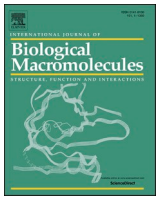
International Journal of Biological Macromolecules 299 (2025) 140113

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

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



Polysaccharide degradation in an Antarctic bacterium: Discovery of glycoside hydrolases from remote regions of the sequence space

Marco Orlando [a], Alessandro Marchetti [a], Luca Bombardi [b], Marina Lotti [a], Salvatore Fusco [b,*], Marco Mangiagalli [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:
GH5
GH50
Computational enzyme modelling
Cold-active enzyme

ABSTRACT

Glycoside hydrolases (GHs) are enzymes involved in the degradation of oligosaccharides and polysaccharides. The sequence space of GHs is rapidly expanding due to the increasing number of available sequences. This expansion paves the way for the discovery of novel enzymes with peculiar structural and functional properties. This work is focused on two GHs, Ps_GH5 and Ps_GH50, from the genome of the Antarctic bacterium *Pseudomonas* sp. ef1. These enzymes are in an unexplored region of the sequence space of their respective GH families, not allowing a reliable sequence-based function prediction. For this reason, a computational pipeline was developed that combines deep learning "dynamic docking" on AlphaFold 3D models with physics-based molecular dynamics simulations to infer their substrate specificity. From *in silico* screening of a repertoire of potential oligosaccharides, only xylooligosaccharides for Ps_GH5 and galactooligosaccharides for Ps_GH50 emerged as catalytically competent substrates. Biochemical characterization agrees with computational simulations indicating that Ps_GH5 is an *endo*-β-xylanase, and Ps_GH50 is active mainly on small galactooligosaccharides. In conclusion, this study identifies two novel GHs subfamilies placed in remote regions of the sequence space and highlights the efficacy of substrate specificity prediction by computational approaches in the discovery of new enzymes.

1. Introduction

Glycoside hydrolases (GHs) catalyze the hydrolysis of glycosidic bonds, playing a crucial role in the uptake of carbon sources represented by oligo- and polysaccharides [1]. Currently, GHs are grouped in 189 families cataloged in the CAZy database (http://www.cazy.org/). The classification system underlying the CAZy database defines families and subfamilies by phylogenetic analysis and the presence of sequences clustered around biochemically characterized members [1,2]. The biochemical data are usually required for the definition of new GH families and subfamilies [1,3].

The number of available GH sequences has increased dramatically due to rapid advances in DNA sequencing technologies and targeted experimental screening [4], providing new opportunities for the discovery of enzymes with novel functions. The exponential increase of protein sequences created a significant discrepancy between the number of uncharacterized entries that have been automatically annotated in CAZy and the number of biochemically characterized enzymes [5]. The main consequence of this discrepancy is the existence of large regions of sequence space that are currently uncharacterized [4,6]. The accuracy of classification and prediction of biochemical properties is therefore challenged when new sequences fall into these remote regions. In this context, sequence/function similarity network (SSN) and phylogenetic approaches may have a limited reliability of functional inference and are unable to infer hitherto undiscovered substrate specificities [7]. Therefore, the biochemical characterization of GH sequences belonging to remote regions of the sequence space is essential to improve the success of functional prediction and classification accuracy [6].

In this frame, GHs from psychrophilic marine microorganisms may serve as a source of enzymes from so far unexplored regions of the sequence space [8,9]. These organisms have evolved to survive under extremely stressful conditions, including low temperatures, high osmotic stress, and low nutrient availability [10,11]. Indeed, psychrophilic marine sediments are characterized by the presence of low

* Corresponding authors.
E-mail addresses: salvatore.fusco@univr.it (S. Fusco), marco.mangiagalli@unimib.it (M. Mangiagalli).

https://doi.org/10.1016/j.ijbiomac.2025.140113
Received 22 October 2024; Received in revised form 13 January 2025; Accepted 18 January 2025
Available online 20 January 2025

amounts of polysaccharides derived from terrestrial plants, such as cellulose and xylan, and/or from marine algae, such as laminarin, alginate, agar and carrageenan [12]. Psychrophilic marine microbes have been reported to secrete extracellular GHs to hydrolyze complex polysaccharides such as cellulases (e.g. GH5, GH6, GH9, GH12), xylanases (e.g. GH8, GH10, GH11 and GH30), agarases (e.g. GH16, GH50 and GH117) and carrageenases (e.g. GH16) [13–15]. However, available information might be still largely incomplete calling for the discovery and classification of additional enzymes and enzyme families.

In this study, a custom workflow that integrated functional- and sequence-based approaches was developed to discover new GHs involved in polysaccharide degradation in *Pseudomonas* sp. ef1 [16]. This bacterium, isolated from an Antarctic microbial consortium, is a source of unusual GHs previously identified from its genome, such as two evolutionary "orphan" GH19s with rare lysozyme activity [17,18]. The enzymes described in this study (Ps_GH5 and Ps_GH50) are situated in unexplored regions of GH5 and GH50 families sequence space. The biochemical characterization reveals that Ps_GH5 is an *endo*-xylanase, while Ps_GH50 is an *exo*-enzyme with low agarolytic activity and high activity on small galactooligosaccharides (GOS). Our *in silico* and functional analyses indicated that these enzymes have uncommon structural properties and may belong to new subfamilies.

2. Materials and methods

2.1. Materials

Glucose, galactose, lactose, xylose, ampicillin, tetracycline, κ-carrageenan, 2,5 dinitrosalicylic (DNS), low-melting point agarose and para-nitrophenyl derivatives were purchased from Merck (Merck, Darmstadt, Germany). Xylan (code: P-XYLNBE), arabinoxylan (code: P-WAXYL), xyloglucan (code: P-XYGLN), mannan (code: P-MANIV) and galactomannan (code: P-GGMMV) were purchased from Megazyme International (Megazyme International, Bray, Ireland). Carboxymethylcellulose (CMC, code: 22525.296) and lactose-galactose trisaccharide (GOS 3X, code: OG32134) were purchased from VWR International (Radnord, USA) and Biosynth Carbosynth (Staad, Switzerland), respectively.

2.2. Biolog Omnilog® screening of the metabolic activity on carbohydrates

Pseudomonas sp. ef1 was grown in Luria-Bertani (LB) broth at 22 °C as described in [12] until an $O.D._{600} \approx 1$ was reached. The cell pellet was harvested by centrifugation at 4000*g* for 10 min at 4 °C, washed twice with physiological solution (0.9 % NaCl), and then resuspended at a final $O.D._{600}$ of 0.4 in IF-0 GN/GP base solution (Biolog, Newark, DE, USA). The cell suspension was supplemented with the Biolog redox mix dye E (1 % *v*/v) and different sterilized solutions of sugars at 0.2 % *w/v* final concentrations. Sugar solutions were prepared by dissolving glucose, galactose, xylose, CMC, xylan, xyloglucan and low melting point agarose, in MilliQ water at the concentration of 1 % (*w/v*). Solutions containing IF-0 GN/GP base, Biolog redox mix dye E and different sugars were used as blanks. The cell suspension supplemented with Biolog redox mix dye E without carbon source was used as a negative control. Biolog 96-well plates containing either 100 μL of cell suspensions or blanks were incubated in the OmniLog® incubator at 25 °C for 48 h. The digital camera was configured to take images every 15 min. Omnilog data on physiological responses, which consists of measures of redox chemistry activity due to cell respiration, was subtracted from the blank contribution, and were analyzed using the OmniLog Software. The activity index (normalized between 0 and 3), a parameter that allows the ranking and comparison of each respiratory curve, was calculated using the *dphenome* function of the DuctApe package (http://combogenomics.github.io/DuctApe/, [19]). Experiments were performed in duplicate and reported as a mean ± standard deviation.

2.3. Bioinformatics analyses

2.3.1. Annotation of Pseudomonas sp. ef1 GH sequences

The sequences of putative GHs were annotated from *Pseudomonas* sp. ef1 genome (GenBank RefSeq Assembly ID: GCF_007293365.1) with *hhmscan* from HHMER v3.3.2 [20], using the family/subfamily profile hidden Markov models from dbCAN2 [21], by applying a restrictive *e*-value of e^{-30}. The predicted molecular weight was determined with Expasy ProtParam [22]. Signal peptides were predicted with SignalP 6.0 [23]. Annotated GHs were scanned for extracellular enzymes putatively involved in the degradation of glucan, xylan, arabinoxylan, xyloglucan and agarose (families 1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 16, 26, 30, 39, 43, 45, 48, 50, 51, 86, 118). Two sequences with a predicted signal peptide were annotated as GH5 (Ps_GH5, UniProt ID: A0A554AFN4) and GH50 (Ps_GH50, UniProt ID: A0A554AKC2) and further analyzed.

2.3.2. Sequence space and evolutionary analysis of selected GHs

The sequence spaces of the selected GHs were created with the modified scripts from protein cartography (https://github.com/Arcadia-Science/ProteinCartography, [24]), a workflow that allows the creation of structure-based maps of protein families.

The sequences of Ps_GH5 and Ps_GH50 were used as a query to search homologous proteins in the ColabFold database (https://colabfold.mmseqs.com/, accessed on 21/11/2023). These sequences were then filtered by using the HHfilter function of the hh-suite package v. 3.3.0 [25] to select proteins with a length between 200 and 1000 amino acids, a minimum of 80 % sequence coverage and 10 % global identity to the query sequences. Maximum 80 % pairwise global sequence identity was allowed between the collected sequences. The final list of protein sequences was created by selecting only those for which a predicted structure was available in the PDB (https://www.rcsb.org/) or in the AlphaFold database (AFDB, https://alphafold.ebi.ac.uk/). As both GH5 and GH50 belong to Clan A, characterized sequences from Clan A with a global sequence identity of >10 % to the query sequences were retrieved from CAZy (http://www.cazy.org, accessed on 21/11/2023) and added to the previous list of filtered sequences. For characterized sequences not available in the AFDB, the AlphaFold (AF) 3D structure was predicted as a monomer by Colabfold 1.5.5 (https://github.com/sokrypton/ColabFold, [26]), using the AlphaFold2_ptm model [27].

Structures were aligned and compared with Foldseek release 8 [28]. Max 2000 sequences/structures were used to create the sequence spaces of each query sequence. Default *protein cartography* settings were chosen for the rest of parameters. The final 2D representation of the protein space was obtained from UMAP analysis of the all-vs-all pair structural similarity matrix. UMAP was chosen as it is often better at preserving global data structure in the final projection. The sequences were plotted as circles and annotated if characterized from known GH families.

For phylogenetic analysis, only the sequences of biochemically characterized GHs were used. The substrate specificity of each selected sequence was manually retrieved from the literature. To remove redundancy, sequences were clustered at 95 % sequence identity threshold with cd-hit 4.8.1 (https://github.com/weizhongli/cdhit, [29]). Only the 3D models exhibiting a high degree of confidence (average plDDT >0.75) were subjected to structural alignment with mTM-align (http://yanglab.nankai.edu.cn/mTM-align, [30]). The resulting structure-based multiple sequence alignment was trimmed to retain only the catalytic domains. Positions with a gap shared by >75 % of sequences were removed. A rooted Maximum Likelihood tree with IQ-Tree v.2.2.2.7 [31] was estimated using this alignment. 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].

2.3.3. In silico prediction of enzyme substrate specificity

The substrate specificity of Ps_GH5 and Ps_GH50 was predicted using a structure-based approach. The 3D models of Ps_GH5 and Ps_GH50 were predicted using AF2 and ColabFold v.1.5.5 (https://github.com/sokrypton/ColabFold) to be monomers. The 3D model of Ps_GH5 was then docked with cello-oligosaccharides (COS: cellobiose, cellotriose, cellotetraose), xylo-oligosaccharides (XOS: xylobiose, xylotriose, xylotetraose), manno-oligosaccharides (MOS: mannobiose, mannotriose, mannotetraose), arabino-oligosaccharides (AOS: arabinobiose, arabinotriose, arabinotetraose), and arabinoxylo-oligosaccharides (AXOS: D-xylose α-1,2 L-arabinofuranoside, D-xylose α-1,3 L-arabinofuranoside), whereas the PS_GH50 model was docked to neoagaro-oligosaccharides (NAOS: neoagarobiose, neoagarotetraose, neoagarohexaose), galacto-oligosaccharides (GOS: lactose, lactose-galactose trisaccharide, lactose-galactose tetrasaccharide). These substrates were generated with glycam (https://glycam.org/cb/) in fully hydrogenated states, manually checked with Avogadro v.1.2.0 [35] and converted into SMILES format with Open Babel v.3.1.0 [36]. The modeled substrates were chosen to be representative of the substrate specificity that is typically observed in evolutionary-related characterized GHs. Docking simulations were performed with DynamicBind v1.0 [37], a deep learning method that allows a "dynamic docking" approach in which both protein and ligand are flexible. The v2 model (https://zenodo.org/records/10183369) was used. The default parameters of DynamicBind v1.0 were used for both Ps_GH5 and Ps_GH50. The only exception was the number of denoised samples, which was increased to 40. Only PoseBuster-validated protein-substrate docking poses were considered [38]. The poses were evaluated by combining the DynamicBind confidence score and two Euclidean distances: between the catalytic nucleophile Oε and the attacked C_1 (distance C—O) and between the catalytic acid/base Oε and the glycosidic O (distance O—O). For each oligosaccharide class, the pose with the highest DynamicBind score and both C—O and O—O distances <4 Å was considered catalytically competent and subjected to energy minimization and refinement through molecular dynamics (MD) simulations in an NPT ensemble. The MD simulation protocol was derived from the Colab Notebook at https://colab.research.google.com/github/pablo-arantes/making-it-rain/blob/main/Protein_ligand.ipynb [39]. Open MM v.7.7 [40] was used as the MD engine. The changes to default parameters were adding a 10 Å padding for the box size, the addition of ions only up to neutralization of the system, a 2 ns NVT equilibration step at 298 K using the Andersen thermostat with the Verlet integrator while position restraining (700 kJ/mol) the heavy atoms of the enzyme and the substrate. 2 ns NPT equilibration while keeping the same position restraining in NVT step. 20 ns of unbiased production MD simulations were run in triplicate, saving the system coordinates every 200 ps, starting from independent equilibration steps. Unbiased MD frames (300 in total) were merged and used to calculate MD simulation statistics representative of enzyme catalysis and to estimate a binding energy using DeepQM [41], a deep learning adaptation of an end-state binding free energy approximation method (linear interaction energy, LIE). The latter is obtained by replacing the standard molecular mechanics potentials (MMPs) with neural network potentials, and by rescaling the predicted energy with a linear fit to the absolute binding energy data used for training [41]. This choice was made because the DeepQM model has been shown to be more accurate than other end-state methods that rely on force field-based MMPs. By jointly evaluating the MD simulation statistics and DeepQM binding affinity, the best enzyme-substrate complexes were selected.

2.4. Production and purification of the recombinant enzymes

Sequences coding for Ps_GH5 and Ps_GH50, were optimized for expression in *Escherichia coli* cells, chemically synthesized (Genscript, Piscataway, NJ, USA) and cloned in frame with a C-terminal 6X His-Tag into pET22 plasmid (EMD, Millipore, Billerica, MA, USA) between *Msc*I and *Xho*I sites. *E. coli* OrigamiB (DE3) cells and *E. coli* BL21 (DE3) cells (EMD, Millipore, Billerica, MA, USA) were used as hosts for the heterologous expression of Ps_GH5 and of Ps_GH50, respectively.

Recombinant enzymes were produced in Zym 5052 medium [42], added of 100 mg/L of ampicillin (Merck, Darmstadt, Germany) for *E. coli* BL21 (DE3) and 100 mg/L tetracycline and ampicillin (Merck, Darmstadt, Germany) for *E. coli* OrigamiB (DE3) cultures. The cultures were incubated at 25 °C for 24 h and then harvested by centrifugation at 4000 x*g* for 10 min at 4 °C. Recombinant enzymes were extracted and purified as described in [12]. Purified Ps_GH5 and Ps_GH50 were suspended in 100 mM sodium phosphate buffer (PB), pH 7. Protein concentration was determined by the Bradford protein assay (Bio-Rad, California, USA) using bovine serum albumin as a standard.

2.5. Quaternary structure

The quaternary structure 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, as described in [12].

2.6. Determination of functional properties

Substrate specificity was assessed using the following substrates: *para*-nitrophenyl β-D-glucopyranoside (pNPG), *ortho*-nitrophenyl β-D-galactopyranoside (oNPGal), *para*-nitrophenyl β-D-xylopyranoside (pNPXyl), *para*-nitrophenyl β-D-mannopyranoside (pNPMan), *para*-nitrophenyl β-D-fucopyranoside (pNPFuc), *para*-nitrophenyl β-D-cellobioside (pNPClb), 4-Nitrophenyl-β-D-glucuronide (pNPGlcA), 4-Nitrophenyl-α-L-arabinofuranoside (pNPAra), xylan, arabinoxylan, CMC, mannan, galactomannan, κ-carrageenan and GOS 3X. Reactions were performed in 100 µL volume at 35 °C in PB with 50 µg of enzyme and 10 mM of nitrophenyl glycoside derivatives or 1 % *w/v* of polysaccharides. The reactions with nitrophenyl glycoside derivatives were stopped, after 15 min, by the addition of 100 µL of 1 M Na_2CO_3 pH 11. The absorbance was measured at either 420 nm for oNPG (ε_{oNP}: 4.6 $mM^{-1} \cdot cm^{-1}$) or at 405 nm for the other substrates (ε_{pNP}: 18.6 $mM^{-1} \cdot cm^{-1}$) with a VICTOR™ X Multilabel Plate Reader (PerkinElmer Inc., USA).

Activity on polysaccharides was determined by the DNS assay. Enzymatic reactions were conducted in a solution 1 % *w/v* of the polysaccharide (in PB) and a final concentration of 0.5 mg/mL of each purified enzyme, at 35 °C, at 800 rpm in a thermal shaker (Eppendorf, Hamburg, Germany). After 1 h of incubation, 400 µL of DNS reagent was added and the reaction was heated at 99 °C for 5 min to inactivate the enzyme and develop the assay [43]. Then 150 µL of Rochelle salt (40 % *w/v*) was added to all reactions to stabilize the final colour. Absorbance was measured at 540 nm using a VICTOR™ X Multilabel Plate Reader (PerkinElmer Inc., USA). The calibration curves were constructed using known concentrations of pure glucose, xylose and galactose to quantify released sugars.

The optimal catalysis conditions were determined in PB using either 1 % *w/v* xylan or 10 mM oNPG as substrates, for Ps_GH5 and Ps_GH50, respectively. The optimal pH of catalysis was measured in the pH range 3.0–10.0 at the optimum temperature (T_{opt}). The T_{opt} was recorded in the temperature range 10–90 °C, at pH 7 for both enzymes. Experiments were performed in triplicate and results are reported as a mean ± standard deviation.

2.7. Analysis of polysaccharide degradation products

The analysis of hydrolysis products from polysaccharides was carried out by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using a Dionex ICS-6000 system equipped with an Ag/AgCl pH reference electrode, and a gold working electrode for the detection (Dionex Corporation, CA, USA). Enzymatic reactions were set up as previously described (see Section 2.5), using xylan and arabinoxylan for Ps_GH5 or low-melting point agarose and

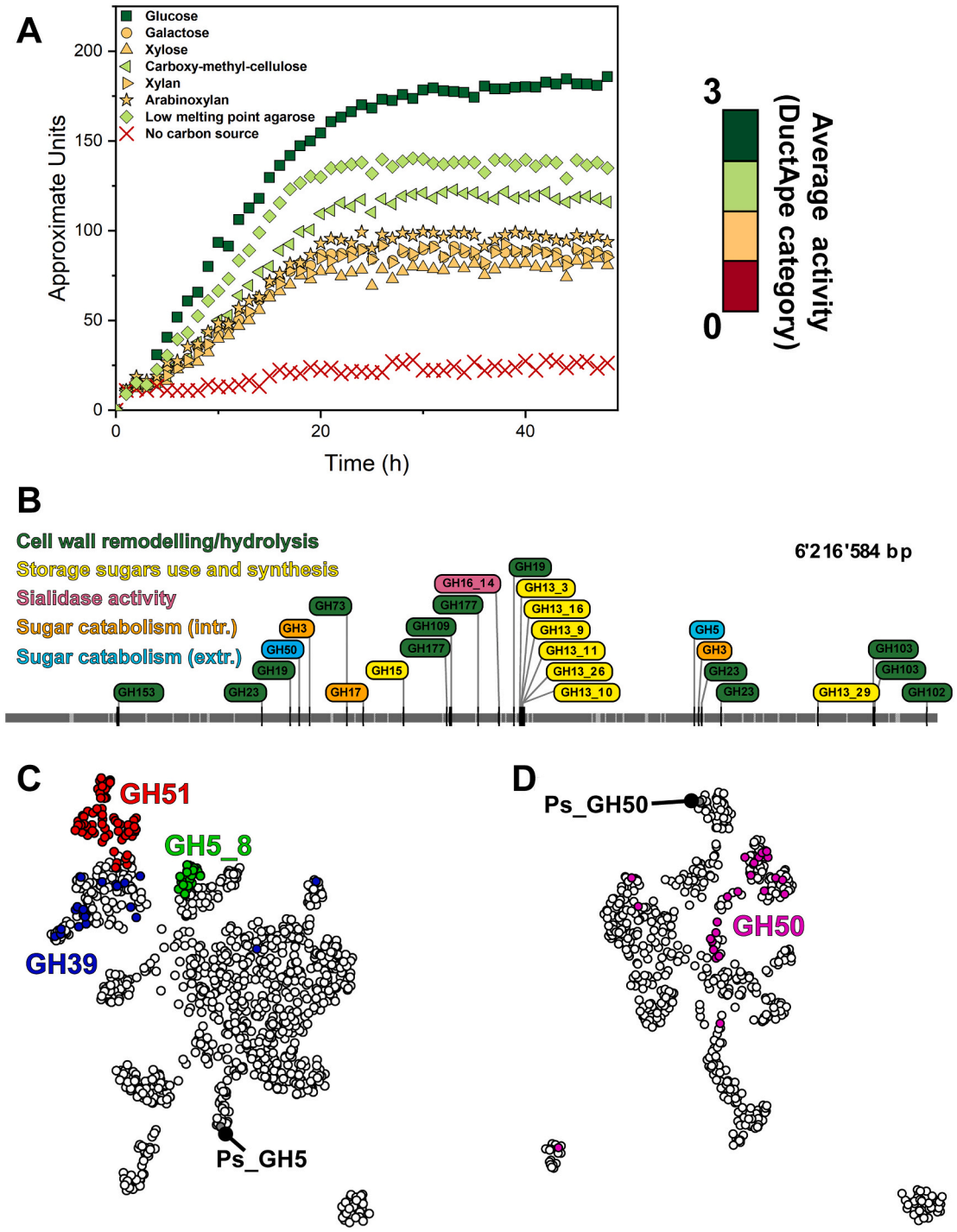


Fig. 1. Identification of GH enzymes contained in the genome of *Pseudomonas* sp. ef1. (A) Metabolic activity of *Pseudomonas* sp. ef1 recorded at OmniLog® in the presence of various sugars. The OmniLog® experiment is performed in duplicate and standard deviations are omitted for clarity ($n = 2$). (B) Localization of putative GH coding genes in the *Pseudomonas* sp. ef1 genome assembly. The annotated GHs are highlighted in different colors, according to their physiological role. "intr.": intracellular; "extr.": extracellular. (C–D) Sequence space of Ps_GH5 (C) and Ps_GH50 (D) shown as UMAP plots. The sequence space was created as described in the Material and Methods section. Empty circles: sequences of uncharacterized GHs; colored circles: sequences of characterized enzymes.

GOS 3X for Ps_GH50. Reactions involving Ps_GH5 were analyzed using a CarboPac PA200–3 mm column (Thermo Fisher Scientific, Waltham, MA, USA), with a flow rate of 0.3 mL/min. The temperature of the auto-sampler and column was set at 10 and 30 °C, respectively. Initial conditions were set to 0.1 M NaOH (eluent A) followed by a linear gradient towards an increasing proportion of a solution of 0.1 M NaOH +0.5 M NaOAc (eluent B). The gradient reached 60 % of solution B in 20 min. On the other hand, digestions involving Ps_GH50 were analyzed on a CarboPac PA210-Fast-4 μm column (Thermo Fisher Scientific, Waltham, MA, USA), with a flow rate of 0.6 mL/min and the temperature of the auto-sampler and column set as previously stated. Elution was achieved by a gradient (from 2 mM to 100 mM) of sodium hydroxide. All samples were centrifuged, diluted with dH_2O and filtered before analysis. Products of digestion were identified and quantified according to standard curves prepared in the range of 25–0.1 mg/L. All measurements were run in quintuplicate and the mean values ± SD are reported.

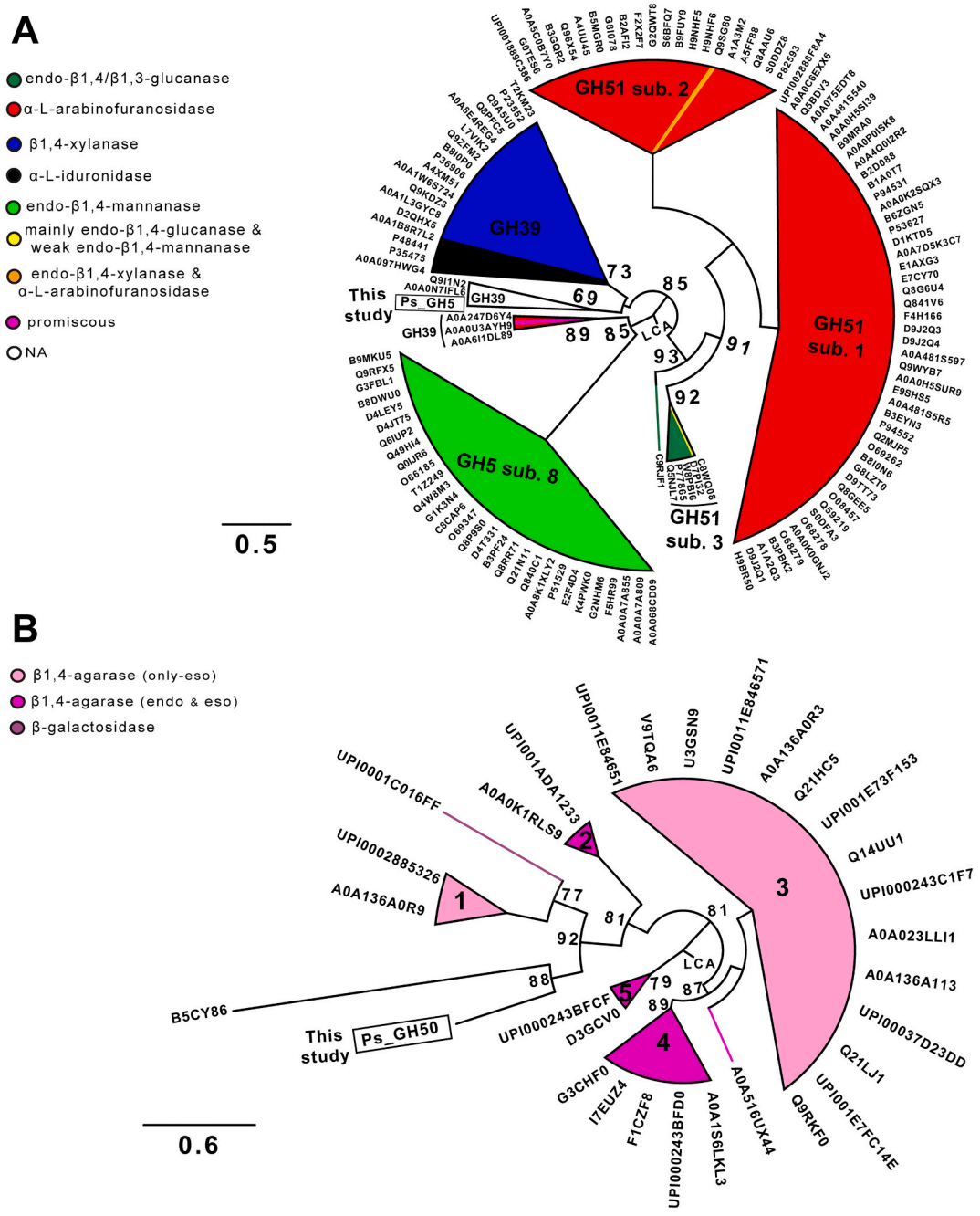


Fig. 2. Evolutionary analysis of Ps_GH5 and Ps_GH50. The rooted maximum likelihood phylogenetic trees of the catalytic domain of Ps_GH5 (A) and Ps_GH50 (B). To calculate the branch support values, 1000 ultra-fast bootstrap replicates were performed; only the values <96 are reported. For the sake of clarity, some subtrees were collapsed to facilitate the visualization of the main clusters. The horizontal bar represents the scale of the fraction of expected substitutions per site. The phylogenetic tree was visualized with FigTree v1.4.4 (https://github.com/rambaut/figtree/releases) and customized. Sequences are labeled with either their Uniprot ID or the UniParc ID (label starting with "UPI").

2.8. Thermal stability assay

Thermal unfolding was determined by recording the circular dichroism (CD) signal at fixed wavelength of 220 nm over a temperature range from 10 °C to 90 °C with a Jasco J815 spectropolarimeter (JASCO Europe, Lecco, Italy). Measurements were carried out at a protein concentration of 4 μM in a 0.1 cm pathlength quartz cuvette and a temperature slope of 1 °C/min.

Long-term thermal stability was evaluated at 35, 45 and 55 °C, by incubating the enzyme in PB, pH 7, at a concentration of 0.5 mg/mL and measuring the residual CD signal (220 nm) and specific activity using xylan and oNPG as substrates for Ps_GH5 and Ps_GH50, respectively. Activity was determined as previously described. Experiments were performed in triplicate.

3. Results

3.1. Identification of extracellular GHs in the genome of Pseudomonas sp. ef1

To guide the search in the genome of *Pseudomonas* sp. ef1 for sequences potentially encoding CAZymes involved in polysaccharide

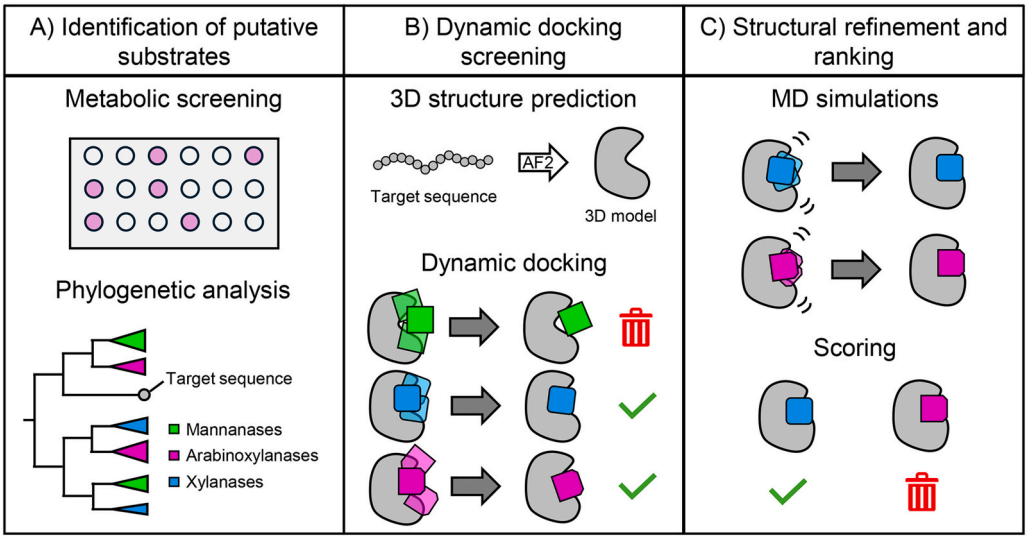


Fig. 3. Computational workflow for predicting substrate specificity. A) The putative substrates to screen are selected based on the results obtained from metabolic screening and phylogenetic analyses (A). The 3D structure of Ps_GH5 and Ps_GH50 was predicted with AF and the 3D models were used to screen the selected substrates employing a "dynamic docking" approach (B). The catalytically competent enzyme-substrate complexes were identified and refined by MD simulations and ML rescoring (LIE using neural network potentials).

degradation, a preliminary screening was designed using the Biolog Omnilog® system. This system was used to test if *Pseudomonas* sp. ef1 can metabolize monosaccharides (glucose, galactose and xylose) and polysaccharides (xylan, arabinoxylan, CMC and low melting point agarose). *Pseudomonas* sp. ef1 turned out to be metabolically active on all tested sugars at varying degrees, with glucose being the substrate on which a greater metabolic activity is observed (activity index: 3), followed by agarose and CMC (activity index: 2) and all other sugars (activity index: 1) (Fig. 1A).

Genome analysis identified 27 genes encoding putative GHs (Fig. 1B), which are associated with various physiological roles, including bacterial wall shaping or hydrolysis, storage sugar synthesis and utilization, and sialidase activity. The Biolog Omnilog data were used to filter the putative GHs identified by the genome analysis. Indeed, only the GHs putatively involved in the degradation of polysaccharides metabolized by *Pseudomonas* sp. ef1 were considered. Following this step, six predicted proteins belonging to five different families were selected: two GH3s (GenBank Protein IDs: WP_041478900, WP_041477342), GH5 (GenBank Protein ID: WP_04147731 4), GH16 (GenBank Protein ID: WP_081932489), GH17 (GenBank Protein ID: WP_041479600), and GH50 (GenBank Protein ID: WP_041478046). Of these, only the GH5 and GH50 sequences contain an N-terminal signal peptide predicted for secretion, suggesting that these enzymes may act extracellularly and are involved in the initial steps of polysaccharide degradation. Therefore, these two enzymes, Ps_GH5 and Ps_GH50, were selected for further sequence and structural analyses to predict their functional specificity.

3.2. Phylogenetic analyses of Ps_GH5 and Ps_GH50

The sequence diversity of Ps_GH5 and Ps_GH50 was investigated by combining the sequence space of their families with phylogenetic analyses. The sequence space of the two selected enzymes was generated using a dataset containing sequences of GHs sharing a global sequence identity higher than 10 % with Ps_GH5 or Ps_GH50 and whose 3D structure is available in PDB or in AFDB. The space of Ps_GH5 contains sequences belonging to GH5 subfamily 8 (GH5_8), GH39 and GH51 (Fig. 1C), suggesting that Ps_GH5 is evolutionarily related also to the GH39 and GH51 families. On the other hand, only GH50 family sequences are contained in the sequence space of Ps_GH50 (Fig. 1D). By mapping the characterized GHs in the sequence space, it is evident that both sequences, especially Ps_GH5, are in an unexplored region, with global sequence identity <30 % to any biochemically characterized homologue available in the CAZy database.

To infer the substrate specificity of the two selected enzymes, we explored the phylogenetic relationship in the frame of characterized GHs only, by using a structure-based alignment of the catalytic domains, to avoid any bias due to recent duplications and horizontal gene transfers of accessory domains. Ps_GH5 is nested in a cluster that includes GH39 enzymes (Fig. 2A), which are mostly β-xylanase in *Bacteria* and α-L-iduronidase in *Eukaryotes*. This cluster is related to GH5_8 (β-mannanase) and GH51 (mostly α-L-arabinofuranosidase), with the latter being the more distant in evolutionary terms. GH39s with catalytic domains evolutionary closest to Ps_GH5 were isolated from *P. aeruginosa* PsIG (PDB ID: 4ZN2, Uniprot ID: Q9I1N2, [44,45], 18 % global sequence identity to Ps_GH5) and from *Bacteroides cellulosilyticus* WH2 (Uniprot ID: A0A0N7IFL6, [46], 16.3 % global sequence identity to Ps_GH5). Although the specificity of such GH39s was not clarified, PsIG has a degradative effect on the complex secreted extracellular matrix of *P. aeruginosa*, which contains a repeating pentasaccharide consisting of D-mannose, D-glucose, and L-rhamnose 3:1:1 [47].

The phylogenetic tree of GH50s reveals five distinct clusters (Fig. 2B). Cluster 3 is the largest and contains only *exo*-β-agarases. Clusters 2, 4 and 5 contain enzymes with both exo- and *endo*-β-agarase activities, while cluster 1 contains enzymes with exo-lytic activity. Only one GH50, related to cluster 1, is active on lactose instead of agarose. The catalytic domain of Ps_GH50 is phylogenetically related to the GH50 enzyme from *Phocaeicola plebeius* DSM 17135 (PDB ID: 5T3B, Uniprot ID: B5CY86, 16.8 % global sequence identity to Ps_GH50). This enzyme does not exhibit the expected β-agarase activity and does not show detectable activity on other related polymeric algal galactans [48]. Concerning GH50 evolution, enzymes with exo-activity are more distant from the root of the tree than those with endo-activity, suggesting that during the evolution of this family an independent loss of the endo-activity occurred. Collectively, Ps_GH5 and Ps_GH50 are evolutionarily located in a region of the sequence space that is poorly explored and characterized, with an unknown specificity for the closest homologs.

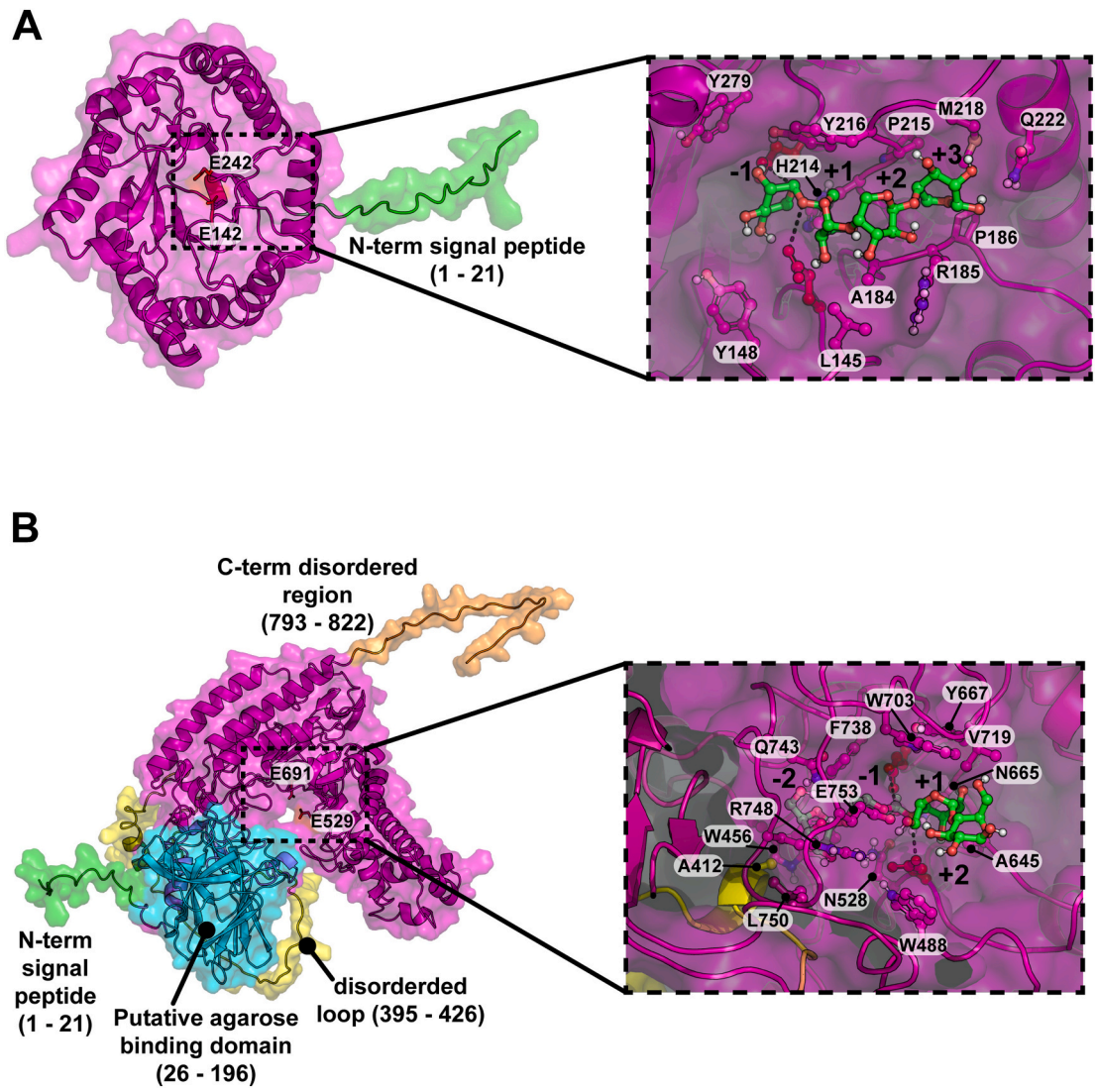


Fig. 4. 3D model of Ps_GH5 and Ps_GH50. Left: 3D model of Ps_GH5 (A) and Ps_GH50 (B) were predicted using AF2 in the absence of substrate. The catalytic domains are colored in magenta, and the catalytic residues are represented as red sticks and labeled with one-letter code. Right: substrate binding site complexed with XOS 4X for Ps_GH5 and GOS 4X for Ps_GH50. The docking poses were extracted from the MD frame with the more negative DeepQM score (higher affinity). Residues that interact with the substrate (distance <4 Å to the substrate) are shown as ball-and-sticks and labeled with one-letter code. Subsites are numbered according to [77]. The docked substrates are visualized in ball-and-sticks with carbon atoms green colored. The C—O and C—C distances are marked by a black dashed line.

3.3. In silico screening of substrates putatively hydrolyzed by Ps_GH5 and Ps_GH50

Considering the orphan evolutionary placement of Ps_GH5 and Ps_GH50, a structural bioinformatics pipeline was employed to predict their substrate specificity (Fig. 3). This pipeline starts from the information derived from the metabolic capacity of *Pseudomonas* sp. ef1 and from phylogenetic analyses and includes the following steps: (i) prediction of the 3D structure of Ps_GH5 and Ps_GH50 with AF (average plDDT of Ps_GH5 and Ps_GH50 are 91.67 and 87.21, respectively, Fig. S1); (ii) screening of different oligosaccharides through a "dynamic docking" approach; (iii) identification of the catalytically competent enzyme-substrate complexes; and (iv) refinement by MD simulations and rescoring with LIE using neural network potentials.

According to AF models, Ps_GH5 (34.0 KDa, 318 AA) is organized in a catalytic domain with the classical $(\alpha/\beta)_8$ TIM barrel fold of clan A (Fig. 4A). In contrast, Ps_GH50 (89.2 KDa, 822 AA) is a two-domain enzyme that includes a putative β sheet-rich N-terminal agarose-binding domain (homologous to the CBM11 domain described in [49]) and a mainly α catalytic domain, an extension of the classical $(\alpha/\beta)_8$ TIM barrel fold of clan A, as also shown in [50]. This architecture closely resembles that of characterized GH50 enzymes with an accessory binding domain, except for two regions, a loop located between amino acids 395 and 426 and residues 793–822 at the C-terminal end (Fig. 4B). Both loops are predicted to be disordered by the consensus method MobiDB-lite 3.0 [51].

These models were docked with a list of oligosaccharides representative of the substrates metabolized by *Pseudomonas* sp. ef1 and those hydrolysed by phylogenetically closest homologs. More in detail, Ps_GH5 was docked with disaccharides (2X), trisaccharides (3X) and tetrasaccharides (4X) derived from arabinan (AOS), cellulose (COS), xylan (XOS) and mannan (MOS). The α-L-arabinofuranosidase activity, typical of some GH51 active on arabinoxylan [52,53], was tested by docking with D-xylose-α-1,2-L-arabinofuranoside and D-xylose-α-1,3-L-arabinofuranoside (AXOS). Ps_GH50 was docked with 2X (neo-agarobiose), 4X and 6X oligosaccharides (NAOS) derived from agarose degradation, and with 2X, 3X and 4X GOS, containing glucose at the non-reducing end. Since loops could play a crucial role in the interaction between enzyme and substrate [54], a full conformational sampling of both substrate and enzyme was considered in the docking step. Ps_GH5

Table 1
MD refinement of docked poses to predict substrate specificity. The binding free energy is estimated by DeepQM score. The DeepQM score and the catalytic distances are averaged over three independent MD simulation runs. The values in the brackets are those from the simulation frame with the more negative DeepQM score (higher predicted affinity). The best metrics for each enzyme are shown in bold. Distances <4 Å were considered catalytically competent (see Materials and methods for further details).

Enzyme	Substrate	DeepQM score[a]	Distance C—O (Å)	Distance O—O (Å)
Ps_GH5	COS 2X	−2.48 ± 0.52 (−4.22)	5.27 ± 0.55 (5.1)	6.39 ± 1.35 (5.1)
	XOS 4X	**−3.06 ± 0.58 (−4.68)**	**3.67 ± 0.53 (3.6)**	4.02 ± 0.40 (**3.6**)
	MOS 4X	−2.89 ± 0.74 (−4.57)	4.83 ± 1.70 (4.4)	4.97 ± 1.43 (4.5)
	AOS 4X	−3.32 ± 1.00 (−6.11)	5.27 ± 1.57 (4.4)	5.65 ± 0.80 (5.4)
	D-xylose α-1,2-L-arabinofuranoside	−2.63 ± 0.80 (−4.2)	4.39 ± 0.76 (4.2)	4.86 ± 1.88 (3.5)
	D-xylose α-1,3-L-arabinofuranoside	−2.72 ± 0.70 (−4.40)	**3.49 ± 0.86** (3.3)	5.07 ± 0.68 (4.4)
Ps_GH50	NAOS 6X	−3.87 ± 0.79 (−6.11)	4.68 ± 1.17 (**3.6**)	5.51 ± 1.05 (4.6)
	GOS 4X	−3.57 ± 0.73 (−5.94)	**4.17 ± 0.42** (3.9)	**4.12 ± 1.07 (3.2)**

[a] The lower the better. Please refer to (40) for further explanations.

and Ps_GH50 can be catalytically docked with most of the tested substrates (Table S1). The only exceptions are COS 3X and COS 4X for Ps_GH5 and NAOS 2X for Ps_GH50. Moreover, Ps_GH50 was competent only for the β-1,4-glycosidic bonds that linked the neoagarobiose units in the tested NAOS 4X and 6X. Specifically, the highest DynamicBind scores are observed with the 4X oligosaccharides for Ps_GH5 and with both the GOS 4X and NAOS 6X for Ps_GH50. To verify if the catalytically competent docking poses of the substrates are physically plausible, after preliminary screening with DynamicBind, these most promising substrates for each oligosaccharide class were selected for MD refinement. The refined enzyme-substrate complexes were evaluated through the measurement of average C—O and C—C catalytic distances, as well as through the analysis of the DeepQM score, which predicts the interaction energy from an end-state simulation more accurately than traditional approaches relying on force field terms [41]. Among the docked substrates, XOS 4X can be simulated in a catalytically competent binding mode for Ps_GH5, while D-xylose α-1,3-L-arabinofuranoside is found in a suboptimal catalytically competent binding mode (Table 1). All the other tested substrates (COS 4X, MOS 4X and AOS 4X) interact in non-catalytic binding modes during MD simulations with Ps_GH5. In the case of Ps_GH50 the GOS 4X remained in a catalytically competent mode only for part of the simulations, including the most energetically favored conformation (Table 1). Consequently, the analyses of C—O and C—C distances indicate a suboptimal binding mode, as evidenced by the average values, which are slightly higher than the threshold value (4.0 Å).

Overall, the bioinformatics pipeline enabled the prediction of the substrate specificity of the two enzymes. It was determined that Ps_GH5 is likely to be active on xylan and possibly also on arabinoxylan polysaccharides, while Ps_GH50 may be more efficient in degrading β-1,4 bonds of GOS than NAOS. To validate these predictions, Ps_GH5 and Ps_GH50 were recombinantly produced in *E. coli* cells and purified by metal-affinity chromatography (Fig. S2). Both enzymes were soluble and in a monomeric state (Table S2).

3.4. Biochemical features of Ps_GH5 and Ps_GH50

The substrate specificity of recombinant Ps_GH5 and Ps_GH50 was investigated using *ortho*- and *para*-nitrophenyl sugars derivatives as well as different polysaccharides reported in the *Material and Methods* section. These sugars are representative of those metabolized by *Pseudomonas* sp. ef1 and those tested *in silico*.

The extent of the hydrolytic activity on polysaccharides was preliminarily determined with the DNS assay. Ps_GH5 shows a faint specific activity on pNPXyl (0.3 ± 0.1 U/mg) and is active in the hydrolysis of xylan and arabinoxylan. On the other hand, Ps_GH50 shows hydrolytic activity towards oNPG (4.4 ± 0.2 U/mg) and only a weak activity on agarose, as indicated by the DNS assay (equivalent sugars released: 19.5 ± 5.7 μg/mL). Therefore, xylan and oNPG were selected as substrates to determine the biochemical properties of Ps_GH5 and Ps_GH50, respectively. Both enzymes are active in a pH range of 5.0 to 9.0, with an optimum pH of 7.0 and 8.0 for Ps_GH5 and Ps_GH50, respectively (Fig. 5A, B). Ps_GH5 exhibits highest activity at 45 °C and retains ≃ 90 % of its activity at 40 °C (Fig. 5C), whereas Ps_GH50 showed the optimal activity at 55 °C (Fig. 5D) and a broader temperature profile compared to Ps_GH5. Both enzymes are active to different extents at 10 °C: Psd_GH5 retains 10 % of its activity, while Ps_GH50 retains 15 %.

Thermal stability experiments, carried out by CD spectroscopy, show that Ps_GH50 (T_m of 62.9 ± 1.2 °C) is more thermostable than Ps_GH5 (T_m of 55.0 ± 0.8 °C, Fig. 5E, F). This behavior was also observed in long-term thermal stability assays performed at 35 and 45 °C. Ps_GH5 is completely inactivated after five and one days of incubation at 35 °C and 45 °C, respectively (black line in Fig. 6A, B). This enzymatic inactivation is due to heat-induced loss of secondary structure, as reported by monitoring the CD signal at 220 nm (black line in Fig. 6C, D). In contrast, Ps_GH50 maintains 60 % of its activity and 75 % of the CD signal after seven days of incubation at 35 °C (red line in Fig. 6A, B), and 35 % of both activity and CD signal after seven days at 45 °C (red line in Fig. 6C, D).

3.5. Determination of degradation products obtained from hydrolysis reactions with Ps_GH5 and Ps_GH50

To gain further insight into the mechanisms of polysaccharide hydrolysis, the degradation products obtained in the presence of Ps_GH5 and Ps_GH50 were analyzed by HPAEC-PAD after 2 and 24 h of incubation. The degradation reactions were carried out at 35 °C, which proved to be a good compromise between the optimal temperature for catalysis and the thermal stability of the enzymes (Figs. 5 and 6). Ps_GH5 is active on xylan, releasing mainly XOS (xylotriose 3X, xylotetraose 4X, and xylopentose 5X) and small amounts of xylose (Fig. 7A, B). A similar degradation pattern was observed in the presence of arabinoxylan (Fig. 7D), although approximately 10 times lower sugars were released with respect to the use of xylan as substrate (Fig. 7E). Overall, these results indicate that Ps_GH5 is an *endo*-xylanase, with the capacity to release only a faint amount of α-L-arabinose units when arabinoxylan is used as a substrate (Fig. 7C, F). These findings are consistent with the *in silico* simulations, which predicted the optimal interaction of Ps_GH5 with XOS 4X and suboptimal interaction with D-xylose α-1,3-L-arabinofuranoside.

As concerning Ps_GH50, the HPAEC-PAD chromatograms of agarose degradation products do not show any peaks, probably due to the very low concentration and/or to the length of the hydrolysis products, which may exceed the column cut-off. On the other hand, Ps_GH50 appears more active in the hydrolysis of the GOS 3X, yielding galactose and glucose as the predominant degradation products after 2 and 24 h of reaction (Fig. 8A and B). Following a 24-h reaction, the molar ratio of galactose to glucose was observed to be 1.7. Additionally, the presence of lactose and galactobiose was detected, indicating that the hydrolysis

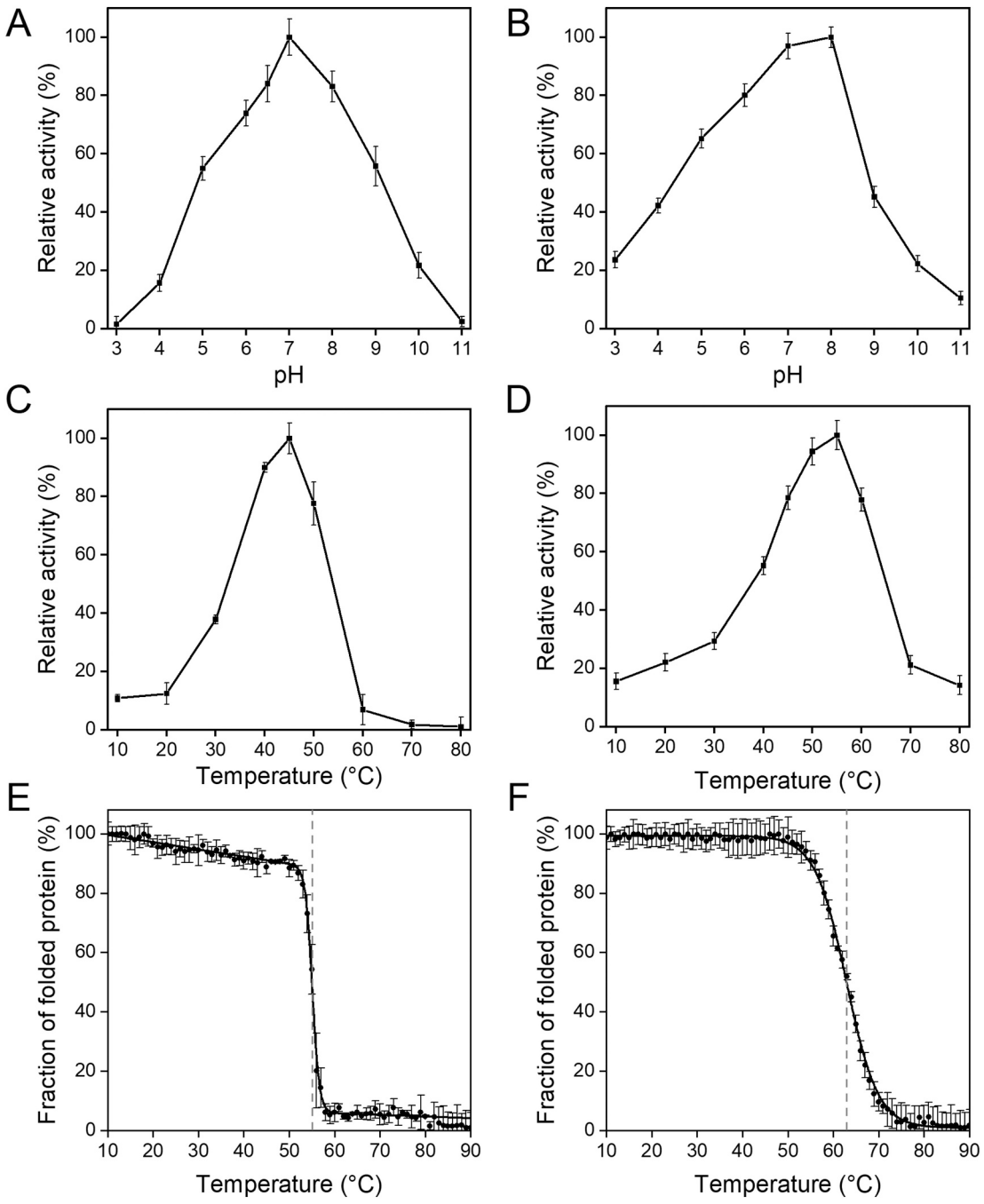


Fig. 5. Biochemical features of Ps_GH5 and Ps_GH50. Effects of pH on the activity of Ps_GH5 (A) and Ps_GH50 (B). Temperature profile of Ps_GH5 (C) and Ps_GH50 (D). The activity of Ps_GH5 and Ps_GH50 were monitored using xylan and oNPG as substrates, respectively. Thermal denaturation profiles of Ps_GH5 (E) and Ps_GH50 (F), determined by CD spectroscopy. Ellipticity values were recorded at 220 nm during heating from 10 to 90 °C. The initial CD signal was taken as 100 % for normalization. The dashed gray line corresponds to T_m temperatures. All the experiments were performed in triplicate and standard deviation is represented ($n = 3$).

reaction is not yet complete. The experimental results and molecular docking simulations are in accordance with the hypothesis that Ps_GH50 is more efficient on GOS and capable of hydrolysing both β1–4 glycosidic bonds between glucose and galactose and between two galactose moieties, potentially releasing galactose, lactose and galactobiose from GOS hydrolysis (Fig. 8C).

3.6. Structural analysis of the substrate binding site of Ps_GH5 and Ps_GH50

The substrate binding sites of Ps_GH5 and Ps_GH50 were predicted by analyzing the interactions between Ps_GH5 and xylotetraose, and Ps_GH50 and GOS 4X in the best (in terms of DeepQM score) docking poses. In both enzymes the substrates interact with two Glu residues at ≃ 6 Å distance to each other (E142 and E242 for Ps_GH5, E691 and E529 for Ps_GH50), which are the catalytic dyad (nucleophile and acid/base, respectively, Fig. 4A, B). This distance is consistent with a retaining hydrolytic mechanism that has been previously reported for other enzymes belonging to the GH5, GH39 and GH50 families (http://www.cazy.org/GH5.html, http://www.cazy.org/GH39.html, http://www.cazy.org/GH50.html).

In Ps_GH5, the substrate binding site is a shallow cleft, comprising at least four distinct subsites (−1, +1, +2 and +3). The structure of the substrate binding site of Ps_GH5 was superimposed on those of two

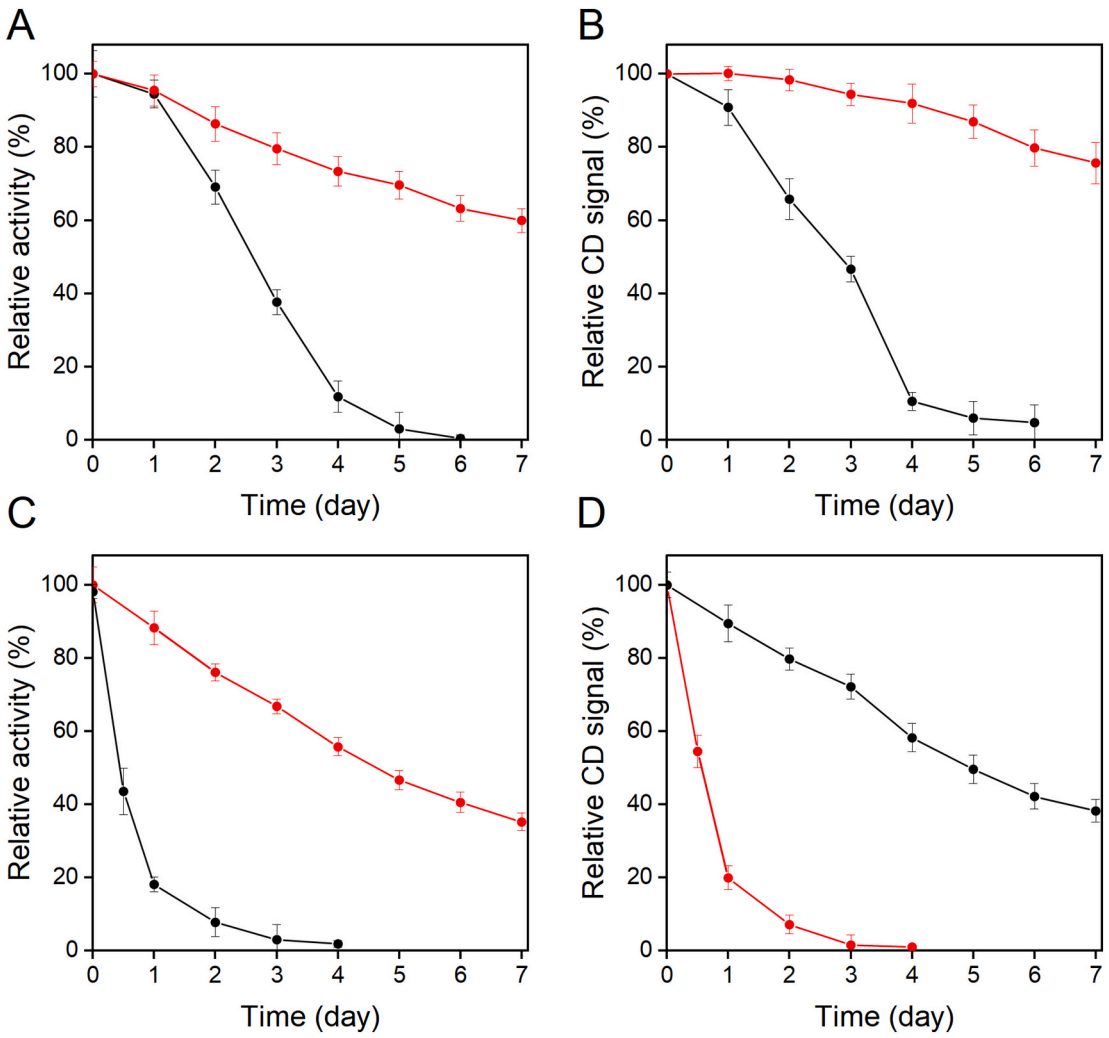


Fig. 6. Long-term thermal stability. Relative activity of Ps_GH5 (black) and Ps_GH50 (red) determined at 35 °C (A) and 45 °C (C). The activity of Ps_GH5 and Ps_GH50 were monitored using xylan and oNPG as substrates, respectively. Relative CD signals of Ps_GH5 (black) and Ps_GH50 (red) recorded at 220 nm after incubation at 35 °C (B) and 45 °C (D). All the experiments were performed in triplicate and the standard deviation is reported (n = 3).

evolutionary closest homologs, namely GH5_8 from *Thermomonospora fusca* (PDB: 2MAN) and GH39 from *P. aeruginosa* (PDB: 4ZN2). The structural comparison shows that the residues at subsites −1 and + 1 exhibit comparable physicochemical properties, whereas the main differences are observed at subsites +2 and +3, which are more hydrophilic in Ps_GH5 than in its closest homologs (Fig. 9A). Notably, these subsites do not contain bulky aromatic residues usually observed in GH5 and GH39 counterparts. It is conceivable that residue R185 in subsite +2 plays a role in substrate positioning or stabilization within the active site (Figs. 9A and 4A).

On the contrary, Ps_GH50 possesses a tunnel-shaped catalytic grove with at least four distinct subsites (−2, −1, +1 and + 2, Fig. 4B). Comparison of the 3D model of Ps_GH50 with and without GOS 4X shows a distinct reorientation of loop 395–426 during the docking step, to form a part of the −2 subsite through the interaction of residue A412 (Fig. 4B). This region, comprising also W456, Q743 and L750, is the most diverse with respect to the closest evolutionary homologs from *Paraglaciecola hydrolytica* (Uniprot ID: A0A136A0R9) and from *Victivallis vadensis* (Uniprot ID: UPI0001C016FF, Fig. 9B): Q743 is present in all the homologs, L750 is present only in Ps_GH50, while W456 is conserved in the non-agarolytic GH50 from *V. vadensis* and absent in *P. hydrolytica* agarase. The physicochemical properties of the other subsites are quite conserved.

4. Discussion

Pseudomonas sp. ef1 is an Antarctic bacterium isolated from the microbial consortium of *Euplotes focardii*. This bacterium has received attention for its ability to synthesize silver nanoparticles and encode two orphan GH19s with lysozyme activity instead of the typical chitinolytic one [17,55]. Despite its ability to metabolize cellulose along with various polysaccharides of terrestrial (xylan and arabinoxylan) and marine (agarose) origin, only two extracellular enzymes, Ps_GH5 and Ps_GH50, have been identified to be involved in polysaccharide degradation. Strikingly, Ps_GH5 and Ps_GH50 are not active on cellulose, suggesting that the genome of *Pseudomonas* sp. ef1 contains other GH(s) that evade the conventional methods of sequence annotation.

Since Ps_GH5 and Ps_GH50 share <30 % sequence identity with the characterized enzymes of their respective families, they lie in an unexplored region of the sequence and structural space. This makes it challenging to infer the substrate specificity of Ps_GH5 and Ps_GH50 using traditional sequence-based enzyme discovery methods (phylogenetic relationships [56], SSNs [7], and BLAST searches) or recently developed deep learning methods (Table 2).

The workflow proposed in this study allows for the prediction of substrate specificity by integrating a “dynamic docking” method with a physics-based MD sampling approach. In our pipeline, the metabolic capacity of *Pseudomonas* sp. ef1 was used to drive genome analysis and

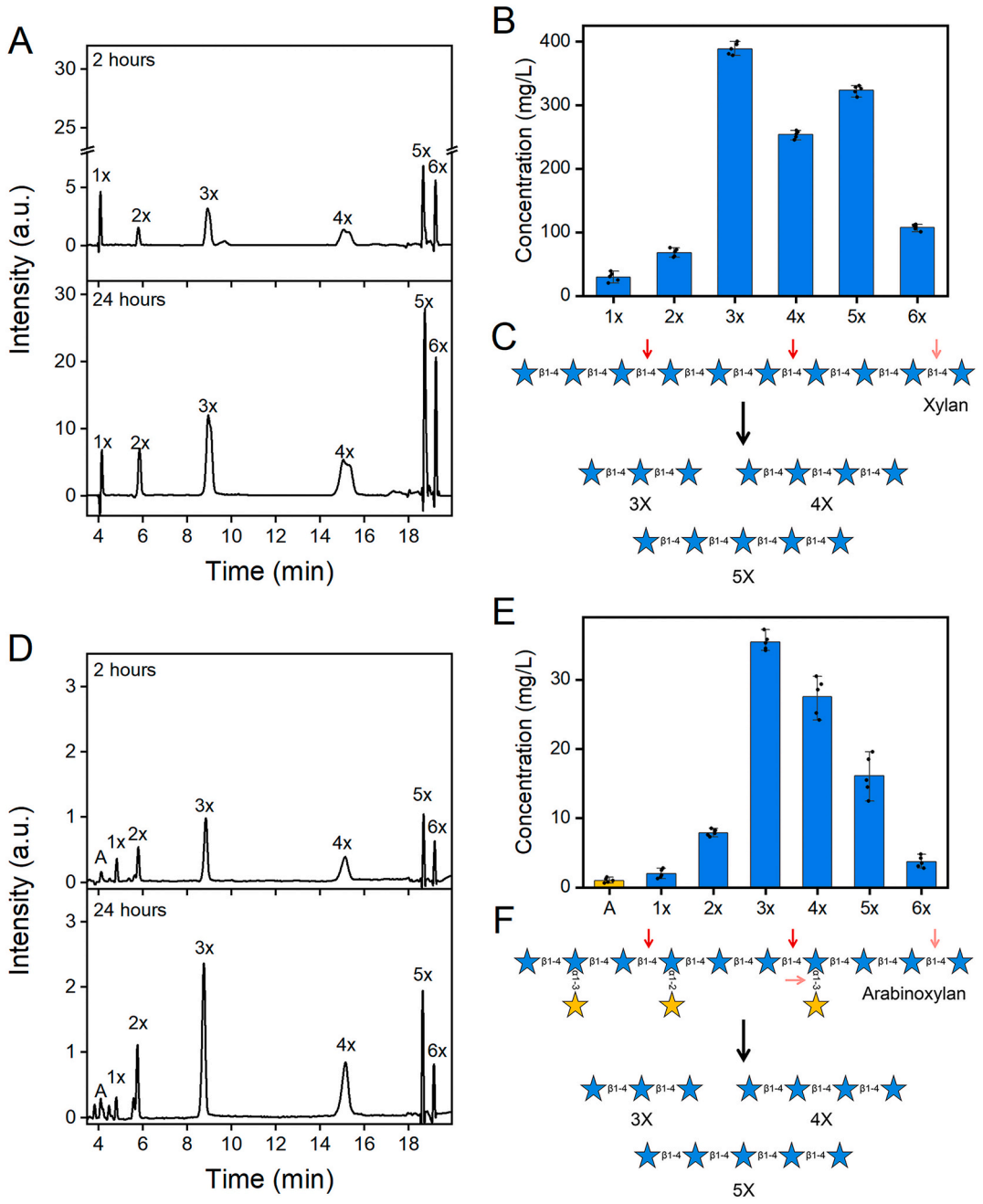


Fig. 7. Xylan and arabinoxylan degradation in the presence of Ps_GH5. (A) HPAEC-PAD chromatogram of xylan degradation products obtained after 2 h and after 24 h of incubation. (B) Quantification of xylan degradation products obtained after 24 h of reaction. (C) Scheme of the degradation pattern of xylan in the presence of Ps_GH5; red arrows indicate the β1–4 glycosidic bonds hydrolyzed by Ps_GH5. (D) HPAEC-PAD chromatogram of arabinoxylan degradation products obtained after 2 h and after 24 h of incubation. (E) Quantification of arabinoxylan degradation products obtained after 24 h of reaction. (F) Scheme of the degradation pattern of arabinoxylan in the presence of Ps_GH5; red arrows indicate the β,1–4 glycosidic bonds hydrolyzed by Ps_GH5, with the probability of hydrolysis proportional to the colour intensity. Reactions were performed with either 1 % *w/v* of xylan or arabinoxylan and 0.5 mg/mL of Ps_GH5 at 35 °C with shaking. All the experiments were performed in quintuplicate and standard deviation is reported in panels B and E ($n = 5$).

to select substrates for the *in silico* screening. To simplify simulation and reduce computational resources, these substrates were represented as medium-chain oligosaccharides. Computational simulations indicate that Ps_GH5 interacts in a catalytically competent manner mainly with XOS 4X and less efficiently with D-xylose-α-1,3-L-arabinofuranoside, representing xylan and arabinoxylan. These simulations are in accordance with the experimental results, which show a more marked activity towards xylan than arabinoxylan. Regarding the substrate specificity of Ps_GH50, although its structure-based prediction is more challenging owing to the presence of a disordered loop 395–426, there is good agreement between the computational simulation and Ps_GH50 activity. Indeed, the experimental results indicated that Ps_GH50 exhibits greater activity on small GOS than agarose, which is in accordance with the best distance metrics in the computational simulation. A comparison of Ps_GH5 and Ps_GH50 with their biochemically characterized homologs reveals that both enzymes exhibit distinctive enzymatic activities that have not been previously reported in their homologs (Table S3). Despite the challenges posed by the low sequence identity and the absence of a systematic approach to studying polysaccharide degradation, it can be concluded that, based on both computational and functional analyses,

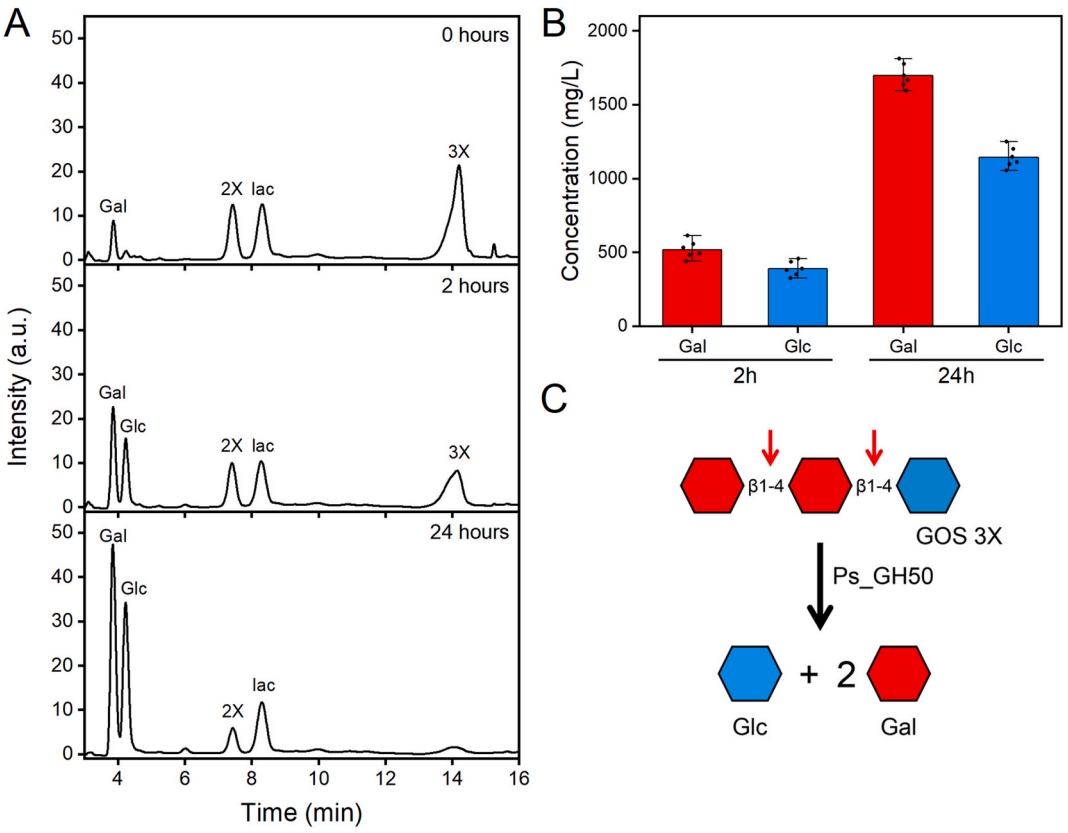


Fig. 8. GOS degradation in the presence of Ps_GH50. (A) HPAEC-PAD chromatograms of GOS degradation at time 0 and after 2 and 24 h of incubation in the presence of Ps_GH50. Reactions were carried out at 35 °C in the presence of 0.5 mg/mL of Ps_GH50 and 1 % *w/v* of GOS. (B) Quantification of galactose and glucose after 2 and 24 h of incubation in the presence of Ps_GH50. (C) Scheme of GOS 3X hydrolysis in the presence of Ps_GH50; red arrows indicate the β1–4 glycosidic bonds attacked by Ps_GH50. All the experiments were performed in quintuplicate and standard deviation is reported in panel B (n = 5).

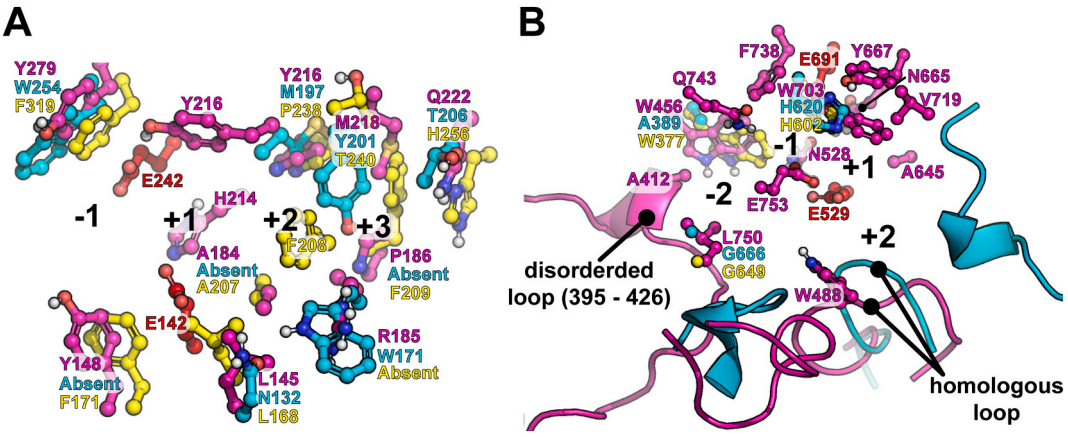


Fig. 9. Comparison of the substrate binding sites of Ps_GH5 (A) and Ps_GH50 (B) with those of their evolutionary closest homologs. The substrate binding residues of Ps_GH5 and Ps_GH50 are represented as in ball-and-sticks and colored in magenta. In panel A the non-conserved residues of GH5 from *T. fusca* (PDB: 2MAN) and GH39 from *P. aeruginosa* (PDB: 4ZN2) are colored in cyan and yellow respectively. In panel B the non-conserved residues of the 3D model of GH50 from *P. hydrolytica* (Uniprot ID: A0A136A0R9) and from *V. vadensis* (Uniprot ID: UPI0001C016FF) are colored in cyan and yellow respectively. Non-conserved loops surrounding the entrance of the substrate-binding tunnel are shown in cartoons. Subsites are numbered according to [77] and based on Fig. 3.

Ps_GH5 and Ps_GH50 are likely to belong to novel subfamilies of GH5/GH39 and GH50, respectively.

This study highlights the efficacy of dynamic docking simulations to infer substrate specificity especially for enzymes with flexible active sites, such as cold-active enzymes [57]. In general, the proposed workflow, which combines dynamic docking and MD simulations, could be applied in a wide range of scenarios to predict and describe the substrate specificity of enzymes, including those poorly represented in training datasets, such as sequences from metagenomic campaigns or from *de novo* design [58–62]. In addition, this computational workflow can be integrated with conventional high-throughput activity-based assays for the discovery of novel catalytic functions [63].

In the context of a general low sequence identity with their closest evolutionary homologs, we focused on the peculiarities of the active sites of the two enzymes, as our workflow allowed us to have a direct interpretation of the inferred substrate specificity. The active site of

Table 2
Predictions of Ps_GH5 and Ps_GH50 substrate specificity by different SOTA approaches (homology and deep learning-based). The values reported for ProSmith and VIPER are the pair score of the enzyme-substrate interaction, the catalytic efficiency was reported for CatPred, the values reported for TurNup and DeepEnzyme refer to the catalytic constant (s^{-1}). The mean and standard deviation values were obtained by running the models with the oligomers of different lengths. EC 3.2.1.4: *endo*-1,4-β-D-glucanase; EC 3.2.1.8: endo-1,4-β-D-xylanase. EC 3.2.1.81: endo-β-agarase. NA: not available.

Enzyme	Blast top-1 hit[a]	CLEAN [78]	Substrate	ProSmith [79]	TurNup [80]	CatPred [81]	DeepEnzyme [82]	VIPER [83]
Ps_GH5	NA	EC 3.2.1.4 (Low confidence)	COS	0.050 ± 0.017	37.99 ± 1.82	6.99 ± 2.25	**5.74 ± 0.002**	0.053 ± 0.020
			XOS	0.017 ± 0.005	42.55 ± 6.07	4.79 ± 1.14	5.28 ± 0.05	0.096 ± 0.033
			MOS	**0.05 ± 0.017**	40.26 ± 1.82	**7.74 ± 2.45**	**5.74 ± 0.002**	0.052 ± 0.018
			ARAO	0.030 ± 0.017	43.67 ± 10.87	3.25 ± 0.59	5.49 ± 0.18	0.134 ± 0.027
			AXOS	0.03 ± 0	**99.25 ± 3.68**	6.68 ± 0.88	5.62 ± 0.027	**0.154 ± 0.07**
Ps_GH50	EC 3.2.1.81	EC 3.2.1.8	NAOS	**0.183 ± 0.176**	119.74 ± 42.96	**17.09 ± 3.79**	**12.96 ± 0.66**	**0.057 ± 0.037**
			GOS	0.063 ± 0.012	**145.35 ± 0**	11.14 ± 3.65	12.17 ± 0.28	0.055 ± 0.017

[a] The Blast search was performed on the Uniprot Swissprot database (Release 2024_04). The prediction is the E. C. number of the top-1 hit according to default Blast ranking, accessed through https://www.uniprot.org/blast. A maximum e-value of e^{-10} was used.

Ps_GH5 is shallower and more hydrophilic than those of its structural homologs from the two closest GH families, namely GH5_8 from *T. fusca* (PDB: 2MAN) and GH39 from *P. aeruginosa* (PDB: 4ZN2). The aforementioned features indicate a prominent role of electrostatic and polar interactions in the correct positioning of the substrate. Although the substrate binding residues are not conserved, a similar prevalence of charged/hydrophilic residues in positions +2 and +3 was observed in GH39 from *Bacteroides cellulosilyticus* [46] and from *Xanthomonas citri* [64]. More in detail, our structural analyses suggest that the R185 residue, situated in subsite +2, is a determinant of substrate specificity. Nevertheless, a high-resolution complex between Ps_GH5 and XOS and additional structural investigations of the enzyme-substrate interaction are necessary to elucidate the function of this residue.

The active site of Ps_GH50 is a tunnel-shaped catalytic groove with four subsites from −2 to +2, as observed in other *exo*-acting GH50s [49,50,65]. In contrast, 3D models of characterized endo-type GH50s display at least eight different subsites, from −4 to +4 [66]. The peculiar specificity of Ps_GH50 for GOS can be attributed to loop 395–426 and the presence of W456 on top of subsite −2, which may contribute to an increase in steric bulk, thereby negatively affecting the required positioning of the 3,6-anhydro-α-L-galactopyranose moiety of the neoagarobiose subunits. In general, the structural reasons for the substrate specificity of GH50s are still unclear due to the paucity of available 3D structures, the scarcity of activity data on GOS, and the conservation of most of the active site residues in non-agarolytic GH50s, such as those from *V. vadensis* [67] and from *Bacteroides plebeius* [48].

As the selected enzymes were identified in the genome of an Antarctic bacterium, the hallmarks of cold adaptation were sought. Both enzymes show activity at 10 °C, and enzymatic inactivation precedes the loss of secondary structure, with a temperature gap of 10 °C and 7.9 °C for Ps_GH5 and Ps_GH50, respectively. These two features represent the main characteristics of cold adaptation as reported for other cold-active enzymes [68–71]. In terms of thermal stability, Ps_GH5 is less thermostable than Ps_GH50, which behaves as a mesophilic enzyme. A similar behavior was observed in other cold-active agarases isolated from the Antarctic *Pseudoalteromonas* sp. NJ21 and *Vibrio natriegens* WPAGA4, which effectively hydrolyse agarose at 40 °C [72,73]. Although this behavior may seem peculiar, it is not unusual to encounter psychrophilic enzymes that couple cold activity with high thermal stability [74–76]. The low sequence identity of both Ps_GH5 and Ps_GH50 with respect to their mesophilic or thermophilic homologs makes it difficult to identify the cold adaptation mechanisms. However, in Ps_GH50, the extended and disordered loops 395–426 and 793–822, which are absent in its homologs, may play a role in activity at low temperatures by increasing the structural flexibility of the enzyme. To test this hypothesis, further work is required to solve the experimental 3D structure and to perform rational design mutagenesis.

5. Conclusion

In this study, a functional and sequence-based workflow was developed to discover two novel enzymes from the genome of *Pseudomonas* sp. ef1 belonging to potentially new subfamilies of the GH5/GH39 and GH50 families. These enzymes are putatively secreted and are involved in the degradation of complex sugars such as xylan and arabinoxylan for Ps_GH5 and GOS for Ps_GH50. In conclusion, Ps_GH5 and Ps_GH50 belong to the repertoire of unusual GHs found in the *Pseudomonas* sp. ef1 genome, which enables the organism to cope with both low temperatures and low nutrient availability.

CRediT authorship contribution statement

Marco Orlando: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Alessandro Marchetti:** Writing – review & editing, Visualization, Investigation, Data curation. **Luca Bombardi:** Writing – review & editing, Visualization, Investigation, Data curation. **Marina Lotti:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Salvatore Fusco:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Data curation, Conceptualization. **Marco Mangiagalli:** Writing – review & editing, Writing – original draft, Validation, Supervision, Data curation, 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 Milan-Bicocca. Marina Lotti reports financial support was provided by University of Milan-Bicocca. Salvatore Fusco reports was provided by Ministero dell'Università e della Ricerca (MUR). 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.

Acknowledgments

The Authors acknowledge Sandra Pucciarelli who kindly provided the *Pseudomonas* sp. e1 strain. This work was supported by the University of Milano-Bicocca with FA (Fondo di Ateneo) to M.M. and M.L., 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. A.M. and M.O: benefits of PhD and postdoctoral fellowships from the University of Milano-Bicocca, respectively.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2025.140113.

Data availability

Data will be made available on request.

References

[1] 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.

[2] M. De Doncker, C. De Graeve, J. Franceus, K. Beerens, V. Křen, H. Pelantová, R. Vercauteren, T. Desmet, Exploration of GH94 sequence space for enzyme discovery reveals a novel Glucosylgalactose phosphorylase specificity, ChemBioChem 22 (2021) 3319–3325, https://doi.org/10.1002/cbic.202100401.

[3] K. Mewis, N. Lenfant, V. Lombard, B. Henrissat, Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization, Appl. Environ. Microbiol. 82 (2016) 1686–1692, https://doi.org/10.1128/AEM.03453-15.

[4] W. Helbert, L. Poulet, S. Drouillard, S. Mathieu, M. Loiodice, M. Couturier, V. Lombard, N. Terrapon, J. Turchetto, R. Vincentelli, B. Henrissat, Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space, Proc. Natl. Acad. Sci. 116 (2019) 6063–6068, https://doi.org/10.1073/pnas.1815791116.

[5] M.-L. Garron, B. Henrissat, The continuing expansion of CAZymes and their families, Curr. Opin. Chem. Biol. 53 (2019) 82–87, https://doi.org/10.1016/j.cbpa.2019.08.004.

[6] A. Martínez Gascueña, H. Wu, R. Wang, C.D. Owen, P.J. Hernando, S. Monaco, M. Penner, K. Xing, G. Le Gall, R. Gardner, D. Ndeh, P.A. Urbanowicz, D.I. R. Spencer, M. Walsh, J. Angulo, N. Juge, Exploring the sequence-function space of microbial fucosidases, Commun. Chem. 7 (2024) 137, https://doi.org/10.1038/s42004-024-01212-4.

[7] B.V.H. Hornung, N. Terrapon, An objective criterion to evaluate sequence-similarity networks helps in dividing the protein family sequence space, PLoS Comput. Biol. 19 (2023) e1010881, https://doi.org/10.1371/journal.pcbi.1010881.

[8] J. Yang, H. Dang, Cloning and characterization of a novel cold-active endoglucanase establishing a new subfamily of glycosyl hydrolase family 5 from a psychrophilic deep-sea bacterium, FEMS Microbiol. Lett. 325 (2011) 71–76, https://doi.org/10.1111/j.1574-6968.2011.02413.x.

[9] G. Dalmaso, D. Ferreira, A. Vermelho, Marine extremophiles: a source of hydrolases for biotechnological applications, Mar. Drugs 13 (2015) 1925–1965, https://doi.org/10.3390/md13041925.

[10] 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.

[11] 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.

[12] A. Marchetti, M. Orlando, L. Bombardi, S. Fusco, M. Mangiagalli, M. Lotti, Evolutionary history and activity towards oligosaccharides and polysaccharides of GH3 glycosidases from an Antarctic marine bacterium, Int. J. Biol. Macromol. 275 (2024) 133449, https://doi.org/10.1016/j.ijbiomac.2024.133449.

[13] 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.

[14] R. Alvarado, S. Leiva, Agar-degrading bacteria isolated from Antarctic macroalgae, Folia Microbiol. (Praha) 62 (2017) 409–416, https://doi.org/10.1007/s12223-017-0511-1.

[15] V. Sánchez Hinojosa, J. Asenjo, S. Leiva, Agarolytic culturable bacteria associated with three antarctic subtidal macroalgae, World J. Microbiol. Biotechnol. 34 (2018) 73, https://doi.org/10.1007/s11274-018-2456-1.

[16] K.P. Ramasamy, A. Telatin, M. Mozzicafreddo, C. Miceli, S. Pucciarelli, Draft genome sequence of a new Pseudomonas sp. strain, ef1, associated with the psychrophilic Antarctic ciliate Euplotes focardii, Microbiol. Resour. Announc. 8 (2019) e00867-19, https://doi.org/10.1128/MRA.00867-19.

[17] M. Orlando, S. Pucciarelli, M. Lotti, Endolysins from Antarctic Pseudomonas display lysozyme activity at Low temperature, Mar. Drugs 18 (2020) 579, https://doi.org/10.3390/md18110579.

[18] M. Orlando, P.C.F. Buchholz, M. Lotti, J. Pleiss, The GH19 engineering database: sequence diversity, substrate scope, and evolution in glycoside hydrolase family 19, PloS One 16 (2021) e0256817, https://doi.org/10.1371/journal.pone.0256817.

[19] M. Galardini, A. Mengoni, E.G. Biondi, R. Semeraro, A. Florio, M. Bazzicalupo, A. Benedetti, S. Mocali, DuctApe: a suite for the analysis and correlation of genomic and OmniLog™ phenotype microarray data, Genomics 103 (2014) 1–10, https://doi.org/10.1016/j.ygeno.2013.11.005.

[20] 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.

[21] 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.

[22] 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.), Proteomics Protoc. Handb, Humana Press, Totowa, NJ, 2005, pp. 571–607, https://doi.org/10.1385/1-59259-890-0:571.

[23] F. Teufel, J.J. Almagro Armenteros, A.R. Johansen, M.H. Gíslason, S.I. Pihl, K. D. Tsirigos, O. Winther, S. Brunak, G. Von Heijne, H. Nielsen, SignalP 6.0 predicts all five types of signal peptides using protein language models, Nat. Biotechnol. 40 (2022) 1023–1025, https://doi.org/10.1038/s41587-021-01156-3.

[24] P. Avasthi, B.M. Bigge, F.M. Celebi, K. Cheveralls, J. Gehring, E. McGeever, G. Mishne, A. Radkov, D.A. Sun, ProteinCartography: comparing proteins with structure-based maps for interactive exploration, Annot, Mapp. Funct. Landsc. Protein Fam. Biol. (2024), https://doi.org/10.57844/ARCADIA-A5A6-1068.

[25] M. Steinegger, M. Meier, M. Mirdita, H. Vöhringer, S.J. Haunsberger, J. Söding, HH-suite3 for fast remote homology detection and deep protein annotation, BMC Bioinformatics 20 (2019) 473, https://doi.org/10.1186/s12859-019-3019-7.

[26] 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.

[27] 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.

[28] I. Barrio-Hernandez, J. Yeo, J. Jänes, M. Mirdita, C.L.M. Gilchrist, T. Wein, M. Varadi, S. Velankar, P. Beltrao, M. Steinegger, Clustering predicted structures at the scale of the known protein universe, Nature 622 (2023) 637–645, https://doi.org/10.1038/s41586-023-06510-w.

[29] L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data, Bioinformatics 28 (2012) 3150–3152, https://doi.org/10.1093/bioinformatics/bts565.

[30] R. Dong, Z. Peng, Y. Zhang, J. Yang, mTM-align: an algorithm for fast and accurate multiple protein structure alignment, Bioinformatics 34 (2018) 1719–1725, https://doi.org/10.1093/bioinformatics/btx828.

[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] 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. Chem. 4 (2012) 17, https://doi.org/10.1186/1758-2946-4-17.

[36] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open babel: an open chemical toolbox, J. Chem. 3 (2011) 33, https://doi.org/10.1186/1758-2946-3-33.

[37] W. Lu, J. Zhang, W. Huang, Z. Zhang, X. Jia, Z. Wang, L. Shi, C. Li, P.G. Wolynes, S. Zheng, DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model, Nat. Commun. 15 (2024) 1071, https://doi.org/10.1038/s41467-024-45461-2.

[38] M. Buttenschoen, G.M. Morris, C.M. Deane, PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences, Chem. Sci. 15 (2024) 3130–3139, https://doi.org/10.1039/D3SC04185A.

[39] P.R. Arantes, M.D. Polêto, C. Pedebos, R. Ligabue-Braun, Making it rain: cloud-based molecular simulations for everyone, J. Chem. Inf. Model. 61 (2021) 4852–4856, https://doi.org/10.1021/acs.jcim.1c00998.

[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] E. Akkus, O. Tayfuroglu, M. Yildiz, A. Kocak, Accurate binding free energy method from end-state MD simulations, J. Chem. Inf. Model. 62 (2022) 4095–4106, https://doi.org/10.1021/acs.jcim.2c00601.

[42] 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.

[43] R. Hu, L. Lin, T. Liu, P. Ouyang, B. He, S. Liu, Reducing sugar content in hemicellulose hydrolysate by DNS method: a revisit, J. Biobaased Mater. Bioenergy 2 (2008) 156–161, https://doi.org/10.1166/jbmb.2008.306.

[44] S. Yu, T. Su, H. Wu, S. Liu, D. Wang, T. Zhao, Z. Jin, W. Du, M.-J. Zhu, S.L. Chua, L. Yang, D. Zhu, L. Gu, L.Z. Ma, PslG, a self-produced glycosyl hydrolase, triggers biofilm disassembly by disrupting exopolysaccharide matrix, Cell Res. 25 (2015) 1352–1367, https://doi.org/10.1038/cr.2015.129.
[45] J. Zhang, H. Wu, D. Wang, L. Wang, Y. Cui, C. Zhang, K. Zhao, L. Ma, Intracellular glycosyl hydrolase PslG shapes bacterial cell fate, signaling, and the biofilm development of Pseudomonas aeruginosa, eLife 11 (2022) e72778. doi:https://doi.org/10.7554/eLife.72778.
[46] A. Ali-Ahmad, M.-L. Garron, V. Zamboni, N. Lenfant, D. Nurizzo, B. Henrissat, J.-G. Berrin, Y. Bourne, F. Vincent, Structural insights into a family 39 glycoside hydrolase from the gut symbiont Bacteroides cellulosilyticus WH2, J. Struct. Biol. 197 (2017) 227–235, https://doi.org/10.1016/j.jsb.2016.11.004.
[47] D. Wang, S.T.A. Naqvi, F. Lei, Z. Zhang, H. Yu, L.Z. Ma, Glycosyl hydrolase from Pseudomonas fluorescens inhibits the biofilm formation of pseudomonads, Biofilm 6 (2023) 100155, https://doi.org/10.1016/j.bioflm.2023.100155.
[48] K. Giles, B. Pluvinage, A.B. Boraston, Structure of a glycoside hydrolase family 50 enzyme from a subfamily that is enriched in human gut microbiome bacteroidetes, Proteins Struct. Funct. Bioinforma. 85 (2017) 182–187, https://doi.org/10.1002/prot.25189.
[49] B. Pluvinage, J.-H. Hehemann, A.B. Boraston, Substrate recognition and hydrolysis by a family 50 exo-β-Agarase, Aga50D, from the marine bacterium Saccharophagus degradans, J. Biol. Chem. 288 (2013) 28078–28088, https://doi.org/10.1074/jbc.M113.491068.
[50] B. Pluvinage, C.S. Robb, R. Jeffries, A.B. Boraston, The structure of *pf* GH50B, an agarase from the marine bacterium *Pseudoalteromonas fuliginea* PS47, Acta Crystallogr, Sect. F Struct. Biol. Commun. 76 (2020) 422–427, https://doi.org/10.1107/S2053230X20010328.
[51] M. Necci, D. Piovesan, D. Clementel, Z. Dosztányi, S.C.E. Tosatto, MobiDB-lite 3.0: fast consensus annotation of intrinsic disorder flavors in proteins, Bioinformatics 36 (2021) 5533–5534, https://doi.org/10.1093/bioinformatics/btaa1045.
[52] M.-H. Beylot, V.A. McKIE, A.G.J. Voragen, C.H.L. Doeswijk-Voragen, H.J. Gilbert, The Pseudomonas cellulosa glycoside hydrolase family 51 arabinofuranosidase exhibits wide substrate specificity, Biochem. J. 358 (2001) 607–614, https://doi.org/10.1042/bj3580607.
[53] D.-H. Im, K. Kimura, F. Hayasaka, T. Tanaka, M. Noguchi, A. Kobayashi, S. Shoda, K. Miyazaki, T. Wakagi, S. Fushinobu, Crystal structures of glycoside hydrolase family 51 α- L -Arabinofuranosidase from *Thermotoga maritima*, Biosci. Biotechnol. Biochem. 76 (2012) 423–428, https://doi.org/10.1271/bbb.110902.
[54] M. Corbella, G.P. Pinto, S.C.L. Kamerlin, Loop dynamics and the evolution of enzyme activity, Nat. Rev. Chem. 7 (2023) 536–547, https://doi.org/10.1038/s41570-023-00495-w.
[55] 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.
[56] A.S.C. Fraser, K.E. Low, J.P. Tingley, G. Reintjes, D. Thomas, H. Brumer, D. W. Abbott, SACCHARIS v2: Streamlining prediction of carbohydrate-active enzyme specificities within large datasets, in: F. Lisacek (Ed.), Protein Bioinforma, Springer US, New York, NY, 2024, pp. 299–330, https://doi.org/10.1007/978-1-0716-4007-4_16.
[57] D. Gioia, M. Bertazzo, M. Recanatini, M. Masetti, A. Cavalli, Dynamic docking: a paradigm shift in computational drug discovery, Molecules 22 (2017) 2029, https://doi.org/10.3390/molecules22112029.
[58] V. Pirillo, M. Orlando, D. Tessaro, L. Pollegioni, G. Molla, An efficient protein evolution workflow for the improvement of bacterial PET hydrolyzing enzymes, Int. J. Mol. Sci. 23 (2021) 264, https://doi.org/10.3390/ijms23010264.
[59] B. Fram, Y. Su, I. Truebridge, A.J. Riesselman, J.B. Ingraham, A. Passera, E. Napier, N.N. Thadani, S. Lim, K. Roberts, G. Kaur, M.A. Stiffler, D.S. Marks, C.D. Bahl, A. R. Khan, C. Sander, N.P. Gauthier, Simultaneous enhancement of multiple functional properties using evolution-informed protein design, Nat. Commun. 15 (2024) 5141, https://doi.org/10.1038/s41467-024-49119-x.
[60] G. Munsamy, R. Illanes-Vicioso, S. Funcillo, I.T. Nakou, S. Lindner, G. Ayres, L. S. Sheehan, S. Moss, U. Eckhard, P. Lorenz, N. Ferruz, Conditional language models enable the efficient design of proficient enzymes (2024), https://doi.org/10.1101/2024.05.03.592223.
[61] A. Lauko, S.J. Pellock, I. Anischanka, K.H. Sumida, D. Juergens, W. Ahern, A. Shida, A. Hunt, I. Kalvet, C. Norn, I.R. Humphreys, C. Jamieson, A. Kang, E. Brackenbrough, A.K. Bera, B. Sankaran, K.N. Houk, D. Baker, Computational design of serine hydrolases (2024), https://doi.org/10.1101/2024.08.29.610411.
[62] M. Braun, A. Tripp, M. Chakatok, S. Kaltenbrunner, M.G. Totaro, D. Stoll, A. Bijelic, W. Elaily, S.Y.Y. Hoch, M. Aleotti, M. Hall, G. Oberdorfer, Computational design of highly active de novo enzymes (2024), https://doi.org/10.1101/2024.08.02.606416.
[63] M. Gantz, S. Neun, E.J. Medcalf, L.D. Van Vliet, F. Hollfelder, Ultrahigh-throughput enzyme engineering and discovery in *in vitro* compartments, Chem. Rev. 123 (2023) 5571–5611, https://doi.org/10.1021/acs.chemrev.2c00910.
[64] M.A.B.D. Morais, C.C. Polo, M.N. Domingues, G.F. Persinoti, R.A.S. Pirolla, F.H. M. De Souza, J.B.D.L. Correa, C.R. Dos Santos, M.T. Murakami, Exploring the molecular basis for substrate affinity and structural stability in bacterial GH39 β-Xylosidases, Front. Bioeng. Biotechnol. 8 (2020) 419, https://doi.org/10.3389/fbioe.2020.00419.
[65] P. Zhang, J. Zhang, L. Zhang, J. Sun, Y. Li, L. Wu, J. Zhou, C. Xue, X. Mao, Structure-based design of agarase AgWH50C from Agarivorans gilvus WH0801 to enhance thermostability, Appl. Microbiol. Biotechnol. 103 (2019) 1289–1298, https://doi.org/10.1007/s00253-018-9540-1.
[66] C. Jiang, T. Zhang, Y. Xu, X. Mao, Characterization of a GH50 β-Agarase: a biotechnological tool for preparing oligosaccharides from agarose and Porphyran, J. Agric. Food Chem. 70 (2022) 9931–9940, https://doi.org/10.1021/acs.jafc.2c02107.
[67] U. Temuujin, W.-J. Chi, J.-S. Park, Y.-K. Chang, J.Y. Song, S.-K. Hong, Identification and characterization of a novel β-galactosidase from Victivallis vadensis ATCC BAA-548, an anaerobic fecal bacterium, J. Microbiol. 50 (2012) 1034–1040, https://doi.org/10.1007/s12275-012-2478-6.
[68] 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.
[69] M. Mangiagalli, S. Brocca, M. Orlando, M. Lotti, The "cold revolution", Present and future applications of cold-active enzymes and ice-binding proteins, New Biotechnol. 55 (2020) 5–11, https://doi.org/10.1016/j.nbt.2019.09.003.
[70] 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.
[71] T. Collins, G. Feller, Psychrophilic enzymes: strategies for cold-adaptation, Essays Biochem. 67 (2023) 701–713, https://doi.org/10.1042/EBC20220193.
[72] M. Zhang, J. Wang, R. Zeng, D. Wang, W. Wang, X. Tong, W. Qu, Agarose-degrading characteristics of a Deep-Sea bacterium vibrio Natriegens WPAGA4 and its cold-adapted GH50 Agarase Aga3420, Mar. Drugs 20 (2022) 692, https://doi.org/10.3390/md20110692.
[73] J. Li, Y. Sha, Expression and enzymatic characterization of a cold-adapted β-agarase from Antarctic bacterium Pseudoalteromonas sp. NJ21, Chinese J. Oceanol. Limnol. 33 (2015) 319–327, https://doi.org/10.1007/s00343-015-4072-3.
[74] T. Oikawa, T. Kazuoka, K. Soda, Paradoxical thermostable enzymes from psychrophile: molecular characterization and potentiality for biotechnological application, J. Mol. Catal. B: Enzym. 23 (2003) 65–70, https://doi.org/10.1016/S1381-1177(03)00073-0.
[75] 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.
[76] G. Bianchi, A. Pessina, D. Ami, S. Signorelli, M. de Divitiis, A. Natalello, M. Lotti, L. Brambilla, S. Brocca, M. Mangiagalli, Sustainable production of a biotechnologically relevant β-galactosidase in Escherichia coli cells using crude glycerol and cheese whey permeate, Bioresour. Technol. 406 (2024) 131063, https://doi.org/10.1016/j.biortech.2024.131063.
[77] G.J. Davies, K.S. Wilson, B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, Biochem. J. 321 (1997) 557–559, https://doi.org/10.1042/bj3210557.
[78] T. Yu, H. Cui, J.C. Li, Y. Luo, G. Jiang, H. Zhao, Enzyme function prediction using contrastive learning, Science 379 (2023) 1358–1363, https://doi.org/10.1126/science.adf2465.
[79] A. Kroll, S. Ranjan, M.J. Lercher, A multimodal transformer network for protein-small molecule interactions enhances predictions of kinase inhibition and enzyme-substrate relationships, PLoS Comput. Biol. 20 (2024) e1012100, https://doi.org/10.1371/journal.pcbi.1012100.
[80] A. Kroll, Y. Rousset, X.-P. Hu, N.A. Liebrand, M.J. Lercher, Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning, Nat. Commun. 14 (2023) 4139, https://doi.org/10.1038/s41467-023-39840-4.
[81] V.S. Boorla, C.D. Maranas, CatPred: a comprehensive framework for deep learning in vitro enzyme kinetic parameters k_{cat}, K m and K i (2024), https://doi.org/10.1101/2024.03.10.584340.
[82] T. Wang, G. Xiang, S. He, L. Su, Y. Wang, X. Yan, H. Lu, DeepEnzyme: a robust deep learning model for improved enzyme turnover number prediction by utilizing features of protein 3D-structures, Brief. Bioinform. 25 (2024) bbae409, https://doi.org/10.1093/bib/bbae409.
[83] M.J. Campbell, VIPER: A General Model for Prediction of Enzyme Substrates (2024), https://doi.org/10.1101/2024.06.21.599972.