Structural determinants of cold activity and glucose tolerance of a family 1 glycoside hydrolase (GH1) from Antarctic *Marinomonas* sp. ef1

Louise Jane Gourlay[1Ψ], Marco Mangiagalli[2Ψ], Elisabetta Moroni[3], Marina Lotti[2*], and Marco Nardini[1*].

[1]Department of Biosciences, University of Milano, Via Celoria 26, Milano, Italy.

[2]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, Milano, Italy

[3] Institute of Chemical Sciences and Technologies, National Research Council of Italy, SCITE-CNR, via Mario Bianco 9, Milano, Italy.

[*]**Corresponding Author:**

Marco Nardini, Department of Biosciences, University of Milano, I-20133, Milano, Italy; Tel: +39 02503 19894; Email: marco.nardini@unimi.it

Marina Lotti, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milano, Italy, Tel: +39 6448 3527; Email: marina.lotti@unimib.it

[Ψ]These authors contributed equally to this work

Running Title: Cold-active β-glucosidase from *Marinomonas* sp. ef1

Abbreviations: BglU, GH1 from the Antarctic bacterium *Micrococcus antarcticus*; CD, circular dichroism; GH1, glycoside hydrolase family 1; HiBG, *Humicola insolens* β-glucosidase; K_d, distribution coefficient; K_i, inhibitory constant; K_M, Michealis constant; M-GH1, *Marinomonas* sp. ef1 β-glucosidase; MD, Molecular Dynamics; MW, molecular weight; NaF, sodium fluoride; NPT, constant number of particles, constant-pressure, constant-temperature; NVT, constant number of particles, constant-volume, constant-temperature; *o*NPGal, *orto*-Nitrophenyl β-D-galactopyranoside; PB, sodium phosphate buffer; PDB, Protein Data Bank; *p*NPClb, *para*-

Nitrophenyl β-D-cellobioside; *p*NPGlu, *para*-Nitrophenyl β-D-glucopyranoside; *p*NPMan, *para*-Nitrophenyl β-D-mannopyranoside; *p*NPXyl, *para*-Nitrophenyl β-D-xylopyranoside; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; SASA, solvent-accessible surface area; SEC, size-exclusion chromatography; T_{opt}, optimal temperature of catalysis; $T_{m,}$ unfolding transition midpoint.

Keywords: cold-active enzyme, β-glucosidase, psychrophiles, crystal structure, glucose-tolerance

Conflict of Interest: The authors declare no conflict of interest.

Abstract

Cold-active enzymes support life at low temperatures due to their ability to maintain high activity in the cold and can be useful in several biotechnological applications. Although information on the mechanisms of enzyme cold adaptation is still too limited to devise general rules, it appears that very diverse structural and functional changes are exploited in different protein families and within the same family. In this context, we studied the cold adaptation mechanism and the functional properties of a member of the glycoside hydrolase family 1 (GH1) from the Antarctic bacterium *Marinomonas* sp. ef1. This enzyme exhibits all typical functional hallmarks of cold adaptation, including high catalytic activity at 5 °C, broad substrate specificity, low thermal stability, and higher lability of the active site compared to the overall structure.

Analysis of the here-reported crystal structure (1.8 Å resolution) and molecular dynamics (MD) simulations suggest that cold activity and thermolability may be due to a flexible region around the active site (residues 298–331), whereas the dynamic behavior of loops flanking the active site (residues 47–61 and 407–413) may favor enzyme-substrate interactions at the optimal temperature of catalysis (T_{opt}) byethering together protein regions lining the active site. Stapling of the N-terminus onto the surface of the β-barrel is suggested to partly counterbalance protein flexibility, thus providing a stabilizing effect. The tolerance of the enzyme to glucose and galactose is accounted for by the presence of a “gatekeeping” hydrophobic residue (Leu178), located at the entrance of the active site.

INTRODUCTION

Glycoside hydrolases (GH) build a large and heterogenous group of hydrolytic enzymes that are highly diverse in substrate specificity, structure, catalytic mechanism, and biological function. GH enzymes have been classified into 184 families to date, based on sequence and structural similarities [1]. The GH1 family includes enzymes active on both β-galactosides and β-glucosides (named β-galactosidases and β-glucosidases, respectively), which have a wide range of biotechnological applications. The GH1 enzymes share a structure organized in a single $(\alpha/\beta)_8$ TIM barrel that contains the two active site glutamic acid residues [2]. β-galactosidases (EC 3.2.1.23) catalyze the hydrolysis of lactose into glucose and galactose or the reverse transglycosylation reaction [3–5]. The catalytic versatility of β-galactosidases makes them widely employed in the food industry, including the production of lactose-free dairy products, and the synthesis of prebiotic galacto-oligosaccharides [5–8]. β-glucosidases (E.C. 3.2.1.21) catalyze the hydrolysis of cellobiose or cello-oligomers into glucose [10]. These enzymes find broad applications, from cellulose degradation, where they act in combination with endoglucanases and cellobiohydrolases, to wine flavor ripening and other reactions relevant in food processing [10]. The use of enzymes in the food industry generally requires low or moderate temperatures, to preserve thermolabile compounds and food quality [11]. To meet this constraint, a useful tool can be the so called cold-active enzymes that may be sourced from psychrophilic microorganisms that thrive in cold environments [3,12].

Although low temperatures decrease the rate of enzymatic reactions, cold-active enzymes still maintain high catalytic activities under such conditions. Furthermore, most of them display lower substrate affinities and thermal stabilities than their mesophilic and thermophilic counterparts [12–14]. The structural basis of enzyme cold adaptation has been investigated by comparing the 3D structures of cold-active enzymes with those of mesophilic and thermophilic homologues. It has emerged that cold-active enzymes display high structural plasticity/flexibility, which can be restricted to specific areas such as the enzyme active site or may regard the entire 3D structure [12,13]. Protein plasticity can be achieved by increasing

the number and extent of loop regions and cavities, modifying protein topology, reducing the compactness of the protein core and the number and strength of intramolecular interactions (H-bonds, hydrophobic interactions, salt bridges, etc.), and by peculiar amino acid compositions of the primary structure, including reduced overall number of proline and arginine residues and a higher number of glycine residues [15–18]. These general features also hold for the few psychrophilic GH1s of which the 3D structures have been solved so far. Interestingly, in the three cold active GH1 of known structure, specific and different mechanisms of cold adaptation have been exploited, ranging from the loss of a single salt bridge to changes in the oligomerization state [19–21].

Here, we report the in-depth structural and functional analysis of a GH1 from the Antarctic *Marinomonas* sp. ef1 (M-GH1). The structural determinants of M-GH1 cold adaptation and glucose tolerance were investigated by combining structural analysis with molecular dynamics (MD) and molecular docking simulations. Our results reveal that the active site and substrate binding region are overall conserved except for two elements: *i)* a dynamic region around the active site, which is likely involved in M-GH1 cold activity; *ii)* a hydrophobic residue located at the entrance of the active site, which may govern glucose tolerance.

RESULTS

M-GH1 shows hallmarks typical of a cold-active enzyme

The gene coding for M-GH1 (NCBI: WP_100188174.1) was identified in the genome of the Antarctic bacterium *Marinomonas* sp. ef1 and the corresponding protein was clasified in the GH1 family by sequence analysis. The M-GH1 construct used for activity measurements and structural work includes the full amino acid residues (1 to 447), and a C-terminal Leu-Glu-6xHis-tag (molecular weight MW 51.7 kDa). M-GH1 was cloned in pET-21a vector, recombinantly produced in *Escherichia coli* cells and purified to homogeneity by affinity chromatography with a final yield of 168 mg per L of culture.

Activity was assessed on a panel of five synthetic substrates. M-GH1 was found to be active on all of them (**Fig. 1A**), with highest activities detected with *orto*-Nitrophenyl β-D-

galactopyranoside (*o*NPGal: specific activity: 48.4 ± 2.5 U/mg), *para*-Nitrophenyl β-D-glucopyranoside (*p*NPGlu: specific activity: 34.9 ± 1.8 U/mg) and *para*-Nitrophenyl β-D-cellobioside (*p*NPClb: specific activity: 31.5 ± 1.3 Umg). Therefore, *o*NPGal was selected as the substrate for further experiments.

M-GH1 displays the highest activity at pH 6.5 and retains ~ 80% of its activity over a pH range of 6.0-7.5 (**Fig. 1B**). The enzyme has an optimal temperature of catalysis (T_{opt}) of 35 °C and maintains ~ 30% of its maximal activity at 5 °C (15.1 ± 1.3 U/mg) (**Fig. 1C**). The loss of the secondary structure triggered by heat (unfolding transition midpoint, T_m: 40.1 ± 1.6 °C) is found to occur at a temperature above the T_{opt} (**Fig. 1D**), suggesting that inactivation precedes denaturation as already reported for several cold-active enzymes. Under optimal catalysis conditions, M-GH1 follows Michaelis-Menten kinetics with both *o*NPGal and *p*NPGlu substrates with a K_M of 5.2 ± 0.4 mM and 0.2 ± 0.1 mM, respectively (**Table 1**).

The long-term temperature stability of M-GH1 was investigated at three different temperatures (5 °C, 25 °C and 35 °C), monitoring changes in residual catalytic activity and in structure over time. At 5 °C, the enzyme is functionally stable for 21 days (**Fig. 2A**), whereas it is completely inactive at 25 °C and 35 °C after 56 h and 5 h, respectively (**Fig. 2B, C**). Changes in secondary structure in the same samples were determined by circular dichroism (CD) spectroscopy by collecting spectra at the beginning and the end of incubation (**Figs. 2D-F)**: black and red lines, respectively). At time 0, M-GH1 presents the typical CD spectrum of α/β structures with two minimum peaks at 222 nm and at 210 nm. After 21 days at 5 °C, the CD spectrum of M-GH1 is still superimposable to that measured at time 0, indicating that this enzyme is structurally stable at low temperatures (**Fig. 2D**). At 25 °C, small changes are observed after 8 h incubation, when the enzyme is still almost fully active (**Fig. 2B**) while after 56 h, when the enzyme activity is negligible (residual activity: 3.1 ± 0.3%), the flattening of CD signal and the formation of a visible precipitate (**Fig. 2E**) is evident. Taken together, results suggest that at 25 °C the enzyme remains active for a long period of time and its slow inactivation is due to protein aggregation. At 35 °C, after 8 h incubation, while enzyme activity is almost zero, the secondary structure assessed by CD spectroscopy (blue line in **Fig. 2F**) is not affected. The onset of aggregation is observed only after 24 h (red line in **Fig. 2F**). Overall, these

data suggest that inactivation is due to the instability of a local region, presumably the catalytic center, and not to aggregation (**Fig. 2F**).

In conclusion, M-GH1 is a cold-active enzyme exhibiting all hallmarks of cold adaptation, namely high catalytic activity at 5 °C, broad substrate specificity, low thermal stability and high lability of its active site compared to the overall structure [13].

M-GH1 adopts the typical fold of GH1s

To correlate M-GH1 function to its structural architecture, we solved its crystal structure at 1.8 Å resolution (**Fig. 3**). Crystals belonged to the orthorhombic $P2_12_12$ space group, with two molecules in the asymmetric unit and a Matthews coefficient of 2.4 $Å^3$/Da (estimated solvent content of 48.8%). Analysis of protein-protein interface made with PISA [22], combined with size-exclusion chromatography (SEC), revealed that M-GH1 is a monomer (**Fig. 3A,B**; estimated MW: 49.5 ± 2.2 kDa). The M-GH1 structure was refined at 1.8 Å resolution, to final R and R_{free} values of 0.17 and 0.19, respectively, with ideal stereochemical parameters (Table 2). Overall, the electron density map was well-defined and covered residues 2 to 446 (chain A) or 445 (chain B) out of 447 residues, with no gaps. The electron density map did not show the C-terminal Leu-Glu-6xHis-tag.

The M-GH1 3D structure displays the canonical TIM barrel $(\beta/\alpha)_8$ fold, consisting of a central eight-stranded β-barrel (β1-5, β7, β10, β11) flanked by several α-helices (α1-15) and additional β-strands (β6, β8-9, β12-13) (**Fig. 3B**, **Fig. 4**). Additional 3_{10} helices (η1-5) are present in connecting loops. Structural alignment carried out with DALI retrieves three enzyme discovered in metagenomic studies, BglM-G1 (PDB: 5NS6, sequence identity: 45.5% [19]), MeBglD2 (PDB: 5XGZ, sequence identity: 47.7% [22]), and Td2F2 (PDB: 3WH5, sequence identity: 43.4% [23]), as the top structural homologs [24]. The overall conservation among these enzymes is demonstrated by a root-mean-square deviation (RMSD) within 1.1 Å (**Fig. 3C**). The most conserved region corresponds to the protein core, while the main differences concern the N-terminal region, the solvent-exposed loop region 298-331 strands β8 and β9, and some other flexible regions (in particular, residues 92-96, 266-281 and 317-322) (**Fig. 3D**). To note, M-GH1 is the only structure in which the N-terminal region folds back onto the surface of the β-barrel, inserting itself between helices α14 and α15 ("stapling" effect), thanks to a backbone kink introduced by

Pro6 (**Fig. 3E,F**). This conformation is stabilized by H-bonds between Thr4 and Asp379, Thr4 and Asn383, and by hydrophobic interactions between Ile3 and Leu5 with residues of the following loop Met11, Leu12, Phe18, of helix α12 Val375, Leu378, Ile382, and of helix α13 Ala437, Leu441 (**Fig. 3F**). Concerning loop 298-331, which is located at the entrance to the active site cavity, the main structural divergence is in a short, Pro-rich region in the homologs identified with DALI (corresponding to M-GH1 residues 307-312). Notably, in M-GH1 this region is devoid of Pro residues, and therefore it is less rigid than in the reference proteins (**Fig. 4**).

The architecture of the active site is conserved among M-GH1 homologs

GH1s display the typical retaining mechanism of catalysis and their active site pocket can be divided into three subsites: the -1 glycone subsite, and the +1 and +2 aglycone subsites [25]. The M-GH1 active site is located at the entrance of the central β-barrel, creating a solvent-accessible pocket with a molecular surface area of 541 $Å^2$ and a volume of 356 $Å^3$ (**Fig. 5A,B**), as calculated with CAST-P [26]. The pocket is lined by atoms belonging to 50 residues; the catalytic residues of M-GH1 are located deep inside the active site pocket, including Glu171 (loop 168-172) and Glu355 (β10), which act as the acid/base and nucleophile, respectively (**Fig. 5A,B** and **Fig. 4**). The active site of M-GH1 contains a Tris molecule, used as the buffer in the crystallization solution. The side chains of catalytic residues Glu171 and Glu355 form hydrogen bonds with the amino group of Tris. Glu171 forms an additional hydrogen bond with a hydroxyl group of Tris *via* its carboxyl group. The other two hydroxyl groups of Tris are hydrogen-bonded by the carboxyl group of Glu408. Two water molecules further contribute to the protein-Tris interaction in the active site. In other GH1 family members, Tris has been described as a good substrate mimetic because it is rich in hydroxyl groups [27,28]. In addition, in M-GH1 Tris perfectly matches the position of ligands bound in the active site (-1 subsite) of the highly structurally similar Td2F2 (glycerol, PDB: 3WH5; D-glucose, PDB: 3WH6; isofagomine, PDB: 3WH7; D-fucose, PDB: 3WH8) [23]) (**Fig. 5C**). All residues involved in sugar recognition at subsite -1 in Td2F2 are conserved in M-GH1, except for Arg325 which is substituted with Trp328 in M-GH1 (**Fig. 4**).

To investigate the structural basis of M-GH1 substrate specificity, for comparison with our biochemical experimental data, we performed *in silico* docking experiments that simulate M-

GH1 binding to all tested substrates (i.e., *o*NPGal *p*NPGlu, *p*NPClb, *para*-Nitrophenyl β-D-xylopyranoside - *p*NPXyl -, and *para*-Nitrophenyl β-D-mannopyranoside - *p*NPMan). For the two best substrates *o*NPGal and *p*NPGlu, the substrate binding residues were identified as Ala23, Gln26, His126, Trp127, Asn170, Glu171, Leu178, His185, Cys174, Asn227, Asn229, Asn297, Tyr299, Trp328, Glu355, Trp401, Glu408, Trp409, and Phe417. In particular, M-GH1 residues Glu408, Trp409, Gln26 and Asn170 contribute to the correct positioning of the sugar moieties of the chromogenic substrates through a network of hydrogen bond interactions (**Fig. 6A,B**). Differences in the binding modes between the two best substrates appear to be mainly due to the positioning of the nitro-phenyl moiety, which points towards the pocket entrance in *p*NPGlu, whereas it forms hydrogen bonds with pocket residues Asn227 and Asn229 in *o*NPGal. T-shape stacking of the nitro-phenyl moiety is established with the side chain of Trp328 for both ligands. The binding free energies for *o*NPGal and *p*NPGlu were predicted to be -8.7 ± 0.1 kcal/mol and -8.1 ± 0.1 kcal/mol, respectively (**Table 3**). As for the other substrates (i.e., *p*NPClb, *p*NPXyl, *p*NPMan) a slightly different position of the sugar moieties, compared to that observed for oNPGal and pNPGlu, involve hydrogen bond interactions with residues Glu408, Gln26, Asn170, and His126 instead of Trp409. Since pNPClb is more extended than the other ligands, the nitrophenyl group, besides establishing a T-shaped stacking interaction with the side chain of Trp328, can also reach the charged guanidinium group of Arg247 (**Fig. 6C,D,E**). The binding free energies for *p*NPClb, *p*NPXyl, *p*NPMan were predicted to be -8 ± 1 kcal/mol, -7.4 ± 0.3 kcal/mol and -6.4 ± 0.2 kcal/mol, respectively (**Table 3**). The different binding modes towards different substrates reflect the substrate specificity reported in **Fig. 1A**.

Cold activity of M-GH1 is achieved by high flexibility of residues 298-331

MD simulations, performed at 5 °C and 35 °C, show that at both temperatures M-GH1 is overall stable over the simulated time, undergoing limited conformational changes, as indicated by all-atom RMSD (1.2-1.8 Å at 5 °C and 1.3-2.1 Å at 35 °C, **Fig. 7A**) and by the analysis of the solvent-accessible surface area of the overall structure, including the active site residues (SASA: 18600 ± 200 $Å^2$ at 5 °C and SASA: 18700 ± 200 $Å^2$ at 35 °C). The most flexible residues were identified by root-mean-square fluctuation (RMSF) analyses. Overall, RMSF profiles are qualitatively similar at

both simulated temperatures, with residues 298-331 constituting the most flexible protein region (maximum RMSF value of 2.2 and 2.1 Å at 5 °C and 35 °C, respectively) (**Fig. 7B**).

Covariance analysis on the C^{α} carbons of the enzyme at 5 °C indicates that the first eigenvector is associated to the motion of the residues 298-331 (**Fig. 7C**), whose dynamic behaviour seems to be correlated to the enhanced flexibility of the adjacent helix α10 (residues 240-250) (**Fig. 7C**). Residues 298-331 form a loop (including β-strands 8 and 9) located at the entrance of the active site **(Fig. 7D)**. Its high flexibility suggests a role in cold-activity, in keeping with the notion that some cold-active enzymes display high structural plasticity near the catalytic site to support enzymatic activity at low temperature [5]. Interestingly, in GH1 from the Antarctic bacterium *Micrococcus antarcticus* (BglU, PDB: 3W53, sequence identity: 42.4%) (**Fig. 4**), a cold-adaptation mechanism based on the plasticity of the loop topologically equivalent to M-GH1 298-331 (called L3 in BglU) has been proposed [21]. Although the two proteins share 42.4% sequence identity and high overall structural similarity (RMSD ~1.1 Å), the two equivalent loops differ in sequence, length, and conformation (due to high number of prolines in L3), indicating differences in the molecular mechanism of cold adaptation (**Fig. 8**).

Covariance analysis on the C^{α} carbons of the enzyme at 35 °C indicates that the first eigenvector is mainly associated with the motions of loop 47-61, a region hosting helix α8 (residues: 170-184), helix11 (residue 266-275), loop 298-331, and loop 407-413 (**Fig. 7C**). Remarkably, loop 47-61 is the most flexible region at 35 °C (RMSF 2.2 Å), and its dynamics seems to affect the mobility of residues Glu408 and Trp409, located in an adjacent protein region (**Fig. 7D**), which play a key role in positioning the sugar moieties of *o*NPGal and *p*NPGlc substrates **(Fig. 6A,B)**. Thus, the substrate promiscuity of M-GH1 may result from the higher flexibility of this region. The mobility of region 298-331 together with helix α10 (residues 240-250) is also confirmed at 35 °C, being associated to the second eigenvector, although with lesser extent relative to 5 °C (**Fig. 7C**).

In conclusion, MD simulations suggest that the high flexibility of region 298-331 at 5 °C might confer cold activity, while the dynamic behavior of loops 47-61 and 407-413 may optimize enzyme-substrate interactions at 35 °C (T_{opt}), tethering together a region that includes helix α8 (residues: 170-184), residues 298-331 and 409-413 (**Fig. 7C,D**).

M-GH1 glucose-tolerance is due to a "gatekeeping" hydrophobic residue.

GH1s can be divided into four classes based on their tolerance to glucose: I) low tolerant β-glucosidase (inhibitory constant $K_i \leq 0.1$ M glucose), II) tolerant β-glucosidase ($K_i \geq 0.1$ M glucose), III) β-glucosidase stimulated by low glucose concentrations and inhibited by high concentrations; and IV) extremely tolerant β-glucosidase ($K_i \geq 1$ M glucose) [29]. We, therefore, investigated the effects of glucose and galactose (the hydrolytic products obtained from β-glucosidase and β-galactosidase activities) on M-GH1 activity. In the presence of *p*NPGlu as a substrate, glucose and galactose act as competitive inhibitors of M-GH1 with a K_i of 161 ± 32 mM and 251 ± 27 mM, respectively (**Fig. 9A**,**B** and **Table 1**). Similar behavior was observed in the presence of *o*NPGal as a substrate with a K_i of 351 ± 49 mM and a K_i of 563 ± 58 mM with glucose and galactose, respectively (**Fig. 9C**,**D** and **Table 1**). On the other hand, no sugar-induced activation was observed (**Fig. 9**), suggesting that M-GH1 can be classified as a glucose-tolerant β-glucosidase (class II).

Although the mechanism of glucose stimulation/inhibition remains elusive, structural analyses and literature data on glucose-tolerant GH1s allowed us to identify the residues likely involved in M-GH1 glucose tolerance [30]. In particular, we focused on the comparison of M-GH1 with the glucose-stimulated β-glucosidase from *H. insolens* (HiBG, PDB: 4MDP [31]), which share 39.9% of sequence identity and reasonable structural homology (RMSD ~1.7 Å over 419 aligned residues, **Fig. 10A**). In HiBG, glucose tolerance has been ascribed to the presence of large hydrophobic residues at the subsite +2 (**Fig. 10 A**, green sticks) which act by narrowing the tunnel that leads to the catalytic acidic residues and, consequently, limit product inhibition due to hindering of glucose entry to subsite -1. In particular, HiBG Trp168 and Leu173 act as the "gatekeepers" of the narrow tunnel at the aglycne-binding site (**Fig. 10B**, left panel) [32]. In M-GH1, the aglycone-binding site region superimposes well with that of HiBG (**Fig. 10A**), apart from the loop between β8 and β9 which is shorter in M-GH1 (**Fig. 4**). The gatekeeping residues that narrow the tunnel between the aglycone-binding site and the active site are conserved as Phe173 (instead of Trp) and Leu178 (**Fig. 4** and **Fig. 10B**, middle panel), although the tunnel is about 1 Å larger in M-GH1 relative to HiBG at its narrowest point, measured in the crystal structures (**Fig. 10B**, left and middle panels). Considering proteins as dynamic entities where the crystal structure represents only one possible static conformation, we analyzed the protein's dynamics to investigate whether other accessible M-GH1 conformations align with the hypothesis presented for HiBG [32]. In particular,

the gatekeeping role of M-GH1 Phe173 and Leu178 in glucose tolerance was analyzed by monitoring their interactions with adjacent residues during MD simulations at 35 °C. Such analysis allows the identification of a network of hydrophobic interactions involving Leu178, Ile326, Trp328, Trp409 and Ala410, tethering together a region that includes helix α8 (residues: 170-184), residues 298-331 and 409-413, thus narrowing the tunnel towards the active site (**Fig. 10B**, right panel). It is important to point out that this analysis concerns the dynamics of these interactions, meaning that not all observed interactions are present simultaneously in every protein conformation sampled during the dynamics. Instead, they continuously form and break. More specifically, the lifetime analysis of these hydrophobic interactions, which were not observed in the static crystal structure, indicates that they are persistent for 57%, 28.3%, and 17% of the simulation time for Ile326-Trp409/Ala410, Leu178-Trp328, and Leu178-Trp409/Ala410, respectively. In contrast, the position of Phe173 during the simulations does not change and its interaction network with adjacent hydrophobic side-chains was found unaltered. Overall, our structural and MD results suggest that the tunnel that leads from the aglycone-binding site to the catalytic acidic residues is narrowed in M-GH1 as in HiBG and that only Leu178 acts as a dynamic gatekeeping residue regulating substrate access to active site, while Phe173 has a more static role.

DISCUSSION

M-GH1 is a *bona fide* cold-active enzyme characterized by high activity at low temperatures, low thermal stability, and a heat labile active site [12]. Although knowledge of cold adaptation mechanisms of GH1s is still limited, three different strategies have been described so far: *i*) presence of a long and flexible loop involved in substrate binding, as observed in BglU [21]; *ii*) loss of a salt bridge, in BglM-G1 [19]; iii) tetramerization, as observed in B7 β-glucosidase form *Exiguobacterium antarcticum* [20]. Compared to these enzymes, M-GH1 displays two specific structural features: an overall high flexibility, particularly at the active site region residues 298-331 and the stapling of the N-terminal region, which contribute to the stabilization of the monomeric structure of the enzyme. Sequence analysis reveals the topological similarity of this flexible M-GH1 region with loop L3 in BglU, suggesting a similar mechanism of cold adaptation but achieved by two different evolutionary routes: the absence of “rigid” Pro residues in the M-GH1 loop, and loop

elongation to counteract the Pro richness in BglU (**Fig. 8**). On the other hand, the N-terminal stapling found in M-GH1 can provide additional intra-molecular interactions resulting in a fine-tuned trade-off between flexibility and stability (**Fig. 3E,F**). This is relevant in M-GH1 that is monomeric, whereas often in psychrophilic enzymes stabilization is achieved through inter-molecular interactions, resulting in oligomerization [16,20,32]. Overall, these pieces of evidence support the idea that cold adaptation is driven by the thermodynamic requirements of cold activity but can be achieved by different structural signatures, depending on the specific protein [14,33].

M-GH1 has broad substrate specificity with β-galactosidase activity being the prominent one. This behavior is uncommon in other GH1 members, in which usually the β-glucosidase activity is higher than the β-galactosidase one [34–38]. We hypothesize that this property is due to the plasticity of a region close to the active site involved in the recognition of sugar moieties (*i.e.* loop 47-61, loop 298-331 and loop 407-413), as shown by MD simulations. The higher affinity of M-GH1 to glucose than to galactose leads to a more stable enzyme-substrate complex, which is reflected in a lower k_{cat} towards glucose compared with galactose.

In addition to its cold-activity, tolerance to glucose and galactose make M-GH1 attractive for biotechnology exploitation. The activity from low to medium temperatures implies that M-GH1 can be used in the transformation of thermolabile compounds [11,14,39], while glucose tolerance allows its use in reactions that generate high glucose concentration, such as the conversion of cellobiose [29]. Although the molecular mechanisms of glucose tolerance in GH1 enzymes are still elusive, it is plausible that this feature depends on the shape and the electrostatic properties of active site entrance [29]. In glucose-stimulated HiBG, two hydrophobic residues (Trp168, Leu173) act as gatekeeping residues for the active site and reduce the size of subsite +2, thus limiting glucose entry to subsite -1 [31]. Sequence analysis indicates that the main difference between M-GH1 and HiBG is the conservative substitution of the gatekeeping Trp with a Phe (Phe173), while the Leu residue (Leu178) is conserved in M-GH1. Accordingly, oth residues contribute- to narrow the tunnel between subsite +2 and the active site in M-GH1, and the MD simulations indicate that the Leu178 residue plays a major dynamic role in regulating substrate access to active site, thus increasing glucose tolerance. This observation aligns with the hypothesis proposed for HiBG [32] which correlates glucose tolerance with a restricted access to the active site.

In conclusion, M-GH1 combines cold activity with glucose tolerance, making it exploitable in various biotechnological applications. Furthermore, our work points out that the same mechanism of cold adaptation can be pursued through alternative structural solution, which are likely related to specific evolutionary history of each enzyme.

Materials and Methods

Recombinant production of M-GH1

The sequence (NCBI:WP_100188174.1), encoding for a putative β-galactosidase (M-GH1), was identified from the genome of *Marinomonas* sp. ef1 using Prokka 1.12 [40]. Sequence analyses were performed using Interpro software [41,42]. The full-length sequence of M-GH1, encoding for residues 1-447, was codon-optimized for the expression in *E. coli* cells, synthetized and cloned in frame with a C-terminal 6xHis tag in the pET-21a vector (Novagen), using *Nde*I and *Xho*I restriction sites (Genscript, Piscataway, NJ, USA). The C-terminal Leu-Glu-6xHis-tag contains the residues encoded by the *Xho*I restriction site (Leu, Glu) and 6xHis tag.

Recombinant M-GH1 was produced in Zym-5052 medium [43] supplemented with 100 mg/L ampicillin. Cells were harvested by centrifugation at 4 °C for 10 min at 5000 *g*, resuspended in lysis buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl and 10 mM imidazole) and lysed using a cell disruptor (Constant Systems Ltd, Daventry, UK) at 172.4 MPa. Recombinant M-GH1 was purified from clarified crude extract (20 min at 6000 *g* at 4 °C) by metal ion affinity chromatography on a nickel-nitrilotriacetic acid agarose resin (Avantor, Radnor, Pennsylvania, US). Elution was carried out with an elution buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl and 10 mM imidazole). Fractions containing the highest protein concentration were pooled and buffer-exchanged using a PD10 Desalting column (GE healthcare, Little Chalfont, UK), pre-equilibrated with 10 mM sodium phosphate buffer (PB) pH 7.0 (assays) or 10 mM Tris-HCl pH 7.0 (crystallization). The protein concentration was determined using the Bradford protein assay (Bio-Rad, California, USA), using bovine serum albumin as a standard, according to the manufacturer's protocols.

SEC analysis

The quaternary structure of M-GH1 was determined by SEC using the NGC Quest 10 Plus Chromatography System (Bio-Rad, California, USA), equipped with a Superdex 75 10/30 column (GE Healthcare, Little Chalfont, UK) with a MW cutoff of 3–75 kDa, following a protocol described earlier [44]. The calibration curve for MW determination was made with the following standard: bovine serum albumin (MW 66.5 kDa), ovalbumin (MW 43 kDa), lipase B of *Candida antarctica* (MW 34.7 kDa), green fluorescence protein (MW 27.5 kDa) and cytochrome C (horse heart, MW 12.4 kDa). For each standard protein the distribution coefficient (K_d) was calculated using the following equation:

where V_E is the elution volume, V_0 the void volume (8.2 mL), which is determined with blue dextran (MW 2000 kDa) and V_T the total volume, determined with Uracil (MW 0.112 kDa). The calibration curve Log(MW) *vs*. K_d was built and the interpolated linear equation used to calculate M-GH1 molecular weight from their K_d values.

M-GH1 activity assays

M-GH1 substrate specificity was assessed for several synthetic substrates purchased from Merck (Merck, Darmstadt, Germany): *o*NPGal, *p*NPGlu, *p*NPClb, *p*NPXyl and *p*NPMan. Assays were performed at 25 °C in 10 mM PB. After 3 min, the reaction was stopped by adding an equal volume of 1 M sodium carbonate buffer, pH 11.0, and the absorbance was measured at a wavelength of 420 nm for *o*NPGal (molar extinction coefficient: 4.6 $mM^{-1} \cdot cm^{-1}$) or 405 nm for the other substrates (molar extinction coefficient: 18.6 $mM^{-1} \cdot cm^{-1}$) using a Jasco V-770 UV/NIR spectrophotometer (JASCO Europe, Lecco, Italy). All the reactions described in this section were initiated upon addition of 0.1 µg of enzyme to 1 mL of reaction. One unit of enzyme activity was defined as the amount of enzyme required to catalyze the formation of 1 µmol per min of *o*- or *p*-Nitrophenol. Experiments were performed in quadruplicate.

The optimum pH and temperatures for M-GH1 activity were determined using *o*NPGal as a substrate, spanning a pH range of 4-10 at 35 °C and a temperature range of 5 °C to 85 °C, at pH 7, respectively. The maximum activity was expressed as 100% for normalization.

Long-term thermal stability of M-GH1 was explored by measuring its residual activity towards *o*NPGal, at 35 °C after incubating the enzyme (0.25 mg/mL) in 10 mM PB, pH 7.0 at 5 °C for 21 days, 25 °C for 56 h and 35 °C for 8 h. Initial activity was given as 100% for normalization.

The kinetic parameters of M-GH1 were determined using different concentrations of *o*NPGal and *p*NPGlu in PB at optimum pH and temperature, in the absence and in the presence of 0.5 M, 1 M or 2 M of glucose or galactose. Kinetics parameters were calculated using the OriginLab software (OriginLab Corporation, Northampton, USA) and nonlinear fitting of the Michaelis-Menten equation.

CD spectroscopy

The unfolding temperature of M-GH1 was determined by measuring the variation of circular dichroism (CD) signal at 208 nm as a function of temperature, ranging from 5 °C to 90 °C, with a JASCO J-815 spectropolarimeter (JASCO Europe, Lecco, Italy), equipped with a Peltier thermoregulation system, using a 0.1-cm pathlength quartz cuvette. Measurements were performed in triplicate with a data pitch of 1 °C and a temperature slope of 0.5 °C/min. The fraction of folded protein was calculated using the following equation:

where θ_T is the molar ellipticity at 208 nm recorded at a given temperature, $\theta_{90°C}$ is the molar ellipticity at 208 nm recorded at 90 °C and $\theta_{5°C}$ is the molar ellipticity at 208 nm recorded at 5 °C. Experiments were performed in triplicate.

Long-term thermostability was assessed by measuring the CD signal at 208 nm after incubating the enzyme in PB, pH 6.0 at 4 °C for 23 days, 25 °C for 3 days and 35 °C for 8 h. CD spectra were collected in the 190 - 260 nm range at the beginning and at the end of the experiments. Measurements were carried out with 0.2-nm data pitch and 20-nm/min scanning speed. All spectra were corrected for buffer contribution, averaged from two independent acquisitions, and smoothed using a third order, least squares polynomial fit.

Crystallization of M-GH1

Purified M-GH1 was crystallized using an Oryx4 crystallization robot (Douglas Instruments). Crystallization screens were performed at 20 °C, mixing 12.4 mg/mL M-GH1 with crystallization solution, with protein volumes of 30%, 50% and 70% in the final drop volume of 400 nL. Drops were deposited in 96-well flat-bottomed CrystalQuick™ sitting drop plates (Greiner Bio-One) containing 100 µL of each of the 96 PACT Premier screen conditions (Molecular Dimensions, Ltd). M-GH1 crystals grew over 2-4 days in 50% protein drop in an optimized condition of PACT condition G1, containing 22% *(w/v)* PEG 3500, 0.1 M sodium fluoride (NaF) and 0.1 M Tris-HCl pH 8.5. Crystals were briefly soaked in cryoprotectant solution (30% *(v/v)* ethylene glycol, 30% *(w/v)* PEG3350, 0.1 M NaF, 0.1 M 0.1 M Tris-HCl pH 8.5 before cryo-cooling in liquid nitrogen for data collection.

X-ray data collection and M-GH1 structure determination

X-ray diffraction data were collected on a single orthorhombic M-GH1 crystal at 100K on the ID30A-3 beamline at the European Synchrotron Radiation Facility (ESRF, France). Reflections were reduced using XDS [45] and scaled with Aimless [46] from the CCP4 suite [47]. The structure was solved by molecular replacement using PHASER [46] and a lower resolution, unpublished M-GH1 crystal structure as a search model. The low-resolution M-GH1 structure was in turn obtained by molecular replacement using the ancestral glycosidase (family 1) as a search model (PDB-code 6Z1H [48]), after trimming non-identical residues to Ala. The initial molecular replacement solution was subjected to subsequent cycles of manual building in Coot [49] and refinement with phenix.refine [50].

The geometry of the final model was checked and validated using Molprobity [51]. Structure solution and refinement statistics are reported in **Table 2**. Structure factors and atomic coordinates have been deposited in the PDB (www.rcsb.org) under entry code 8PUO.

MD simulations

MD simulations were carried out using the AMBER software package (version 18) [52,53]; with the ff14SB force field [54], using its GPU-accelerated *pmemd.cuda* utility during equilibration and production, and *sander* otherwise [55]. Simulations employed the experimental crystal structure

of M-GH1; residues were modeled in their standard protonation states at physiological pH, as predicted by PROPKA, version 3.2 [56]: this resulted in Glu171 side chain being protonated and His residues 8, 126, 132, 185, 200, 206, 303, 343, 435 being protonated on Nε2, and His 64,65, 201, 381 being protonated on Nδ1. Hydrogen atoms were added using the *tleap* utility in AmberTools (version 19) [52].

The solute was explicitly solvated in a triclinic box with 44871 TIP3P water molecules [56,57], keeping a minimum distance of 1.1 nm between the solute and the edges of the solvent box, and then rendered electroneutral by the addition of sodium counterions. 100 mM NaCl was added. Crystallographic waters were taken from the crystal strcture of M-GH1. The global system consisted of 141706 atoms.

To remove any bad contacts between solute and solvent, the system was minimized with position restraints on the solute coordinates except for hydrogen atoms, with 150 steps of steepest descent followed by 150 steps of conjugate gradient. A cutoff of 10 Å was used for Coulomb and van der Waals interactions. The whole system was then minimized with 150 steps of steepest descent followed by 150 steps of conjugate gradient without restraints.

The temperature of the system was then increased to 278 K or 308 K in the NVT (constant number of particles, constant-volume, constant-temperature) ensemble running 20 ps of MD with weak positional restraints on the on C^{α} atoms of the protein with the Langevin thermostat [57] (harmonic restraints on C^{α} atoms: k = 5.0 kcal mol^{-1} $Å^{-2}$, collision frequency 0.75 ps^{-1}, 2 fs time-step).

The systems were then equilibrated at 278 K and 308 K for 5 ns with a 2 fs time step in periodic boundary conditions in the NPT (constant number of particles, constant-pressure, constant-temperature) ensemble to set the pressure of the system to 10^5 Pa. The electrostatic interactions were treated using the particle mesh Ewald method with a cutoff of 10 Å [58]. The same cutoff was used even for short-range Lennard-Jones interactions. Bonds involving hydrogen atoms were constrained with the SHAKE algorithm [59].

Three independent MD replicas (different random seeds) were carried out for M-GH1 at 278 K and 308 K. Long-range electrostatic interactions were treated by particle mesh Ewald method [58] with a cutoff distance for the Van der Waals and Coulomb interactions of 10 Å. Each replica's production stage was 500 ns in length, with a 2 fs time-step with SHAKE algorithm [59] in

the NVT ensemble (with a temperature of 278 K and 308 K enforced *via* Langevin's thermostat [57]; collision frequency 1 ps^{-1}, and a 1 atm pressure).

Analyses of MD trajectories were carried out with the *cpptraj* program distributed within the AmberTools suite (version 19) [52] or with code written in-house.

Docking calculations

*o*NPGal and *p*NPGlu, *p*NPClb, *p*NPXyl, *p*NPMan were docked with the Glide docking program (Schrödinger Release 2022-3: Glide, Schrödinger, LLC, New York, NY, 2021) on the crystal structure of M-GH1 [60].

The Protein Preparation Workflow in Maestro (Schrödinger Release 2022-3: Maestro, Schrödinger, LLC, New York, NY, 2021) was used to prepare the receptor protein for docking calculations. The hydrogen bonding network was optimized with default settings. An all-atom restrained minimization was performed on the structure with a termination criterion, based on the RMSD of 0.3 Å on the heavy atoms relative to their initial location, with the OPLS_2005 force field [61].

Ligand preparation was carried out by processing the 3D structures of all ligands with the LigPrep module of Schrodinger (Schrödinger Release 2022-3: LigPrep, Schrödinger, LLC, New York, NY, 2021). The minimization step was performed using the OPLS_2005 force field [61].

The grid which represents the active site of the receptor protein for Glide ligand docking calculations was generated on the prepared M-GH1 structure. The position of the grid was determined via pairwise structure alignment with the hydrolase SghA (PDB: 6RJO [62]) with which M-GH1 shares 47% sequence identity (backbone atom RMSD of 1.6 Å). SghA was crystalized in complex with 2-(hydroxymethyl)phenyl β-D-glucopyranoside, therefore the position of the ligand resulting from structure alignment was used to center the grid on the M-GH1 binding site. No modifications were applied to default settings.

Rigid receptor protein and flexible ligand docking calculations were performed in standard precision mode with the OPLS_2005 force field [61]; adjustment of the receptor protein to the presence of the ligand was explored with the Induced Fit Docking protocol (standard mode) (Schrödinger Release 2022-3**:** Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2021; Prime, Schrödinger, LLC, New York, NY, 2021); nonplanar conformations of

amide bonds were penalized, van der Waals radii were scaled by 0.80, and the partial charge cutoff was fixed to 0.15. No further modifications were applied to default settings.

Statistical analyses

Unless otherwise indicated, experiments were performed in triplicate. Statistical analyses were performed using OriginLab software (OriginLab Corporation, Northampton, USA). p-values were determined using an unpaired two-tailed t-test.

Author Contributions

L.J.G., M.M. and E.M. planned and performed the experiments; L.J.G., M.M. E.M., M.L. and M.N. analyzed the data; L.J.G., M.M. E.M., M.L. and M.N. wrote the paper.

Acknowledgements

L.J.G acknowledges "Linea 3 – SEED" funding for the SMART project from the Università degli Studi di Milano. This work was partly supported by Fondo di Ateneo of the University of Milano-Bicocca (2023-ATE-0015, M.M. and 2018-ATE-0320, M.L.). Financial support from the Italian Ministry of University and Research (MUR) through grant "Dipartimenti di Eccellenza 2017" to University of Milano Bicocca, Department of Biotechnology and Biosciences, and through grant "Dipartimenti di Eccellenza 2017" and "Dipartimenti di Eccellenza 2023" to University of Milano, Department of Biosciences, is also acknowledged.

Data availability statement

The structural data that support these findings are openly available in the wwPDB at https://doi.org/10.2210/pdb8PUO/pdb.

References

1 Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B & Terrapon N (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* **50**, D571–D577.

2 Marana S (2006) Molecular basis of substrate specificity in family 1 glycoside hydrolases. *IUBMB Life* **58**, 63–73.

3 Bojarová P & Křen V (2009) Glycosidases: a key to tailored carbohydrates. *Trends Biotechnol* **27**, 199–209.

4 Husain Q (2010) Beta galactosidases and their potential applications: a review. *Crit Rev Biotechnol* **30**, 41–62.
5 Xavier JR, Ramana KV & Sharma RK (2018) β-galactosidase: Biotechnological applications in food processing. *J Food Biochem* **42**, e12564.
6 Lu L, Guo L, Wang K, Liu Y & Xiao M (2020) β-Galactosidases: A great tool for synthesizing galactose-containing carbohydrates. *Biotechnol Adv* **39**, 107465.
7 Torres DPM, Gonçalves M do PF, Teixeira JA & Rodrigues LR (2010) Galacto-Oligosaccharides: Production, Properties, Applications, and Significance as Prebiotics. *CRFSFS* **9**, 438–454.
8 Duan F, Sun T, Zhang J, Wang K, Wen Y & Lu L (2022) Recent innovations in immobilization of β-galactosidases for industrial and therapeutic applications. *Biotechnol Adv* **61**, 108053.
9 Ketudat Cairns JR & Esen A (2010) β-Glucosidases. *Cell Mol Life Sci* **67**, 3389–3405.
10 Wang Y, Zhang C, Li J & Xu Y (2013) Different influences of β-glucosidases on volatile compounds and anthocyanins of Cabernet Gernischt and possible reason. *Food Chem* **140**, 245–254.
11 Mangiagalli M, Brocca S, Orlando M & Lotti M (2020) The "cold revolution". Present and future applications of cold-active enzymes and ice-binding proteins. *N Biotechnol* **55**, 5–11.
12 Collins T & Margesin R (2019) Psychrophilic lifestyles: mechanisms of adaptation and biotechnological tools. *Appl Microbiol Biotechnol* **103**, 2857–2871.
13 Feller G & Gerday C (2003) Psychrophilic enzymes: Hot topics in cold adaptation. *Nature Rev Microbiol* **1**, 200–208.
14 Mangiagalli M & Lotti M (2021) Cold-Active β-Galactosidases: Insight into Cold Adaptation Mechanisms and Biotechnological Exploitation. *Marine Drugs* **19**, 43.
15 Aguilar CF, Sanderson I, Moracci M, Ciaramella M, Nucci R, Rossi M & Pearl LH (1997) Crystal structure of the β-glycosidase from the hyperthermophilic archeon *sulfolobus solfataricus*: Resilience as a key factor in thermostability. *J Mol Biol* **271**, 789–802.
16 Brocca S, Ferrari C, Barbiroli A, Pesce A, Lotti M & Nardini M (2016) A bacterial acyl aminoacyl peptidase couples flexibility and stability as a result of cold adaptation. *FEBS J* **283**, 4310–4324.
17 Mangiagalli M, Lapi M, Maione S, Orlando M, Brocca S, Pesce A, Barbiroli A, Camilloni C, Pucciarelli S, Lotti M & Nardini M (2021) The co-existence of cold activity and thermal stability in an Antarctic GH42 β-galactosidase relies on its hexameric quaternary arrangement. *FEBS J* **288**, 546–565.
18 Carrea G & Colombo G (2000) Coupling high enzyme activity and stability: a challenging target. *Trends Biotechnol* **18**, 401–402.
19 Mhaindarkar D, Gasper R, Lupilov N, Hofmann E & Leichert LI (2018) Loss of a conserved salt bridge in bacterial glycosyl hydrolase BgIM-G1 improves substrate binding in temperate environments. *Commun Biol* **1**, 171.
20 Zanphorlin LM, de Giuseppe PO, Honorato RV, Tonoli CCC, Fattori J, Crespim E, de Oliveira PSL, Ruller R & Murakami MT (2016) Oligomerization as a strategy for cold adaptation: Structure and dynamics of the GH1 β-glucosidase from *Exiguobacterium antarcticum* B7. *Sci Rep* **6**, 23776.
21 Miao L-L, Hou Y-J, Fan H-X, Qu J, Qi C, Liu Y, Li D-F & Liu Z-P (2016) Molecular Structural Basis for the Cold Adaptedness of the Psychrophilic β-Glucosidase BglU in *Micrococcus antarcticus*. *Appl Environ Microbiol* **82**, 2021–2030.
22 Matsuzawa T, Watanabe M & Yaoi K (2017) Improved thermostability of a metagenomic glucose-tolerant β-glycosidase based on its X-ray crystal structure. *Appl Microbiol Biotechnol* **101**, 8353–8363.
23 Matsuzawa T, Jo T, Uchiyama T, Manninen JA, Arakawa T, Miyazaki K, Fushinobu S & Yaoi K (2016) Crystal structure and identification of a key amino acid for glucose tolerance, substrate specificity, and transglycosylation activity of metagenomic β-glucosidase Td2F2. *FEBS J* **283**, 2340–2353.
24 Holm L (2020) Using Dali for Protein Structure Comparison. In *Methods Mol Biol* pp. 29–42.
25 Davies GJ, Wilson KS & Henrissat B (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* **321**, 557–559.

26 Tian W, Chen C, Lei X, Zhao J & Liang J (2018) CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res* **46**, W363–W367.
27 Isorna P, Polaina J, Latorre-García L, Cañada FJ, González B & Sanz-Aparicio J (2007) Crystal Structures of Paenibacillus polymyxa β-Glucosidase B Complexes Reveal the Molecular Basis of Substrate Specificity and Give New Insights into the Catalytic Machinery of Family I Glycosidases. *J Mol Biol* **371**, 1204–1218.
28 Wang X, He X, Yang S, An X, Chang W & Liang D (2003) Structural Basis for Thermostability of β-Glycosidase from the Thermophilic Eubacterium *Thermus nonproteolyticus* HG102. *J Bacteriol* **185**, 4248–4255.
29 Salgado JCS, Meleiro LP, Carli S & Ward RJ (2018) Glucose tolerant and glucose stimulated B-glucosidases – A review. *Bioresource Technol* **267**, 704–713.
30 Mariano D, Pantuza N, Santos LH, Rocha REO, de Lima LHF, Bleicher L & de Melo-Minardi RC (2020) Glutantβase: a database for improving the rational design of glucose-tolerant β-glucosidases. *BMC Mol and Cell Biol* **21**, 50.
31 De Giuseppe PO, Souza TDACB, Souza FHM, Zanphorlin LM, Machado CB, Ward RJ, Jorge JA, Furriel RDPM & Murakami MT (2014) Structural basis for glucose tolerance in GH1 β-glucosidases. *Acta Crystallogr D* **70**, 1631–1639.
32 Skálová T, Dohnálek J, Spiwok V, Lipovová P, Vondráčková E, Petroková H, Dušková J, Strnad H, Králová B & Hašek J (2005) Cold-active β-Galactosidase from Arthrobacter sp. C2-2 Forms Compact 660kDa Hexamers: Crystal Structure at 1.9Å Resolution. *J Mol Biol* **353**, 282–294.
33 Marchetti A, Orlando M, Mangiagalli M & Lotti M (2023) A cold-active esterase enhances mesophilic properties through Mn^{2+} binding. *FEBS J* **290**, 2394–2411.
34 Sun J, Wang W, Yao C, Dai F, Zhu X, Liu J & Hao J (2018) Overexpression and characterization of a novel cold-adapted and salt-tolerant GH1 β-glucosidase from the marine bacterium Alteromonas sp. L82. *J Microbiol* **56**, 656–664.
35 Fan H-X, Miao L-L, Liu Y, Liu H-C & Liu Z-P (2011) Gene cloning and characterization of a cold-adapted β-glucosidase belonging to glycosyl hydrolase family 1 from a psychrotolerant bacterium Micrococcus antarcticus. *Enzyme Microbial Technol* **49**, 94–99.
36 Zhao W, Peng R, Xiong A, Fu X, Tian Y & Yao Q (2012) Expression and characterization of a cold-active and xylose-stimulated β-glucosidase from Marinomonas MWYL1 in Escherichia coli. *Mol Biol Rep* **39**, 2937–2943.
37 Crespim E, Zanphorlin LM, De Souza FHM, Diogo JA, Gazolla AC, Machado CB, Figueiredo F, Sousa AS, Nóbrega F, Pellizari VH, Murakami MT & Ruller R (2016) A novel cold-adapted and glucose-tolerant GH1 β-glucosidase from *Exiguobacterium antarcticum* B7. *Int J Biol Macromol* **82**, 375–380.
38 Wierzbicka-Woś A, Bartasun P, Cieśliński H & Kur J (2013) Cloning and characterization of a novel cold-active glycoside hydrolase family 1 enzyme with β-glucosidase, β-fucosidase and β-galactosidase activities. *BMC Biotechnol* **13**, 22.
39 Sarmiento F, Peralta R & Blamey JM (2015) Cold and Hot Extremozymes: Industrial Relevance and Current Trends. *Front Bioeng Biotechnol* **3**, 148.
40 Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069.
41 Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD & Bateman A (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412–D419.
42 Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH & Bateman A (2023) InterPro in 2022. *Nucleic Acids Res* **51**, D418–D427.
43 Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* **41**, 207–234.

44 Rinaldi C, Pizzul P, Casari E, Mangiagalli M, Tisi R & Longhese MP (2023) The Ku complex promotes DNA end-bridging and this function is antagonized by Tel1/ATM kinase. *Nucleic Acids Res* **51**, 1783–1802.
45 Kabsch W (2010) *XDS*. *Acta Crystallogr D* **66**, 125–132.
46 Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D* **69**, 1204–1214.
47 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* **67**, 235–242.
48 Gamiz-Arco G, Gutierrez-Rus LI, Risso VA, Ibarra-Molero B, Hoshino Y, Petrović D, Justicia J, Cuerva JM, Romero-Rivera A, Seelig B, Gavira JA, Kamerlin SCL, Gaucher EA & Sanchez-Ruiz JM (2021) Heme-binding enables allosteric modulation in an ancient TIM-barrel glycosidase. *Nat Commun* **12**, 380.
49 Emsley P & Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126–2132.
50 Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* **68**, 352–367.
51 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* **66**, 12–21.
52 Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B & Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688.
53 Case DA, Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham TE & Cruzeiro, V. W. D. (2018) Amber 2018.
54 Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE & Simmerling C (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713.
55 Salomon-Ferrer R, Götz AW, Poole D, Le Grand S & Walker RC (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **9**, 3878–3888.
56 Søndergaard CR, Olsson MHM, Rostkowski M & Jensen JH (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K_a Values. *J Chem Theory Comput* **7**, 2284–2295.
57 Loncharich RJ, Brooks BR & Pastor RW (1992) Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **32**, 523–535.
58 Darden T, York D & Pedersen L (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* **98**, 10089–10092.
59 Miyamoto S & Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* **13**, 952–962.
60 Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P & Shenkin PS (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* **47**, 1739–1749.
61 Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, Murphy R, Philipp DM, Repasky MP, Zhang LY, Berne BJ, Friesner RA, Gallicchio E & Levy RM (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem* **26**, 1752–1780.

62 Wang C, Ye F, Chang C, Liu X, Wang J, Wang J, Yan X-F, Fu Q, Zhou J, Chen S, Gao Y-G & Zhang L-H (2019) Agrobacteria reprogram virulence gene expression by controlled release of host-conjugated signals. *Proc Natl Acad Sci USA* **116**, 22331–22340.
63 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.
64 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.
65 Maiti R, Van Domselaar GH, Zhang H & Wishart DS (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* **32**, W590–W594.

Table 1. Kinetic parameters of M-GH1 measured in the absence (K_M and k_{cat}) and in the presence (K_i) of glucose and galactose.

	K_M (mM)	k_{cat} (s^{-1})	K_i (mM)	
			Glucose	Galactose
*o*NPGal	5.2 ± 0.4	17.9 ± 0.3	351 ± 49	563 ± 58
*p*NPGlu	0.2 ± 0.1	13.2 ± 0.4	161 ± 32	251 ± 27

Table 2. X-ray data collection and refinement statistics.

	M-GH1 (PDB: 8PUO)
Crystal	
Space group	$P2_12_12$
Unit cell dimensions	103.2, 201.1, 47.8
a, *b*, *c* (Å); α, β, γ (°)	90.0, 90.0, 90.0
Data collection	
Beamline	ESRF ID30A-3
Wavelength (Å)	0.9677
Resolution (Å)	67-1.8 (1.8-1.9)*
Total reflections	976400 (117976)
Unique reflections	91533 (12892)
$^{\Psi}R_{merge}$	0.083 (0.986)
$^{\#}R_{meas}$	0.087 (1.038)
$I/\sigma(I)$	16.2 (2.3)
$^{+}CC_{1/2}$	0.999 (0.805)
Completeness (%)	98.3 (96.0)
Average redundancy	10.7 (9.2)
Wilson B-factor (Å^2)	20.5
Refinement	
Resolution (Å)	45.9-1.8
No. reflections	91394
R / R_{free}	0.165 / 0.189
No. molecules (non-H atoms) in the asymmetric unit	
Protein	2 (7129)
Water	625
Tris	2 (16)
Ethylene glycol	24 (96)
Average B factors (Å^2)	
Protein	27.15
Water	35.8
Tris	36.9
Ethylene glycol	37.2
RMSD from ideal values	
Bond lengths (Å)	0.018
Bond angles (°)	1.27
Clashscore	2.8
Ramachandran	
Favored (%)	99.1
Allowed (%)	100

* Values in parenthesis correspond to the high-resolution shell. For cross-validation, 5% experimental reflections were randomly selected to calculate the R_{free} value.

Ψ $R_{merge} = \Sigma_h \Sigma_i |<I_h> - I_{h,i}|/\Sigma_h \Sigma_i I_{h,i}$

$R_{meas} = \Sigma_h [N_h/(N_h - 1)]^{1/2} \Sigma_i|<I_h> - I_{h,i}|/\Sigma_h \Sigma_i I_{h,i}$, where N_h is the data multiplicity.

+ $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data.

Table 3. Free energies of binding for M-GH1 substrates.

Ligand	Predicted binding free energy (kcal/mol)
*o*NPGal	-8.7 ± 0.1
*p*NPGlu	-8.1 ± 0.1
*p*NPClb	-8 ± 1
*p*NPXyl	-7.4 ± 0.3
*p*NPMan	-6.4 ± 0.2

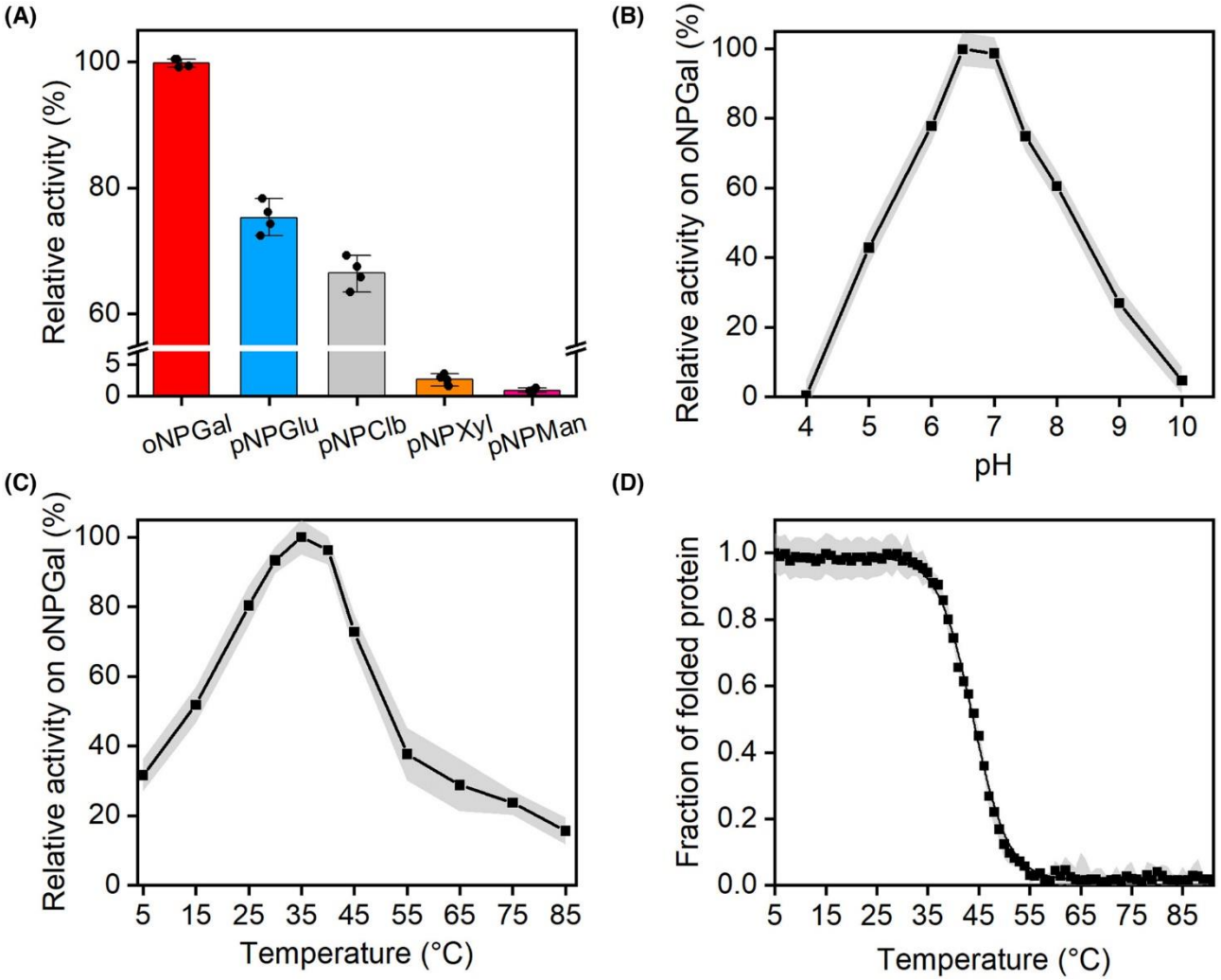

Fig. 1. Biochemical characterization of M-GH1. (A) *Marinomonas* sp. ef1 β-glucosidase (M-GH1) substrate specificity**.** Reactions were carried out on *orto*-Nitrophenyl β-D-galactopyranoside (*o*NPGal), *para*-Nitrophenyl β-D-glucopyranoside (*p*NPGlu), *para*-Nitrophenyl β-D-cellobioside (*p*NPClb), *para*-Nitrophenyl β-D-xylopyranoside (*p*NPXyl), and *para*-Nitrophenyl β-D-mannopyranoside (*p*NPMan), at 25 °C in 10mM sodium phosphate buffer (PB), pH 6.5; specific activity towards *o*NPGal was taken as 100%. Error bars indicate standard deviations of four independent experiments (n=4). (B) Effects of pH on M-GH1 activity at 35 °C and (C) effects of temperature on M-GH1 activity in 10 mM PB at pH 7.0. Activity was measured using *o*NPGal as a substrate. (D) Thermal stability of M-GH1. Ellipticity values were recorded at 208 nm over an increasing temperature gradient from 5 to 90 °C. The fraction of folded protein was calculated as described in the Material and Methods section. All experiments were performed in triplicate after incubating the enzyme in 10 mM PB, pH 6.0. The shadowed area refers to the standard deviation of the data (n=3).

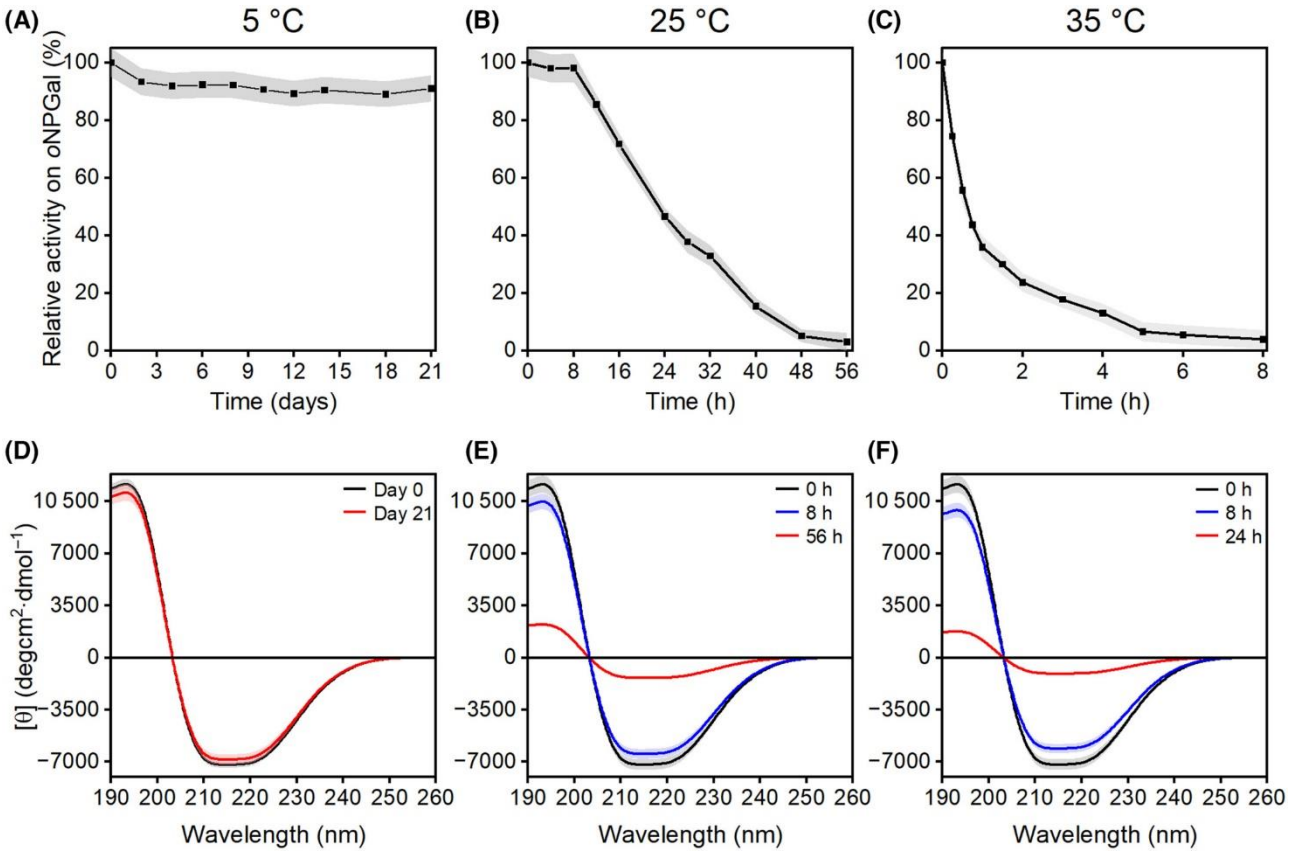

Fig. 2. Long-term M-GH1 thermal stability. Residual activity towards *orto*-Nitrophenyl β-D-galactopyranoside (*o*NPGal) was monitored after incubating the enzyme at 5 °C (A), 25 °C (B) and 35 °C (C). The activity measurements at time 0 were normalized as 100%. Far UV circular dichroism (CD) spectra were collected at 5 °C (D), 25 °C (E) and 35 °C (F) at the beginning (black line), after 8 h (blue line) and at the end (red line) of each kinetics. All experiments were performed in triplicate and the shadowed area refers to the standard data deviation (n=3).

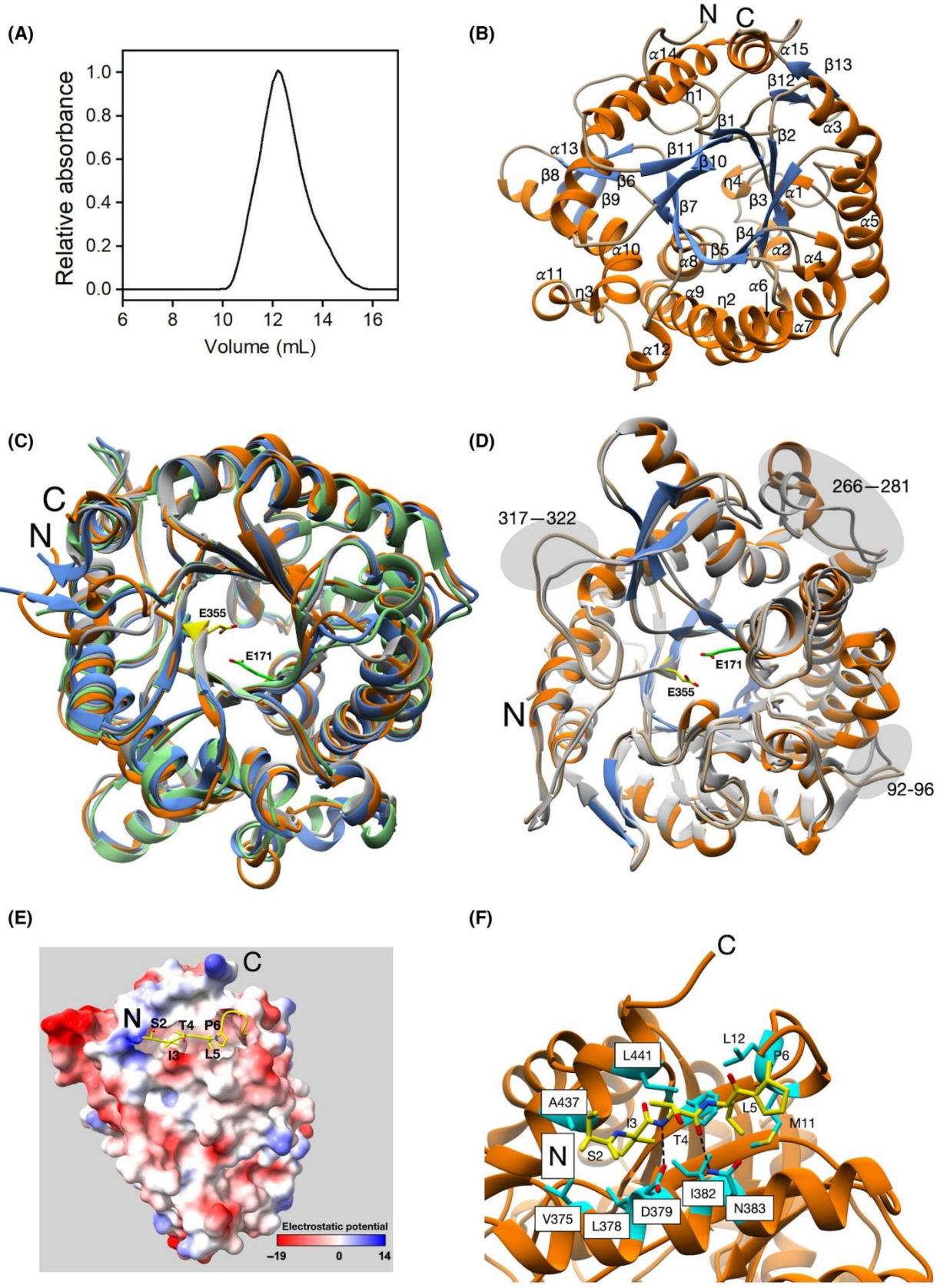
(A)
Relative absorbance
1.0
0.8
0.6
0.4
0.2
0.0
6
8
10
12
14
16
Volume (mL)
(B)
N
C
α14
α15
β13
β12
η1
β1
β2
α3
α13
β11
β10
β8
β6
η4
β9
β7
β3
α1
α5
β4
α10
β5
α2
α4
α8
α11
η3
α9
η2
α6
α7
α12
(C)
C
N
E355
E171
(D)
266—281
317—322
E171
E355
N
92-96
(E)
C
N
S2
T4
P6
I3
L5
Electrostatic potential
−19
0
14
(F)
C
L12
L441
P6
A437
L5
M11
I3
T4
N
S2
I382
V375
D379
N383
L378

Fig. 3. Oligomerization state and 3D structure of M-GH1. (A) Size-exclusion chromatography (SEC) analysis of *Marinomonas* sp. ef1 β-glucosidase (M-GH1). The gel filtration profile revealed a single peak at the elution volume of 12.2 mL, indicating that M-GH1 is in a monomeric state in solution (molecular weight of 49.5 ± 2.2 kDa). Experiments were performed in triplicate (n=3). (B) Overall secondary structure organization of M-GH1 shown in ribbons, with secondary structure labels. (C) Secondary structure ribbon representation of chain A of M-GH1 (orange) superimposed with BglM-G1 (47% sequence identity and a RMSD of 1.6 Å; grey; PDB: 5NS6 [19]), MeBglD2 (48% identity and 1.1 Å RMSD; light blue; PDB: 5XGZ [22]) and glucose-tolerant compost genome GH1 β-glucosidase Td2F2 (44% identity and 1.2 Å RMSD; green; PDB: 3WH5 [22]). M-GH1 structural homologs were identified using the DALI server [24]. The N- and C-termini of M-GH1 are indicated, and the catalytic Glu171 and Glu355 are shown in green and yellow sticks, respectively. (D) Superposition of M-GH1 with BglM-G1, (grey, PDB: 5NS6 [19]). The most evident structural deviations are highlighted in grey shading and the residue numbers are shown. The catalytic Glu171 and Glu355 residues are shown in green and yellow sticks, respectively. The N-terminus of M-GH1 is shown, whereas the C-terminus is not indicated because it hidden by the structure. (E) Electrostatic surface representation (red and blue coloring indicate negative and positive regions, respectively) of the M-GH1 monomer, illustrating the "stapling" of the first six N-terminal residues (yellow sticks) that fold back towards the core of the protein. The N and C termini are labeled and the first five N-terminal residues are labeled in single letter code. The Coulombic electrostatic potential color coding is shown. (F) Stapling interactions. For simplicity only M-GH1 secondary structure elements are shown in orange ribbons. Residues that form stabilizing interactions with N-terminal residues (yellow sticks) are indicated as cyan sticks and labeled using a one-letter code. H-bonds between Thr4 and Asp379 (3.1 Å) and Asn383 (2.9 Å) are indicated as dashes. Panels B-F were prepared using UCSF Chimera [63].

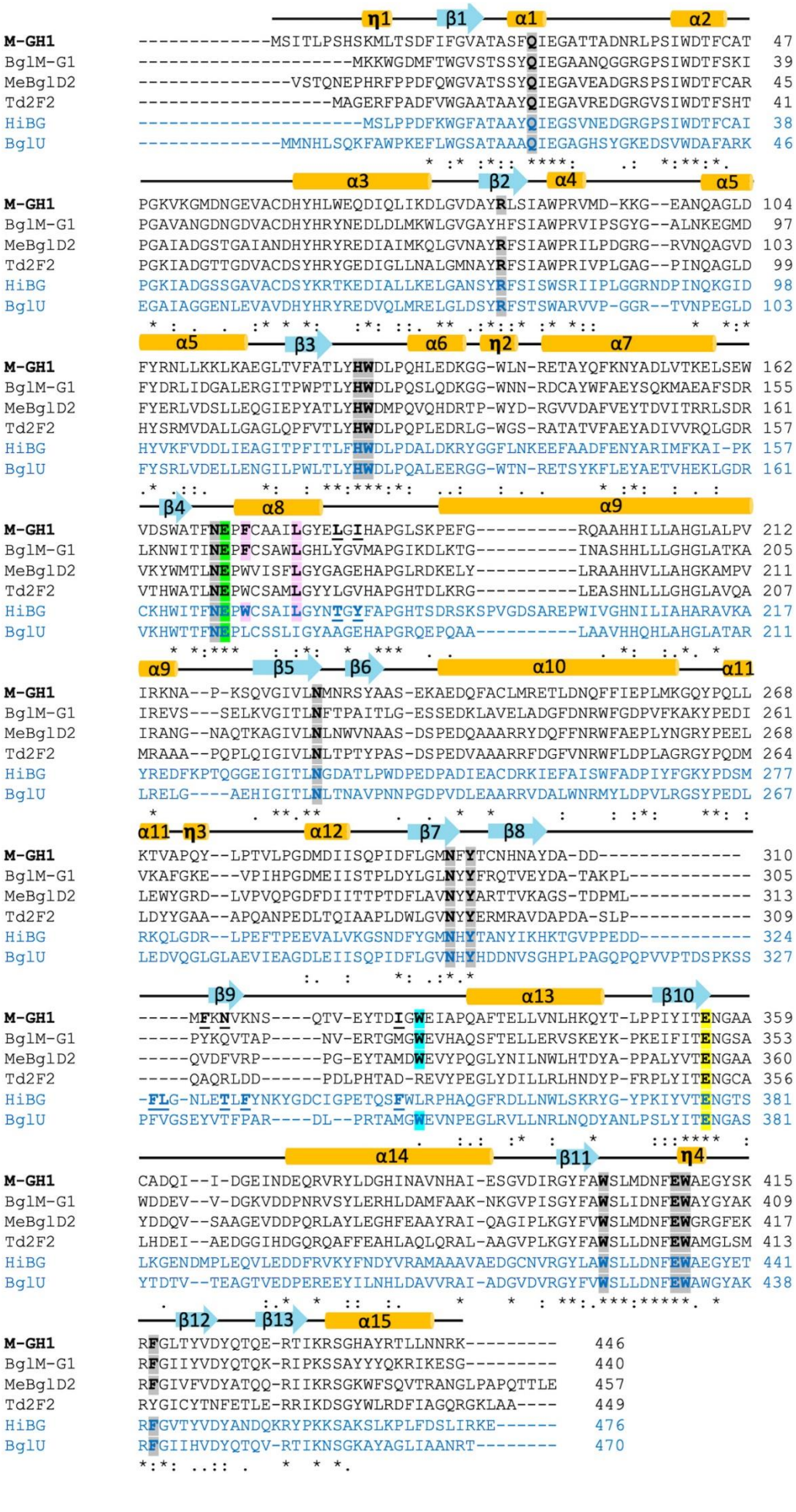
η1 β1 α1 α2
M-GH1 ----------MSITLPSHSKMLTSDFIFGVATASFQIEGATTADNRLPSIWDTFCAT 47
BglM-G1 ----------MKKWGDMFTWGVSTSSYQIEGAANQGGRGPSIWDTFSKI 39
MeBglD2 ----------VSTQNEPHRFPPDFQWGVATSSYQIEGAVEADGRSPSIWDTFCAR 45
Td2F2 ----------MAGERFPADFVWGAATAAYQIEGAVREDGRGVSIWDTFSHT 41
HiBG ----------MSLPPDFKWGFATAAYQIEGSVNEDGRGPSIWDTFCAI 38
BglU ----------MMNHLSQKFAWPKEFLWGSATAAAQIEGAGHSYGKEDSVWDAFARK 46
α3 β2 α4 α5
M-GH1 PGKVKGMDNGEVACDHYHLWEQDIQLIKDLGVDAYRLSIAWPRVMD-KKG--EANQAGLD 104
BglM-G1 PGAVANGDNGDVACDHYHRYNEDLDLMKWLGVGAYHFSIAWPRVIPSGYG--ALNKEGMD 97
MeBglD2 PGAIADGSTGAIANDHYHRYREDIAIMKQLGVNAYRFSIAWPRILPDGRG--RVNQAGVD 103
Td2F2 PGKIADGTTGDVACDSYHRYGEDIGLLNALGMNAYRFSIAWPRIVPLGAG--PINQAGLD 99
HiBG PGKIADGSSGAVACDSYKRTKEDIALLKELGANSYRFSISWSRIIPLGGRNDPINQKGID 98
BglU EGAIAGGENLEVAVDHYHRYREDVQLMRELGLDSYRFSTSWARVVP-GGR--TVNPEGLD 103
α5 β3 α6 η2 α7
M-GH1 FYRNLLKKLKAEGLTVFATLYHWDLPQHLEDKGG-WLN-RETAYQFKNYADLVTKELSEW 162
BglM-G1 FYDRLIDGALERGITPWPTLYHWDLPQSLQDKGG-WNN-RDCAYWFAEYSQKMAEAFSDR 155
MeBglD2 FYERLVDSLLEQGIEPYATLYHWDMPQVQHDRTP-WYD-RGVVDAFVEYTDVITRRLSDR 161
Td2F2 HYSRMVDALLGAGLQPFVTLYHWDLPQPLEDRLG-WGS-RATATVFAEYADIVVRQLGDR 157
HiBG HYVKFVDDLIEAGITPFITLFHWDLPDALDKRYGGFLNKEEFAADFENYARIMFKAI-PK 157
BglU FYSRLVDELLENGILPWLTLYHWDLPQALEERGG-WTN-RETSYKFLEYAETVHEKLGDR 161
β4 α8 α9
M-GH1 VDSWATFNEPFCAAILGYELGIHAPGLSKPEFG----------RQAAHHILLAHGLALPV 212
BglM-G1 LKNWITINEPFCSAWLGHLYGVMAPGIKDLKTG----------INASHHLLLGHGLATKA 205
MeBglD2 VKYWMTLNEPWVISFLGYGAGEHAPGLRDKELY----------LRAAHHVLLAHGKAMPV 211
Td2F2 VTHWATLNEPWCSAMLGYYLGVHAPGHTDLKRG----------LEASHNLLLGHGLAVQA 207
HiBG CKHWITFNEPWCSAILGYNTGYFAPGHTSDRSKSPVGDSAREPWIVGHNILIAHARAVKA 217
BglU VKHWTTFNEPLCSSLIGYAAGEHAPGRQEPQAA----------LAAVHHQHLAHGLATAR 211
α9 β5 β6 α10 α11
M-GH1 IRKNA--P-KSQVGIVLNMNRSYAAS-EKAEDQFACLMRETLDNQFFIEPLMKGQYPQLL 268
BglM-G1 IREVS---SELKVGITLNFTPAITLG-ESSEDKLAVELADGFDNRWFGDPVFKAKYPEDI 261
MeBglD2 IRANG--NAQTKAGIVLNLNWVNAAS-DSPEDQAAARRYDQFFNRWFAEPLYNGRYPEEL 268
Td2F2 MRAAA--PQPLQIGIVLNLTPTYPAS-DSPEDVAAARRFDGFVNRWFLDPLAGRGYPQDM 264
HiBG YREDFKPTQGGEIGITLNGDATLPWDPEDPADIEACDRKIEFAISWFADPIYFGKYPDSM 277
BglU LRELG----AEHIGITLNLTNAVPNNPGDPVDLEAARRVDALWNRMYLDPVLRGSYPEDL 267
α11 η3 α12 β7 β8
M-GH1 KTVAPQY--LPTVLPGDMDIISQPIDFLGMNFYTCNHNAYDA-DD-------------- 310
BglM-G1 VKAFGKE---VPIHPGDMEIISTPLDYLGLNYYFRQTVEYDA-TAKPL------------ 305
MeBglD2 LEWYGRD--LVPVQPGDFDIITTPTDFLAVNYYARTTVKAGS-TDPML------------ 313
Td2F2 LDYYGAA--APQANPEDLTQIAAPLDWLGVNYYERMRAVDAPDA-SLP------------ 309
HiBG RKQLGDR--LPEFTPEEVALVKGSNDFYGMNHYTANYIKHKTGVPPEDD----------- 324
BglU LEDVQGLGLAEVIEAGDLEIISQPIDFLGVNHYHDDNVSGHPLPAGQPQPVVPTDSPKSS 327
β9 α13 β10
M-GH1 -----MFKNVKNS----QTV-EYTDIGWEIAPQAFTELLVNLHKQYT-LPPIYIITENGAA 359
BglM-G1 -----PYKQVTAP-----NV-ERTGMGWEVHAQSFTELLERVSKEYK-PKEIFITENGSA 353
MeBglD2 -----QVDFVRP------PG-EYTAMDWEVYPQGLYNILNWLHTDYA-PPALYVTENGAA 360
Td2F2 -----QAQRLDD------PDLPHTAD-REVYPEGLYDILLRLHNDYP-FRPLYITENGCA 356
HiBG -FLG-NLETLFYNKYGDCIGPETQSFWLRPHAQGFRDLLNWLSKRYG-YPKIYVTENGTS 381
BglU PFVGSEYVTFPAR----DL--PRTAMGWEVNPEGLRVLLNRLNQDYANLPSLYITENGAS 381
α14 β11 η4
M-GH1 CADQI--I-DGEINDEQRVRYLDGHINAVNHAI-ESGVDIRGYFAWSLMDNFEWAEGYSK 415
BglM-G1 WDDEV--V-DGKVDDPNRVSYLERHLDAMFAAK-NKGVPISGYFAWSLIDNFEWAYGYAK 409
MeBglD2 YDDQV--SAAGEVDDPQRLAYLEGHFEAAYRAI-QAGIPLKGYFVWSLMDNFEWGRGFEK 417
Td2F2 LHDEI--AEDGGIHDGQRQAFFEAHLAQLQRAL-AAGVPLKGYFAWSLLDNFEWAMGLSM 413
HiBG LKGENDMPLEQVLEDDFRVKYFNDYVRAMAAAVAEDGCNVRGYLAWSLLDNFEWAEGYET 441
BglU YTDTV--TEAGTVEDPEREEYILNHLDAVVRAI-ADGVDVRGYFVWSLLDNFEWAWGYAK 438
β12 β13 α15
M-GH1 RFGLTYVDYQTQE-RTIKRSGHAYRTLLNNRK--------- 446
BglM-G1 RFGIIYVDYQTQK-RIPKSSAYYYQKRIKESG--------- 440
MeBglD2 RFGIVFVDYATQQ-RIIKRSGKWFSQVTRANGLPAPQTTLE 457
Td2F2 RYGICYTNFETLE-RRIKDSGYWLRDFIAGQRGKLAA---- 449
HiBG RFGVTYVDYANDQKRYPKKSAKSLKPLFDSLIRKE------ 476
BglU RFGIIHVDYQTQV-RTIKNSGKAYAGLIAANRT-------- 470

Fig. 4. Structure-based sequence alignment of M-GH1 with protein homologs. Schematic representation of the secondary structure organization of *Marinomonas* sp. ef1 β-glucosidase (M-GH1) in relation to amino acid sequences (taken from the PDB structures), and multiple alignment (carried out using Clustal Omega [64]) with the top three closest structural homologs (black font): BglM-G1 (PDB: 5NS6 [19]) MeBglD2 (PDB: 5XGZ [22]) and Td2F2 (PDB: 3WH5 [23]). Additionally, cold-active GH1 from the Antarctic bacterium *Micrococcus antarcticus* (BglU, PDB: 3W53 [21]) and glucose-tolerant *Humicola insolens* β-glucosidase (HiBG, PDB: 4MDP [31]) were also aligned, both shown in blue font. The sequence alignment was manually corrected based on 3D structure comparison. Catalytic Glu171 is shaded in green and Glu355 in yellow. Active site residues (glycone subsite) and Trp328 are highlighted in grey and cyan, respectively. Residues belonging to the aglycone subsite of HiBG and the corresponding residues in M-GH1 are underlined. Putative "gatekeeping" residues F173 and L178, at the entrance of the substrate binding pocket, are highlighted in pink. Identical and similar residues are indicated by asterisks and dots, respectively.

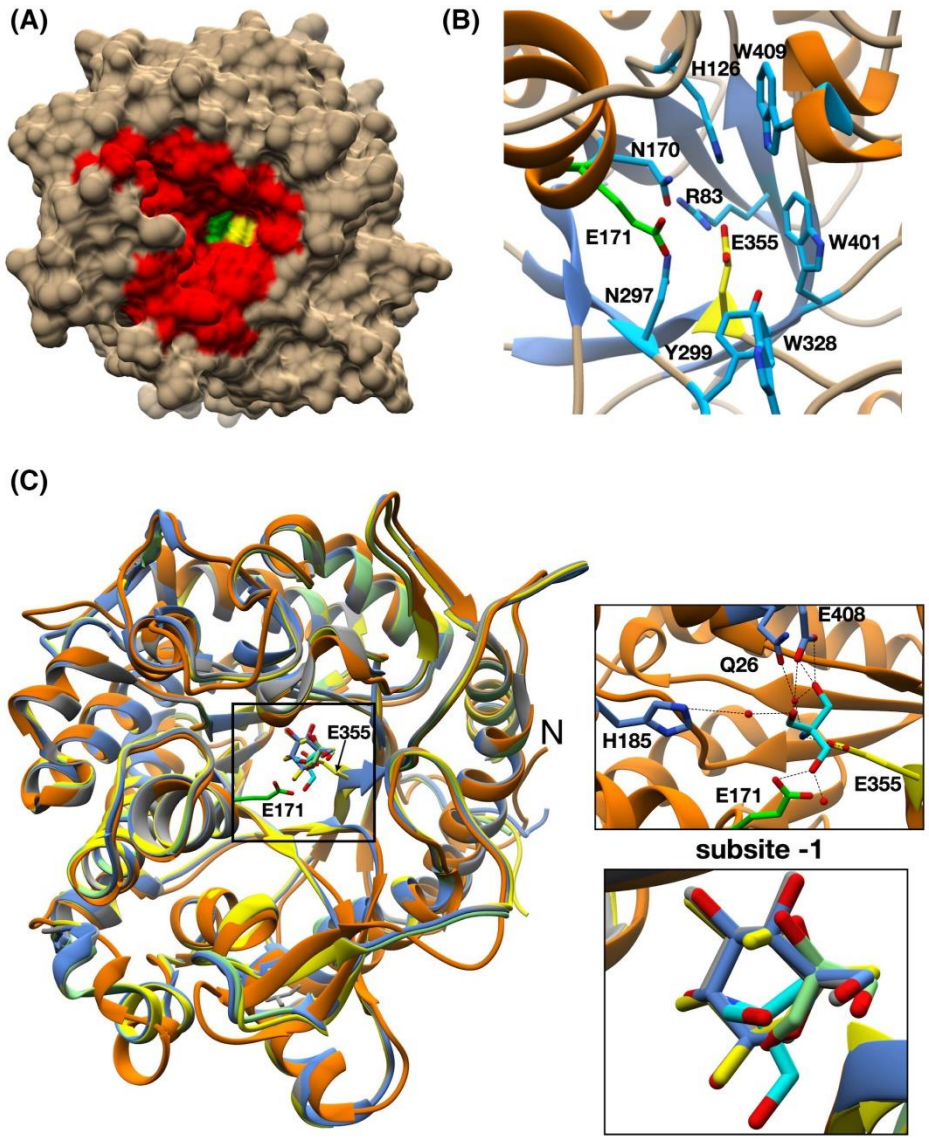

Fig. 5. The GH1 active site and comparison with ligand-bound structures of compost genome GH1 β-glucosidase Td2F2. (A) Surface representation of the *Marinomonas* sp. ef1 β-glucosidase (M-GH1) monomer. The active site pocket is highlighted in red. Green and yellow coloring refer to the location of Glu171 and Glu355, respectively. (B) Detailed view of the active site with conserved residues (blue sticks), surrounding the catalytic Glu171 (green sticks) and Glu355 (yellow sticks). Residues are labeled using the one-letter code. The structure is oriented as in panel A. (C). Secondary structure ribbon representation of Chain B of M-GH1 (orange) with a Tris molecule (derived from the crystallization buffer; cyan sticks) present in the active site, superposed with Td2F2 bound to glycerol (green; PDB: 3WH5), D-glucose (blue; PDB: 3WH6), isofagomine (yellow; PDB: 3WH7) and D-fucose (grey; PDB: 3WH8) [23]. Superpositions were made using SuperPose [65]. Ligands of M-GH1 homologs are shown as sticks and colored in line with the color of their corresponding receptor protein structure. Oxygen and nitrogen atoms are colored red and blue, respectively. The catalytic residues Glu171 and Glu355 are indicated as a reference. The upper inset highlights the H-bonds (bond length range 2.5–3.3 Å; black dashes) made between the Tris molecule and M-GH1 active site residues that are labeled accordingly. Water molecules that participate in stabilizing H-bonds are also shown as red spheres. The lower inset underlines the superimposable ligand positions between all structures. This figure was prepared using UCSF Chimera [63]).

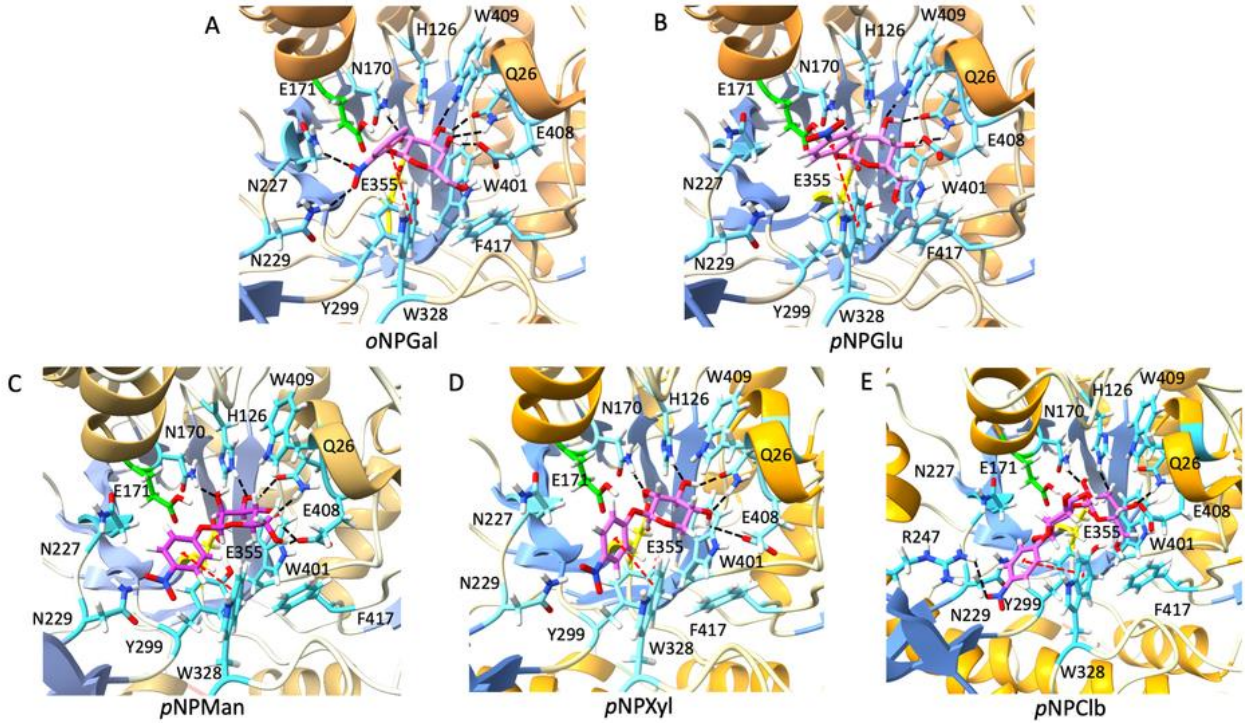

Fig.6. Binding of the tested substrates substrates to M-GH1. *In silico* docking of tested substrates was carried out on the *Marinomonas* sp. ef1 β-glucosidase (M-GH1) crystal structure: (A) *orto*-Nitrophenyl β-D-galactopyranoside (*o*NPGal), (B) *para*-Nitrophenyl β-D-glucopyranoside (*p*NPGlu), (C) *para*-Nitrophenyl β-D-mannopyranoside (*p*NPMan), (D) *para*-Nitrophenyl β-D-xylopyranoside (*p*NPXyl), and (E) *para*-Nitrophenyl β-D-cellobioside (*p*NPClb) bound at the active site. Residues lining the active site are shown in sticks and labeled using a one-letter code. Hydrogen bond interactions are marked by black dashed lines (distance between donor and acceptor atoms within 2.5–3.3 Å). T-shaped π stacking interaction of the nitro-phenyl moiety of the substrates with the side chain of Trp328 is marked by a red dashed line. Figures were prepared using UCSF Chimera [63].

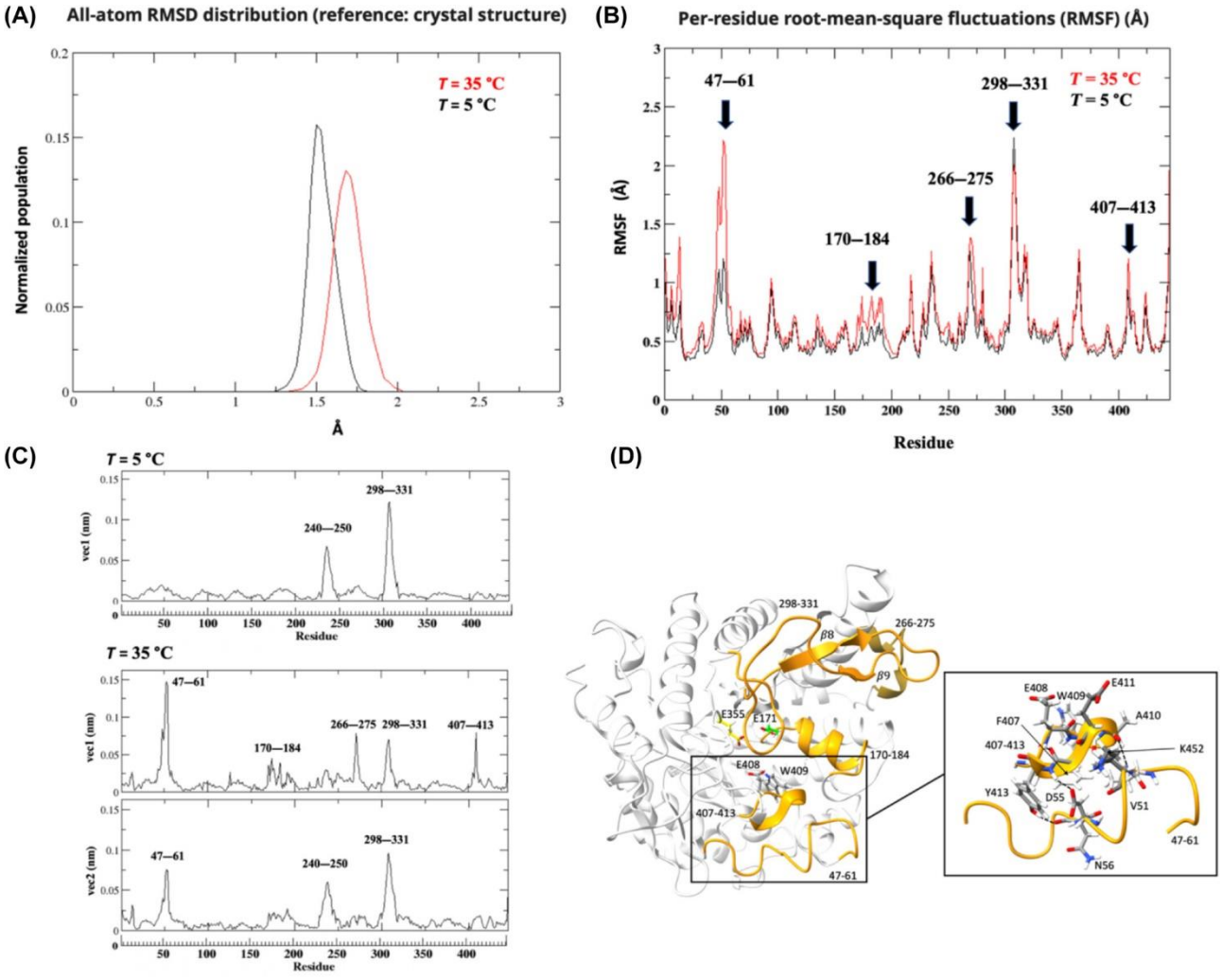

Fig. 7. MD simulations and covariance analysis of M-GH1. (A) Distributions of all-atom root-mean-square deviation (RMSD) values for MD simulations carried out at 5 °C (black line) and 35 °C (red line), using the *Marinomonas* sp. ef1 β-glucosidase (M-GH1) crystal structure as reference. (B) Per-residue root-mean-square fluctuations (RMSF) (Å) for MD simulations at 5 °C (black line) and 35 °C (red line). Arrows and labels highlight protein regions with high RMSF. (C) Covariance analysis of the Cα carbons of M-GH1. Projection of the RMSFs of Cα carbons of each residue on the first eigenvector for the simulation at 5 °C (upper panel) and on the first and second eigenvectors for the simulation at 35 °C (lower panel).

Three independent MD replicas (different random seeds) were carried out for each temperature (n=3). (D) Mapping RMSF fluctuations of all atoms onto the M-GH1 crystal structure. Regions that display fluctuations exceeding 0.7 Å, as observed in the RMSF profile, are highlighted in orange with corresponding residue numbers specified. Catalytic residues Glu171 and Glu355 are represented by green and yellow sticks, respectively. Residues E408 and W409 are shown in gray sticks. The inset displays interacting residues of loops 47-61 and 407-413. Black dashed lines mark hydrogen bond interactions (distance between donor and acceptor atoms within 2.5–3.3 Å). Figures were prepared using UCSF Chimera [63].

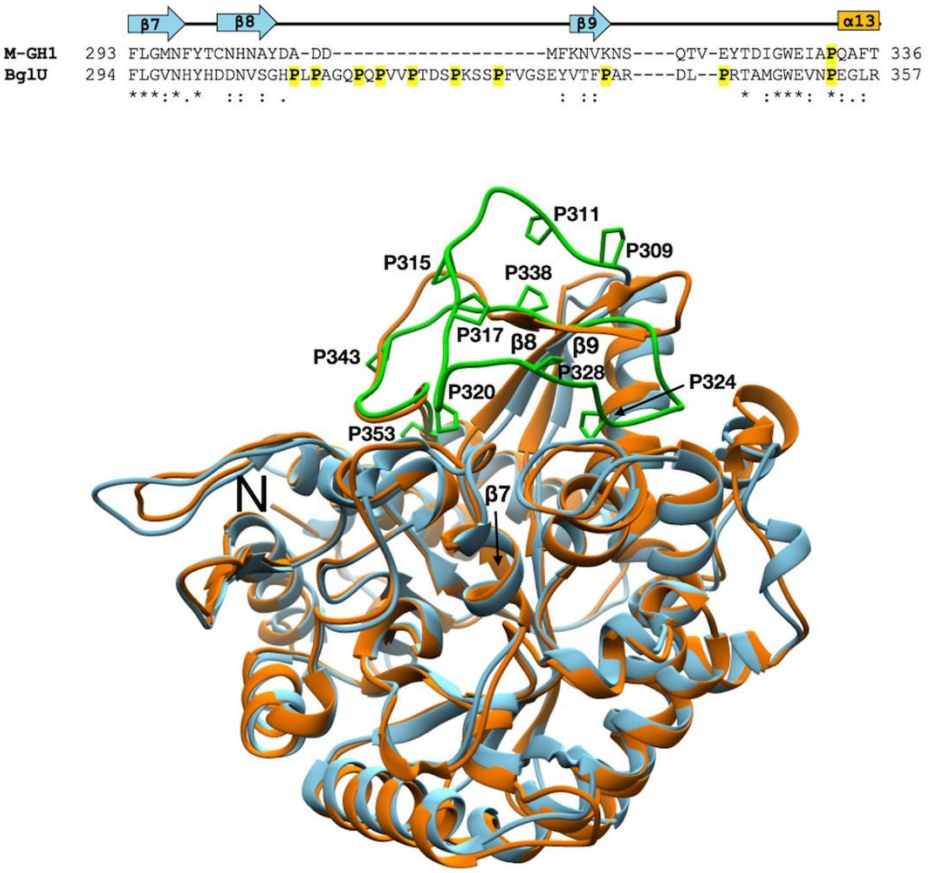

Fig. 8. Sequence and structure comparisons of GH1 and BglU. (Upper panel) Sequence alignment of region 298-336 of *Marinomonas* sp. ef1 β-glucosidase (M-GH1) with that of GH1 from the Antarctic bacterium *Micrococcus antarcticus* (BglU: residues 294 to 357) performed with Clustal Omega [64]. Secondary structure elements were extracted from the 3D structure of M-GH1, and Pro residues are highlighted in yellow. (Lower panel) Secondary structure ribbon representation of Chain B of M-GH1 (orange) superposed with BlgU (blue) (PDB: 3W53 [21]). The N-terminus of M-GH1 and β-strands 7, 8 and 9 are labeled. The C-terminus is not indicated because hidden by the structure. The BlgU Pro-rich loop is shown in green. Proline residues are labeled using the one-letter code. This figure was prepared using Chimera UCSF [63]).

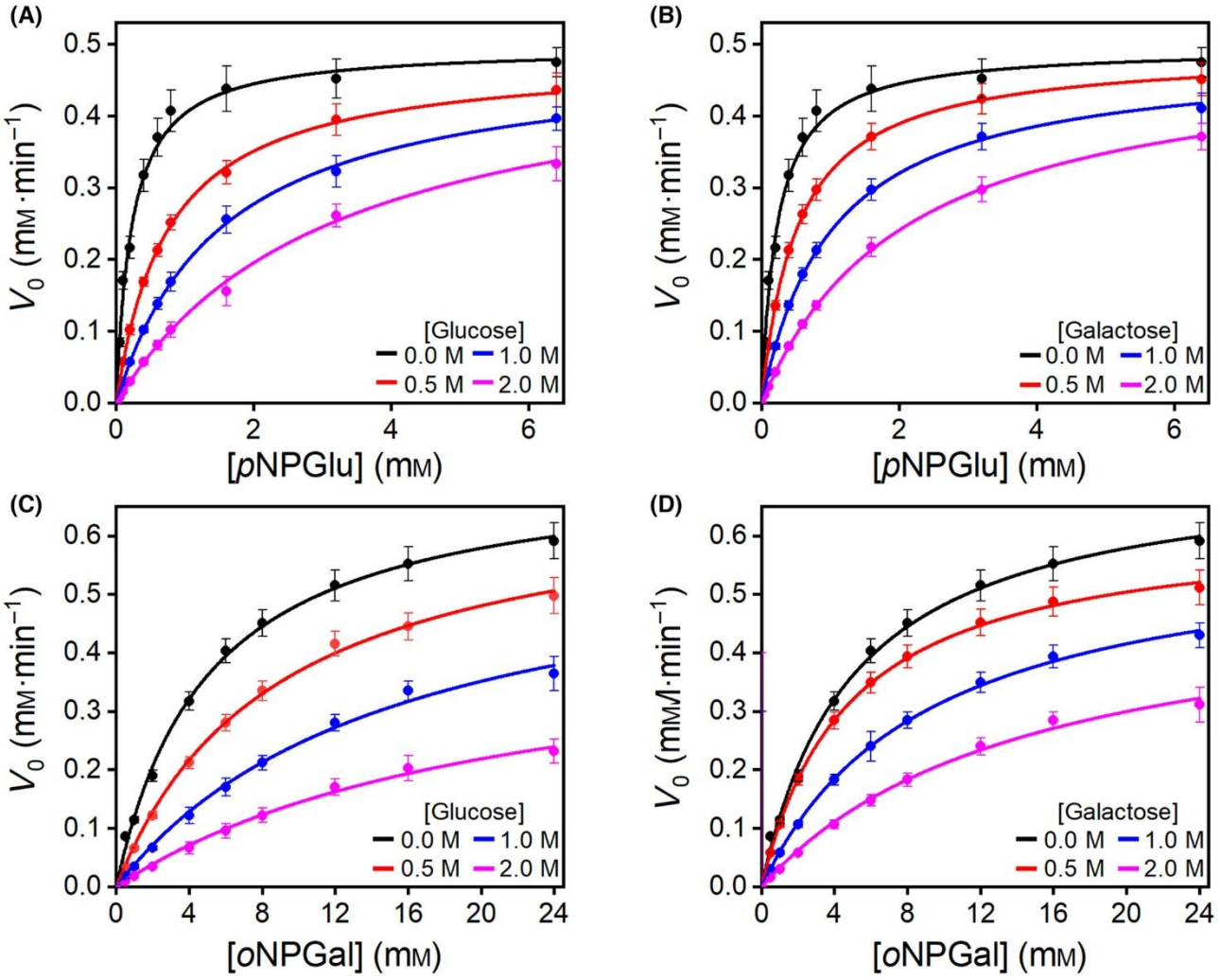

Fig. 9. Inhibitory effects of glucose and galactose on M-GH1 activity. Steady-state kinetics of *Marinomonas* sp. ef1 β-glucosidase (M-GH1) were carried out in the absence and in the presence of different concentrations of glucose (A, C) and galactose (B, D), as reported in the figure legend. Michaelis-Menten kinetics were recorded using increasing concentrations of *para*-Nitrophenyl β-D-glucopyranoside (*p*NPGlu) (A, B) or *orto*-Nitrophenyl β-D-galactopyranoside (*o*NPGal) (C, D) as substrates. Error bars indicate standard deviations on three independent experiments (n=3).

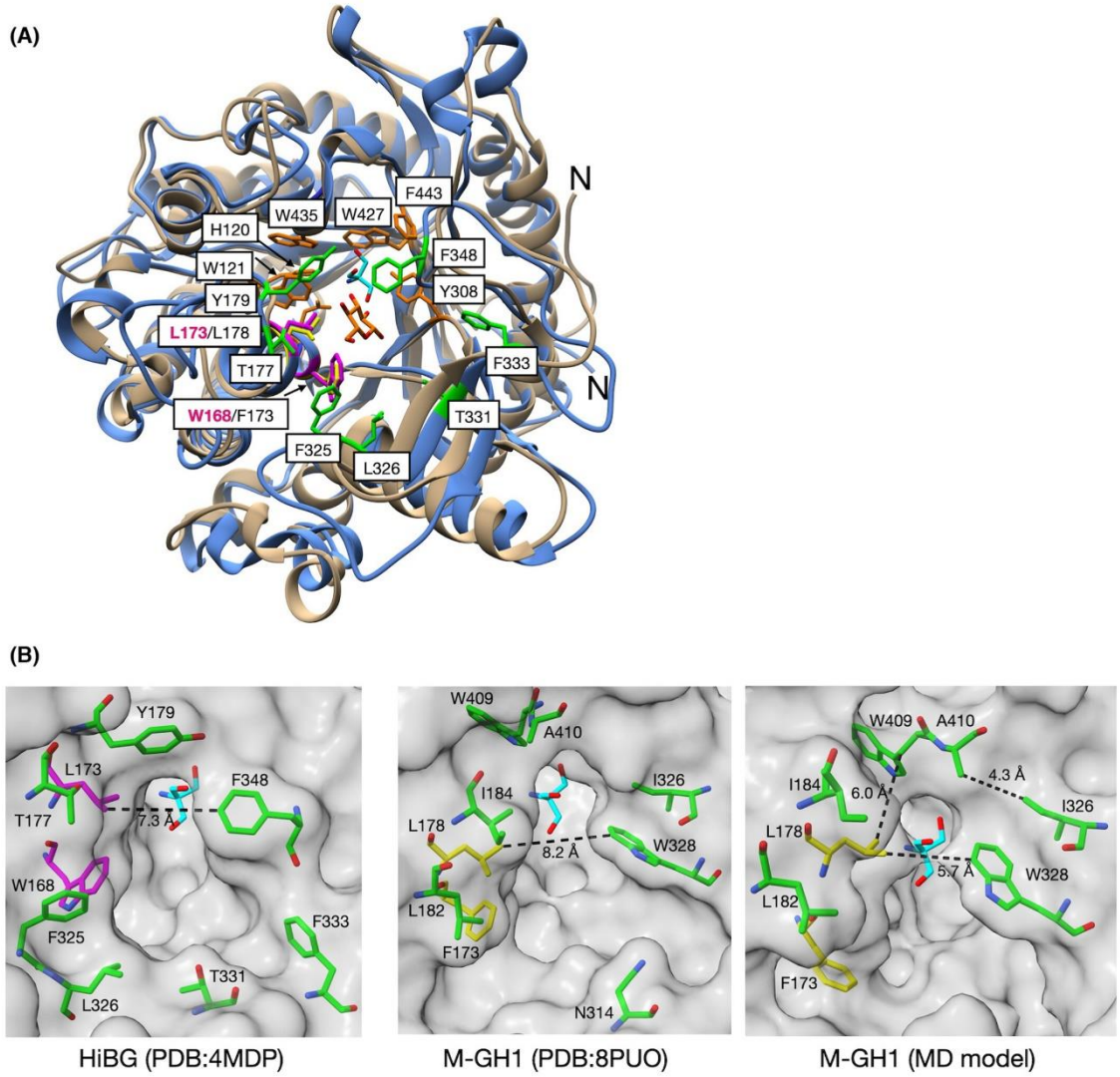

Fig. 10. Superposition of M-GH1 with glucose bound *H. insolens* β-glucosidase and the gate-keeping role of L178. (A) Secondary structure ribbon representation of Chain A of *Marinomonas* sp. ef1 β-glucosidase (M-GH1; PDB: 8PUO, brown) superposed with glucose-bound *H. insolens* β-glucosidase (HiBG; PDB: 4MDP, blue [31]). HiBG-bound glucose (orange sticks), and the Tris molecule bound in the active site of M-GH1 (cyan sticks) are shown. Residues of the glycone and aglycone subsites of HiBG are shown in orange and green sticks, respectively, and labeled with the one-letter code. HiBG gate-keeping residues Trp168 and Leu173 (magenta sticks and magenta labeling), and Phe173 and Leu178 of M-GH1 (yellow sticks) are also shown. Nitrogen and oxygen atoms are shown for ligands in blue and red, respectively. (B) Surface representation (transparent gray surface) and structural comparison of the active-site entrances of HiBG (PDB: 4MDP), and M-GH1 (PDB: 8PUO and a representative protein conformation from MD simulations). Residues that delineate the respective active-site entrances are shown in green sticks with the "gatekeeping" residues in magenta and yellow sticks for of HiBG and M-GH1, respectively. The active site cavity in HiBG and M-GH1 is identified by the position of the bound-Tris molecule, shown in cyan sticks. Distances indicating the minimal width of the tunnel connecting the aglycone-binding site and the active site are shown as dashed lines. Additionally, the closest distances between the C atoms of the side chains of Ile326-Ala410, and Leu178-Trp409 are also shown in the right panel. The protein orientation shown in panel B is similar to that of panel A. This figure was prepared using UCSF Chimera [63].