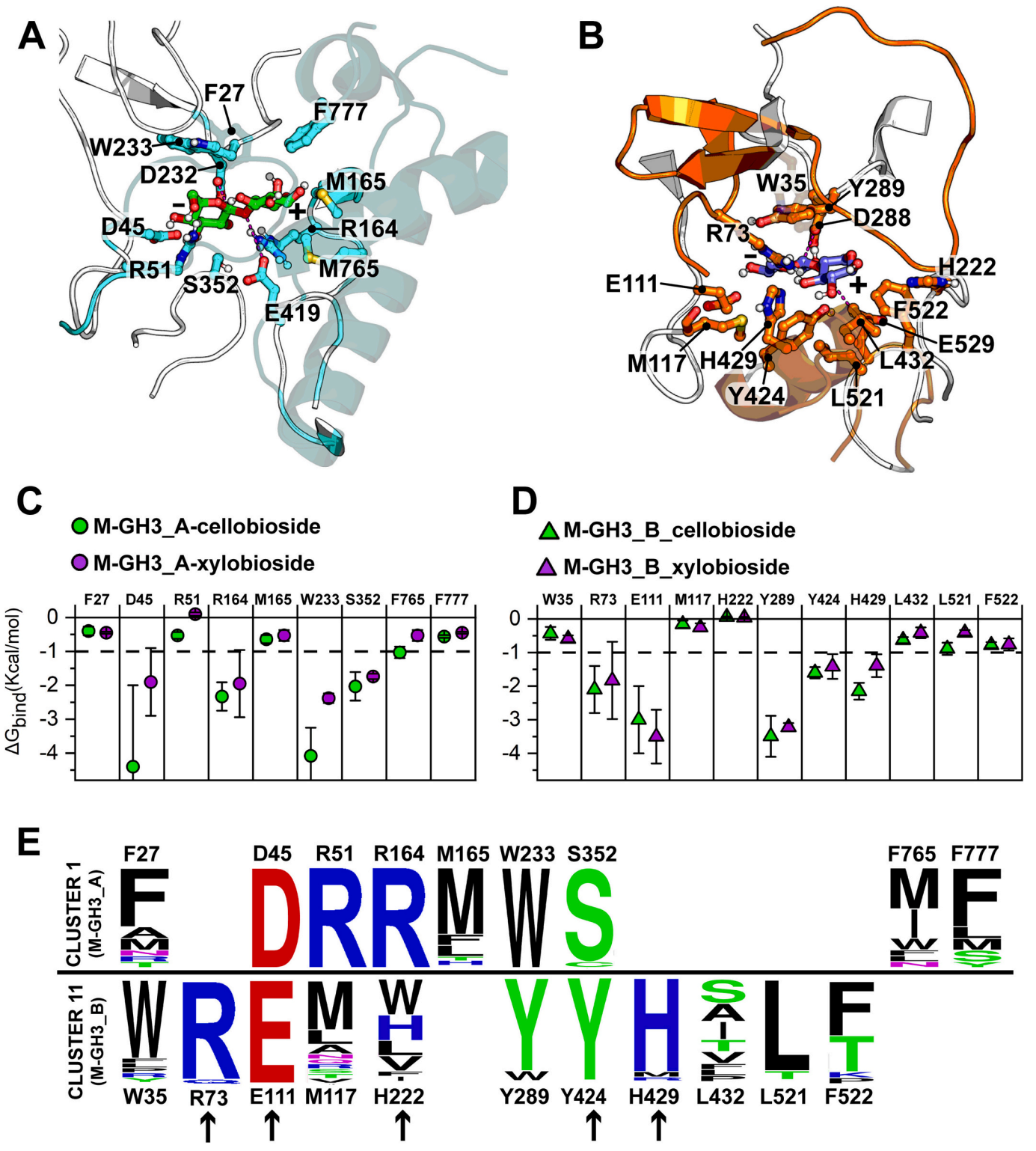


Fig. 8. Interactions of cellobiose and xylobiose substrates to M-GH3s. *In silico* docking analysis of M-GH3_A with cellobiose (A) and of M-GH3_B with xylobiose (B) Estimation of end-state MM-GBSA average binding free energy (ΔG_{bind}) of M-GH3_A (C) and M-GH3_B (D) in complexes with cellobiose and xylobiose. Only residues for which the ΔG_{bind} was < -0.5 kcal/mol for at least one enzyme-substrate combination are reported. A dashed line indicates the threshold used to identify hotspots of interaction. The error bars report SD from three independent MD AdaptivePELE simulations. (E) The evolutionary conservation of interacting residues in the M-GH3 cluster 1 and 11 are visualized as sequence logos. The most relevant sites for comparison are indicated by black arrows.

catalytically more favorable position than cellobiose (Table 3). These residues are conserved and are involved in the coordination of a xylobiose molecule co-crystallized with the GH3 from *T. maritima* (PDB: 7ZB3) (Fig. S4B).

4. Discussion

Marinomonas sp. ef1 is a psychrotolerant bacterium isolated from the microbial consortium of *Euplotes focardii*, an Antarctic marine ciliate

Table 3
AdaptivePELE MD simulation statistics. The reported values (Å) are averaged over simulation replicates. The RMSF is averaged over backbone atoms of residues. The "lid" structure is highlighted in Fig. 3. RMSD: root mean square deviation; RMSF: root mean square fluctuation.

MD statistics	M-GH3_A in complex with cellobioside	M-GH3_A in complex with xylobioside	M-GH3_B in complex with cellobioside	M-GH3_B in complex with xylobioside
Heavy-atoms substrate RMSD to initial docked pose	1.9 ± 0.4	2.3 ± 0.7	1.7 ± 0.7	2.1 ± 0.4
Distance between catalytic aspartate Oγ and the attacked C_1 of the modeled substrate	3.3 ± 0.1	3.3 ± 0.2	4.7 ± 0.5	3.7 ± 0.3
Distance between catalytic acid/base $O_δ$ and the glycosidic O	4.3 ± 0.4	4.0 ± 0.5	4.0 ± 0.7	5.0 ± 0.4
Backbone RMSF	0.7 ± 0.1	0.7 ± 0.1	0.7 ± 0.1	0.7 ± 0.1
"Lid" loop backbone RMSF	1.6 ± 0.3	1.5 ± 0.1	2.0 ± 0.2	1.6 ± 0.2

[22,55]. In addition to the strict temperature requirements, it is known that the cold environment imposes adaptation to low nutrient availability [2]. Therefore, among hydrolytic enzymes, GHs play an important role in the degradation of environmentally poly- and oligosaccharides [12]. *Marinomonas* sp. ef1 has 34 genes coding for putative GHs classified in 19 different families. An insight into their heterogeneity is provided by the characterization of the M-GH1 and M-GH42 and the observation of their different thermal and properties and their substrate specificity towards β-galactosidic (M-GH1 and M-GH42) and β-glucosidic bonds (M-GH1) [56,57].

This study focused on GH3 enzymes identified in the genome of *Marinomonas* sp. ef1 and conserved in phylogenetically related *Marinomonas* spp. isolated from cold environments. GH3 is one of the largest families in the CAZy database and includes enzymes with β-glucosidase, β-xylosidase and *N*-acetylhexosaminidase activities. Usually these enzymes show *exo*-activity and act in synergy with *endo*-glucosidase (e.g. GH5, GH6 and GH7) and endo-xylanase (e.g. GH8, GH10, GH11 and GH30) in the degradation of polysaccharides [58]. Since GH3s have been frequently found in genomes and metagenomes isolated from hot environments [59–61], it can be supposed that this enzyme family plays a key role in polysaccharide degradation and adaptation to extreme environments.

Typically, cold-active enzymes are characterized by activity at low temperatures, low thermal stability, and undergo thermal inactivation before any significant change in their secondary structure; this behavior is evidenced by the so-called temperature gap (T_{GAP}), namely the difference between T_M and T_{opt} [5,62,63]. Our results indicate that M-GH3_A and M-GH3_B display contrasting thermal and catalytic properties. M-GH3_A is a *bona fide* cold-active enzyme with 20 % of activity at 10 °C, and a T_{GAP} of 16.8 °C. M-GH3_B while retaining 5 % of activity at 10 °C, exhibits mesophilic properties, i.e. high long-term thermal stability and temperature of inactivation coincident with that triggering the loss of secondary structure (T_{GAP}: - 0.5 °C). Overall, the biochemical features of these two enzymes combined with their evolutionary history suggest that M-GH3s have different phylogenetic origins and were probably acquired during the evolution of *Marinomonas* species by separate events of horizontal gene transfer and subsequently lost in some lineages.

In terms of substrate specificity, both M-GH3s are exo-acting enzymes, with M-GH3_A being a promiscuous β-glucosidase, and M-GH3_B a β-xylosidase with a narrow substrate specificity. The divergence in the substrate specificity of these two enzymes is probably due to the different shape of the catalytic chamber and its entrance, as suggested by our 3D models. More in detail, the catalytic chamber of M-GH3_A is predicted to be wider but has a narrower entrance than that of M-GH3_B, resulting in a negligible activity towards polysaccharides and a broad substrate specificity towards relatively small molecules such as cellobiose and cellotetraose. The narrower entrance of M-GH3_A is likely the result of the interaction between the C-terminal domain and the $(β/α)_8$ TIM barrel domain, which is enabled by the length and flexibility of the linker connecting the fibronectin-like and C-terminal domains. Structural and sequence analysis suggests that the size of the catalytic chamber and the C-terminal domain may serve as distinctive traits shaping the evolutionary trajectory of M-GH3. Intriguingly, the phylogenetic analysis of characterized GH3s reveals the existence of many distinct subfamilies, also grouped in at least four classes, based on substrate specificity towards β-glucans (clusters 1–7), xylans (clusters 8 and 11), and *N*-acetyl-β-D hexosaminides (cluster 14). Overall, our results indicate that substrate specificity within the GH3 family can be predicted by phylogenetic analysis. It should be noted that substrate specificity data are not available for all GH3s, and atypical activities, such as β-glucuronidase [64], may be underestimated. Although the correlation between the phylogenetically conserved residues that form the catalytic chamber and the interaction model between the enzyme and the substrate appears to be a promising tool for assessing the specificity of new GH3s and their classification, further structural studies are necessary to strengthen this approach.

M-GH3s lack a signal peptide for secretion, suggesting intracellular activity. They likely play a crucial role in the intracellular hydrolysis of oligosaccharides resulting from the degradation or breakdown of cellulose, xylan, and arabinoxylan by extracellular enzymes secreted by *Marinomonas* sp. ef1 or other bacteria belonging to the microbial consortium of *Euplotes focardii* [65,66]. Genome analysis revealed various operons that probably govern the catabolism of xylose and arabinose. Notably, canonical polysaccharide utilization loci [10,67] were found to be absent. Overall, our research indicates that *Marinomonas* sp. ef1 possesses a variety of GHs involved in the hydrolysis of glycosidic bonds, such as β-galactosidic, β-glucosidic and β-xylosidic sugar bonds [56,57].

In conclusion, this study presents a new method for annotating genes that may encode hydrolytic enzymes of the GH3 family. Additionally, it clarifies the physiological function of these enzymes in the adaptation of Antarctic bacteria, as demonstrated by *Marinomonas* sp. ef1.

CRediT authorship contribution statement

Alessandro Marchetti: Writing – review & editing, Writing – original draft, Visualization, Investigation, Data curation. **Marco Orlando:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Data curation. **Luca Bombardi:** Writing – review & editing, Visualization, Investigation, Data curation. **Salvatore Fusco:** Writing – review & editing, Supervision, Methodology, Data curation. **Marco Mangiagalli:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Data curation, Conceptualization. **Marina Lotti:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. Marco Mangiagalli reports financial support was provided by University

of Milano-Bicocca. Marina Lotti reports financial support was provided by University of Milano-Bicocca. Salvatore Fusco reports financial support was provided by University of Verona. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the University of Milano-Bicocca with FA (Fondo di Ateneo) to M.M. and M.L., PhD and postdoctoral fellowships from the University of Milano-Bicocca to A.M. and M.O., and MUR—Italian Ministry of University and Research—grant number "CUP B31I18000230006" and by Next Generation EU in the framework of National Biodiversity Future Center, grant number "CUP B33C22000660001" to S.F. and L.B. The authors also thank Cristian Pietro Maierna for his contribution to protein purification and biochemical experiments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2024.133449.

References

[1] N. Merino, H.S. Aronson, D.P. Bojanova, J. Feyhl-Buska, M.L. Wong, S. Zhang, D. Giovannelli, Living at the extremes: extremophiles and the limits of life in a planetary context, Front. Microbiol. 10 (2019) 780, https://doi.org/10.3389/fmicb.2019.00780.

[2] C. Lauritano, C. Rizzo, A. Lo Giudice, M. Saggiomo, Physiological and molecular responses to main environmental stressors of microalgae and bacteria in polar marine environments, Microorganisms 8 (2020) 1957, https://doi.org/10.3390/microorganisms8121957.

[3] S. Mocali, C. Chiellini, A. Fabiani, S. Decuzzi, D. de Pascale, E. Parrilli, M.L. Tutino, E. Perrin, E. Bosi, M. Fondi, A. Lo Giudice, R. Fani, Ecology of cold environments: new insights of bacterial metabolic adaptation through an integrated genomic-phenomic approach, Sci. Rep. 7 (2017) 839, https://doi.org/10.1038/s41598-017-00876-4.

[4] G. Feller, Protein stability and enzyme activity at extreme biological temperatures, J. Phys. Condens. Matter 22 (2010) 323101, https://doi.org/10.1088/0953-8984/22/32/323101.

[5] T. Collins, G. Feller, Psychrophilic enzymes: strategies for cold-adaptation, Essays Biochem. 67 (2023) 701–713, https://doi.org/10.1042/EBC20220193.

[6] M. Santiago, C.A. Ramírez-Sarmiento, R.A. Zamora, L.P. Parra, Discovery, molecular mechanisms, and industrial applications of cold-active enzymes, Front. Microbiol. 7 (2016), https://doi.org/10.3389/fmicb.2016.01408.

[7] M. Fabiano, R. Danovaro, Enzymatic activity, bacterial distribution, and organic matter composition in sediments of the Ross Sea (Antarctica), Appl. Environ. Microbiol. 64 (1998) 3838–3845, https://doi.org/10.1128/AEM.64.10.3838-3845.1998.

[8] C. Rizzo, R. Malavenda, B. Gerçe, M. Papale, C. Syldatk, R. Hausmann, V. Bruni, L. Michaud, A. Lo Giudice, S. Amalfitano, Effects of a simulated acute oil spillage on bacterial communities from Arctic and Antarctic marine sediments, Microorganisms 7 (2019) 632, https://doi.org/10.3390/microorganisms7120632.

[9] S. Emil Ruff, D. Probandt, A.-C. Zinkann, M.H. Iversen, C. Klaas, L. Würzberg, N. Krombholz, D. Wolf-Gladrow, R. Amann, K. Knittel, Indications for algae-degrading benthic microbial communities in deep-sea sediments along the Antarctic Polar Front, Deep-Sea Res. II Top. Stud. Oceanogr. 108 (2014) 6–16, https://doi.org/10.1016/j.dsr2.2014.05.011.

[10] T. Dutschei, I. Beidler, D. Bartosik, J. Seeßelberg, M. Teune, M. Bäumgen, S. Q. Ferreira, J. Heldmann, F. Nagel, J. Krull, L. Berndt, K. Methling, M. Hein, D. Becher, P. Langer, M. Delcea, M. Lalk, M. Lammers, M. Höhne, J. Hehemann, T. Schweder, U.T. Bornscheuer, Marine *Bacteroidetes* enzymatically digest xylans from terrestrial plants, Environ. Microbiol. 25 (2023) 1713–1727, https://doi.org/10.1111/1462-2920.16390.

[11] A.A. Salmeán, W.G.T. Willats, S. Ribeiro, T.J. Andersen, M. Ellegaard, Over 100-year preservation and temporal fluctuations of cell wall polysaccharides in marine sediments, Front. Plant Sci. 13 (2022) 785902, https://doi.org/10.3389/fpls.2022.785902.

[12] M. Bäumgen, T. Dutschei, U.T. Bornscheuer, Marine polysaccharides: occurrence, enzymatic degradation and utilization, ChemBioChem 22 (2021) 2247–2256, https://doi.org/10.1002/cbic.202100078.

[13] D. Li, Z. Zhang, P. Zhang, J. Wang, Y. Liu, Y. Li, Deterministic assembly processes shaping habitat-specific glycoside hydrolase composition, Glob. Ecol. Biogeogr. 33 (2024) 189–202, https://doi.org/10.1111/geb.13768.

[14] E. Drula, M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat, N. Terrapon, The carbohydrate-active enzyme database: functions and literature, Nucleic Acids Res. 50 (2022) D571–D577, https://doi.org/10.1093/nar/gkab1045.

[15] D. Faure, The family-3 glycoside hydrolases: from housekeeping functions to host-microbe interactions, Appl. Environ. Microbiol. 68 (2002) 1485–1490, https://doi.org/10.1128/AEM.68.4.1485-1490.2002.

[16] D. Dodd, S. Kiyonari, R.I. Mackie, I.K.O. Cann, Functional diversity of four glycoside hydrolase family 3 enzymes from the rumen bacterium *Prevotella bryantii* $B_1$4, J. Bacteriol. 192 (2010) 2335–2345, https://doi.org/10.1128/JB.01654-09.

[17] W.-Y. Jeng, N.-C. Wang, M.-H. Lin, C.-T. Lin, Y.-C. Liaw, W.-J. Chang, C.-I. Liu, P.-H. Liang, A.H.-J. Wang, Structural and functional analysis of three β-glucosidases from bacterium Clostridium cellulovorans, fungus Trichoderma reesei and termite Neotermes koshunensis, J. Struct. Biol. 173 (2011) 46–56, https://doi.org/10.1016/j.jsb.2010.07.008.

[18] J.N. Varghese, M. Hrmova, G.B. Fincher, Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase, Structure 7 (1999) 179–190, https://doi.org/10.1016/S0969-2126(99)80024-0.

[19] R.N. Florindo, V.P. Souza, L.R. Manzine, C.M. Camilo, S.R. Marana, I. Polikarpov, A.S. Nascimento, Structural and biochemical characterization of a GH3 β-glucosidase from the probiotic bacteria Bifidobacterium adolescentis, Biochimie 148 (2018) 107–115, https://doi.org/10.1016/j.biochi.2018.03.007.

[20] T. Pozzo, J.L. Pasten, E.N. Karlsson, D.T. Logan, Structural and functional analyses of β-glucosidase 3B from Thermotoga neapolitana: a thermostable three-domain representative of glycoside hydrolase 3, J. Mol. Biol. 397 (2010) 724–739, https://doi.org/10.1016/j.jmb.2010.01.072.

[21] B. Deflandre, C. Jadot, S. Planckaert, N. Thiébaut, N. Stulanovic, R. Herman, B. Devreese, F. Kerff, S. Rigali, Structure and function of BcpE2, the most promiscuous GH3-family glucose scavenging beta-glucosidase, MBio 13 (2022) e00935-22, https://doi.org/10.1128/mbio.00935-22.

[22] M. Zannotti, K.P. Ramasamy, V. Loggi, A. Vassallo, S. Pucciarelli, R. Giovannetti, Hydrocarbon degradation strategy and pyoverdine production using the salt tolerant Antarctic bacterium *Marinomonas* sp. ef1., RSC Adv. 13 (2023) 19276–19285. doi:https://doi.org/10.1039/D3RA02536E.

[23] M.S. John, J.A. Nagoth, M. Zannotti, R. Giovannetti, A. Mancini, K.P. Ramasamy, C. Miceli, S. Pucciarelli, Biogenic synthesis of copper nanoparticles using bacterial strains isolated from an Antarctic consortium associated to a psychrophilic marine ciliate: characterization and potential application as antimicrobial agents, Mar. Drugs 19 (2021) 263, https://doi.org/10.3390/md19050263.

[24] M. Wistrand, E.L. Sonnhammer, Improved profile HMM performance by assessment of critical algorithmic features in SAM and HMMER, BMC Bioinformatics 6 (2005) 99, https://doi.org/10.1186/1471-2105-6-99.

[25] H. Zhang, T. Yohe, L. Huang, S. Entwistle, P. Wu, Z. Yang, P.K. Busk, Y. Xu, Y. Yin, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation, Nucleic Acids Res. 46 (2018) W95–W101, https://doi.org/10.1093/nar/gky418.

[26] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: J. M. Walker (Ed.), The Proteomics Protocols Handbook, Humana Press, Totowa, NJ, 2005, pp. 571–607, https://doi.org/10.1385/1-59259-890-0:571.

[27] A.J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees, BMC Evol. Biol. 7 (2007) 214, https://doi.org/10.1186/1471-2148-7-214.

[28] A. Rambaut, A.J. Drummond, D. Xie, G. Baele, M.A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7, Syst. Biol. 67 (2018) 901–904, https://doi.org/10.1093/sysbio/syy032.

[29] K. Katoh, H. Toh, Parallelization of the MAFFT multiple sequence alignment program, Bioinformatics 26 (2010) 1899–1900, https://doi.org/10.1093/bioinformatics/btq224.

[30] J. Rozewicki, S. Li, K.M. Amada, D.M. Standley, K. Katoh, MAFFT-DASH: integrated protein sequence and structural alignment, Nucleic Acids Res. (2019) gkz342, https://doi.org/10.1093/nar/gkz342.

[31] B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, Mol. Biol. Evol. 37 (2020) 1530–1534, https://doi.org/10.1093/molbev/msaa015.

[32] C.C. Dang, B.Q. Minh, H. McShea, J. Masel, J.E. James, L.S. Vinh, R. Lanfear, nQMaker: estimating time nonreversible amino acid substitution models, Syst. Biol. 71 (2022) 1110–1123, https://doi.org/10.1093/sysbio/syac007.

[33] D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation, Mol. Biol. Evol. 35 (2018) 518–522, https://doi.org/10.1093/molbev/msx281.

[34] F. Lemoine, J.-B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, T. De Oliveira, O. Gascuel, Renewing Felsenstein's phylogenetic bootstrap in the era of big data, Nature 556 (2018) 452–456, https://doi.org/10.1038/s41586-018-0043-0.

[35] J. Zimmermann, C. Kaleta, S. Waschina, gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models, Genome Biol. 22 (2021) 81, https://doi.org/10.1186/s13059-021-02295-1.

[36] B. Taboada, K. Estrada, R. Ciria, E. Merino, Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes, Bioinformatics 34 (2018) 4118–4120, https://doi.org/10.1093/bioinformatics/bty496.

[37] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger,

M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.
[38] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all, Nat. Methods 19 (2022) 679–682, https://doi.org/10.1038/s41592-022-01488-1.
[39] Y. Zhang, Z. Zhang, B. Zhong, S. Misra, J. Tang, DiffPack: A Torsional Diffusion Model for Autoregressive Protein Side-chain Packing, 2023, https://doi.org/10.48550/ARXIV.2306.01794.
[40] P. Eastman, J. Swails, J.D. Chodera, R.T. McGibbon, Y. Zhao, K.A. Beauchamp, L.-P. Wang, A.C. Simmonett, M.P. Harrigan, C.D. Stern, R.P. Wiewiora, B.R. Brooks, V.S. Pande, OpenMM 7: rapid development of high performance algorithms for molecular dynamics, PLoS Comput. Biol. 13 (2017) e1005659, https://doi.org/10.1371/journal.pcbi.1005659.
[41] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, J. Chem. Theory Comput. 11 (2015) 3696–3713, https://doi.org/10.1021/acs.jctc.5b00255.
[42] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, J. Chemother. 4 (2012) 17, https://doi.org/10.1186/1758-2946-4-17.
[43] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174, https://doi.org/10.1002/jcc.20035.
[44] A. Jakalian, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, J. Comput. Chem. 23 (2002) 1623–1641, https://doi.org/10.1002/jcc.10128.
[45] A.T. McNutt, P. Francoeur, R. Aggarwal, T. Masuda, R. Meli, M. Ragoza, J. Sunseri, D.R. Koes, GNINA 1.0: molecular docking with deep learning, J. Chemother. 13 (2021) 43, https://doi.org/10.1186/s13321-021-00522-2.
[46] D. Lecina, J.F. Gilabert, V. Guallar, Adaptive simulations, towards interactive protein-ligand modeling, Sci. Rep. 7 (2017) 8466, https://doi.org/10.1038/s41598-017-08445-5.
[47] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, Nucleic Acids Res. 35 (2007) W522–W525, https://doi.org/10.1093/nar/gkm276.
[48] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, *MMPBSA.py*: an efficient program for end-state free energy calculations, J. Chem. Theory Comput. 8 (2012) 3314–3321, https://doi.org/10.1021/ct300418h.
[49] D. Ubbiali, M. Orlando, M. Kovačič, C. Iacobucci, M.S. Semrau, G. Bajc, S. Fortuna, G. Ilc, B. Medagli, S. Oloketuyi, P. Storici, A. Sinz, R. Grandori, A. de Marco, An anti-HER2 nanobody binds to its antigen HER2 via two independent paratopes, Int. J. Biol. Macromol. 182 (2021) 502–511, https://doi.org/10.1016/j.ijbiomac.2021.04.032.
[50] F.W. Studier, Protein production by auto-induction in high-density shaking cultures, Protein Expr. Purif. 41 (2005) 207–234, https://doi.org/10.1016/j.pep.2005.01.016.
[51] D. Touchette, I. Altshuler, C. Gostinčar, P. Zalar, I. Raymond-Bouchard, J. Zajc, C. P. McKay, N. Gunde-Cimerman, L.G. Whyte, Novel Antarctic yeast adapts to cold by switching energy metabolism and increasing small RNA synthesis, ISME J. 16 (2022) 221–232, https://doi.org/10.1038/s41396-021-01030-9.
[52] N.L. Mohamad, S.M. Mustapa Kamal, M.N. Mokhtar, Xylitol biological production: a review of recent studies, Food Rev. Int. 31 (2015) 74–89, https://doi.org/10.1080/87559129.2014.961077.
[53] T. Yanai, Y. Takahashi, E. Katsumura, N. Sakai, K. Takeshita, R. Imaizumi, H. Matsuura, S. Hongo, T. Waki, S. Takahashi, M. Yamamoto, K. Kataoka, T. Nakayama, S. Yamashita, Structural insights into a bacterial β-glucosidase capable of degrading sesaminol triglucoside to produce sesaminol: toward the understanding of the aglycone recognition mechanism by the C-terminal lid domain, The Journal of Biochemistry 174 (2023) 335–344, https://doi.org/10.1093/jb/mvad048.
[54] A. Gazizov, A. Lian, C. Goverde, S. Ovchinnikov, N.F. Polizzi, AF2BIND: Predicting Ligand-binding Sites Using the Pair Representation of AlphaFold2, 2023, https://doi.org/10.1101/2023.10.15.562410.
[55] M.S. John, J.A. Nagoth, K.P. Ramasamy, P. Ballarini, M. Mozzicafreddo, A. Mancini, A. Telatin, P. Liò, G. Giuli, A. Natalello, C. Miceli, S. Pucciarelli, Horizontal gene transfer and silver nanoparticles production in a new Marinomonas strain isolated from the Antarctic psychrophilic ciliate Euplotes focardii, Sci. Rep. 10 (2020) 10218, https://doi.org/10.1038/s41598-020-66878-x.
[56] M. Mangiagalli, M. Lapi, S. Maione, M. Orlando, S. Brocca, A. Pesce, A. Barbiroli, C. Camilloni, S. Pucciarelli, M. Lotti, M. Nardini, The co-existence of cold activity and thermal stability in an Antarctic GH42 β-galactosidase relies on its hexameric quaternary arrangement, FEBS J. 288 (2021) 546–565, https://doi.org/10.1111/febs.15354.
[57] L.J. Gourlay, M. Mangiagalli, E. Moroni, M. Lotti, M. Nardini, Structural determinants of cold activity and glucose tolerance of a family 1 glycoside hydrolase (GH1) from Antarctic *Marinomonas* sp. ef1, FEBS J. (2024) febs.17096, https://doi.org/10.1111/febs.17096.
[58] Y. Miao, P. Li, G. Li, D. Liu, I.S. Druzhinina, C.P. Kubicek, Q. Shen, R. Zhang, Two degradation strategies for overcoming the recalcitrance of natural lignocellulosic xylan by polysaccharides-binding GH 10 and GH 11 xylanases of filamentous fungi, Environ. Microbiol. 19 (2017) 1054–1064, https://doi.org/10.1111/1462-2920.13614.
[59] N.J. Reichart, R.M. Bowers, T. Woyke, R. Hatzenpichler, High potential for biomass-degrading enzymes revealed by hot spring metagenomics, Front. Microbiol. 12 (2021) 668238, https://doi.org/10.3389/fmicb.2021.668238.
[60] R. Ameri, J.L. García, A.B. Derenfed, N. Pradel, S. Neifar, S. Mhiri, M. Mezghanni, N.Z. Jaouadi, J. Barriuso, S. Bejar, Genome sequence and Carbohydrate Active Enzymes (CAZymes) repertoire of the thermophilic Caldicoprobacter algeriensis TH7C1T, Microb. Cell Factories 21 (2022) 91, https://doi.org/10.1186/s12934-022-01818-0.
[61] A. Strazzulli, B. Cobucci-Ponzano, R. Iacono, R. Giglio, L. Maurelli, N. Curci, C. Schiano-di-Cola, A. Santangelo, P. Contursi, V. Lombard, B. Henrissat, F. M. Lauro, C.M.G.A. Fontes, M. Moracci, Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments, FEBS J. 287 (2020) 1116–1137, https://doi.org/10.1111/febs.15080.
[62] S. Gault, P.M. Higgins, C.S. Cockell, K. Gillies, A meta-analysis of the activity, stability, and mutational characteristics of temperature-adapted enzymes, Biosci. Rep. 41 (2021) BSR20210336, https://doi.org/10.1042/BSR20210336.
[63] M. Mangiagalli, M. Lotti, Cold-active β-galactosidases: insight into cold adaptation mechanisms and biotechnological exploitation, Mar. Drugs 19 (2021) 43, https://doi.org/10.3390/md19010043.
[64] S. Neun, P. Brear, E. Campbell, T. Tryfona, K. El Omari, A. Wagner, P. Dupree, M. Hyvönen, F. Hollfelder, Functional metagenomic screening identifies an unexpected β-glucuronidase, Nat. Chem. Biol. 18 (2022) 1096–1103, https://doi.org/10.1038/s41589-022-01071-x.
[65] S. Pucciarelli, R.R. Devaraj, A. Mancini, P. Ballarini, M. Castelli, M. Schrallhammer, G. Petroni, C. Miceli, Microbial consortium associated with the Antarctic marine ciliate Euplotes focardii: an investigation from genomic sequences, Microb. Ecol. 70 (2015) 484–497, https://doi.org/10.1007/s00248-015-0568-9.
[66] M.S. John, J.A. Nagoth, K.P. Ramasamy, A. Mancini, G. Giuli, A. Natalello, P. Ballarini, C. Miceli, S. Pucciarelli, Synthesis of bioactive silver nanoparticles by a Pseudomonas strain associated with the Antarctic psychrophilic protozoon Euplotes focardii, Mar. Drugs 18 (2020) 38, https://doi.org/10.3390/md18010038.
[67] L. Reisky, A. Préchoux, M.-K. Zühlke, M. Bäumgen, C.S. Robb, N. Gerlach, T. Roret, C. Stanetty, R. Larocque, G. Michel, T. Song, S. Markert, F. Unfried, M. D. Mihovilovic, A. Trautwein-Schult, D. Becher, T. Schweder, U.T. Bornscheuer, J.-H. Hehemann, A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan, Nat. Chem. Biol. 15 (2019) 803–812, https://doi.org/10.1038/s41589-019-0311-9.