



High-resolution crystal structure of a polyextreme GH43 glycosidase from *Halothermothrix orenii* with α -L-arabinofuranosidase activity

Noor Hassan,^{a,b} Lokesh D. Kori,^{c,d} Rosaria Gandini,^{a,b} Bharat K. C. Patel,^c Christina Divne^{a,b} and Tien Chye Tan^{a,b*}

Received 1 January 2015
Accepted 16 February 2015

Keywords: glycoside hydrolase; five-bladed β -propeller; arabinofuranosidase; *Halothermothrix orenii*; halothermophile.

PDB reference: HoAraf43, 4qqs

Supporting information: this article has supporting information at journals.iucr.org/f

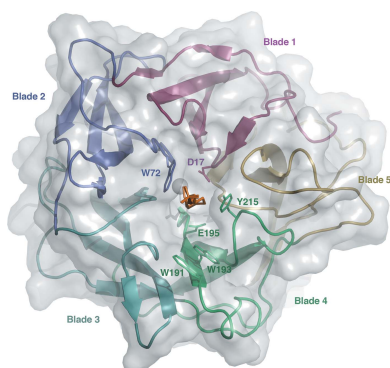
^aSchool of Biotechnology, KTH Royal Institute of Technology, Stockholm, Sweden, ^bDepartment of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden, ^cMicrobial Gene Research and Resources Facility, School of Natural Sciences, Griffith University, Brisbane, QLD 4111, Australia, and ^dDepartment of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA. *Correspondence e-mail: tantc@kth.se

A gene from the heterotrophic, halothermophilic marine bacterium *Halothermothrix orenii* has been cloned and overexpressed in *Escherichia coli*. This gene encodes the only glycoside hydrolase of family 43 (GH43) produced by *H. orenii*. The crystal structure of the *H. orenii* glycosidase was determined by molecular replacement and refined at 1.10 Å resolution. As for other GH43 members, the enzyme folds as a five-bladed β -propeller. The structure features a metal-binding site on the propeller axis, near the active site. Based on thermal denaturation data, the *H. orenii* glycosidase depends on divalent cations in combination with high salt for optimal thermal stability against unfolding. A maximum melting temperature of 76°C was observed in the presence of 4 M NaCl and Mn²⁺ at pH 6.5. The gene encoding the *H. orenii* GH43 enzyme has previously been annotated as a putative α -L-arabinofuranosidase. Activity was detected with *p*-nitrophenyl- α -L-arabinofuranoside as a substrate, and therefore the name *HoAraf43* was suggested for the enzyme. In agreement with the conditions for optimal thermal stability against unfolding, the highest arabinofuranosidase activity was obtained in the presence of 4 M NaCl and Mn²⁺ at pH 6.5, giving a specific activity of 20–36 $\mu\text{mol min}^{-1} \text{mg}^{-1}$. The active site is structurally distinct from those of other GH43 members, including arabinanases, arabinofuranosidases and xylanases. This probably reflects the special requirements for degrading the unique biomass available in highly saline aqueous ecosystems, such as halophilic algae and halophytes. The amino-acid distribution of *HoAraf43* has similarities to those of mesophiles, thermophiles and halophiles, but also has unique features, for example more hydrophobic amino acids on the surface and fewer buried charged residues.

1. Introduction

Carbohydrate-based polymers offer a large renewable natural resource, and extensive efforts have been invested in investigating how biomass may be used to harness energy through its conversion to fermentable sugars by environmentally sustainable enzymatic processes. The most prominent decomposers of lignocellulosic biomass are bacteria and fungi, which typically produce intricate and complex networks of hydrolytic (glycoside hydrolases; GHs) and oxidative enzymes for the complete solubilization of cellulose, hemicelluloses and lignins. In this context, the genomes of thermophilic anaerobic microbes present a treasure trove of genetic information, including a large repertoire of genes encoding intracellular and extracellular glycoside hydrolases, some of which could possess useful biochemical properties suitable for exploitation in improved or new enzymatic bioprocesses.

The marine bacterium *Halothermothrix orenii* of the order Haloanaerobiales is interesting as an obligate



OPEN ACCESS

halothermophile, and is also the first truly thermophilic halophile to be described whose complete genome has been sequenced (Mijts & Patel, 2001; Mavromatis *et al.*, 2009). This strictly anaerobic bacterium was first isolated from the 40–60 cm bottom sediment layer of the warm saline lake Chott El Guettar in Tunisia (Cayol *et al.*, 1994). This habitat supports the optimal growth of *H. orenii* at a temperature of 60°C (range 45–68°C) with 1.7 M NaCl (range 0.7–3.4 M) and within the pH range 5.5–8.2, with optimal growth between pH 6.5 and 7.0 (Cayol *et al.*, 1994).

The *H. orenii* genome contains only a single gene encoding a family 43 glycoside hydrolase (CAZy GH43; <http://www.cazy.org>; Lombard *et al.*, 2014), which contrasts with, for instance, the genome of the soil saprophyte *Cellvibrio japonicus*, which contains 14 GH43 genes (DeBoy *et al.*, 2008). The GH43 family mainly includes hemicellulases, and the large number of GH43 genes in *C. japonicus* may reflect the need for a wide variety of hemicellulases for the efficient degradation of complex plant cell-wall material. Since information is lacking regarding substrate availability in the various sediment layers of Chott El Guettar, we may only speculate that this saline environment offers biomass of a less variable nature, probably mostly of algal origin, which could explain the lack of catalytic diversification of GH43 enzymes in *H. orenii*. To date, the crystal structures of 28 unique GH43 members have been determined, mainly including bacterial β -xylosidases, xylanases, α -L-arabinofuranosidases and arabinanases (<http://www.cazy.org>). Based on sequence similarity to other GH43 arabinofuranosidases (EC 3.2.1.55), the gene encoding the *H. orenii* glycosidase (Gene ID 7314382; UniProt accession No. B8CZV1) was suggested to be a putative α -L-arabinofuranosidase (Kori, 2012). Here, we report the cloning, overexpression, high-resolution crystal structure determination and preliminary biochemical characterization of the *H. orenii* GH43 glycosidase.

2. Experimental procedures

2.1. Cloning, expression and purification

The gene encoding the *H. orenii* glycosidase was cloned into the pNIC28-Bsa4 vector with a cleavable His₆ tag and a *Tobacco etch virus* (TEV) protease cleavage site at the N-terminus (Savitsky *et al.*, 2010) using ligation-independent cloning (Doyle, 2005). The recombinant plasmid was transformed into *Escherichia coli* strain BL21(DE3) and the protein was overexpressed for 16–18 h at 18°C following induction with 0.2 mM isopropyl β -D-1-thiogalactopyranoside. The cells were lysed in 20 mM HEPES pH 7.0, 150 mM NaCl (lysis buffer) using an Avestin Emulsiflex-C3 system. The gene product was purified from the lysate using Ni-NTA agarose resin (Invitrogen) eluted with lysis buffer containing 350 mM imidazole. The His₆ tag and imidazole were removed by dialysis of the sample overnight in a dialysis bag (12–14 kDa molecular-weight cutoff) containing TEV protease and 20 mM HEPES pH 7.0, 150 mM NaCl. The sample was rerun through the Ni-NTA resin and the flowthrough containing the target

protein was collected and concentrated. To remove remaining Ni²⁺ ions, 20 mM EDTA and 20 mM EGTA were added and a final step of gel filtration was performed using a HiLoad 26/60 Superdex 200 prep-grade column (GE Healthcare Life Sciences) equilibrated with 20 mM HEPES pH 7.0, 150 mM NaCl. Eluted fractions containing the *H. orenii* glycosidase were collected in a 96-well deep-well microplate, pooled and concentrated using a Vivaspinn centrifugal concentrator (10 kDa molecular-weight cutoff).

2.2. Stability assay using Thermofluor screening

The effects of various parameters on the thermal stability against unfolding of the GH43 glycosidase were investigated using Thermofluor screening (Ericsson *et al.*, 2006). Specifically, the influence of pH (5.5–9.0), metal ions (2 mM Mn²⁺, Mg²⁺, Ca²⁺ and Na⁺) and salt (0, 0.1, 0.5, 1.0, 2.0, 4.0 and 5.0 M NaCl) on the melting temperature was assessed. Reaction mixtures consisted of 1 μ l 23 mg ml⁻¹ protein in 10 mM bis-tris buffer pH 5.5, 1 μ l SYPRO Orange dye 25 \times stock solution (diluted 1:40 with water) and 23 μ l screen solution to give a total volume of 25 μ l reaction mixture per well. The reaction mixtures were added to a 96-well thin-walled PCR plate (Bio-Rad) and the plate was sealed with Optical Quality Sealing Tape (Bio-Rad) and heated in an iCycler iQ Real Time PCR Detection System (Bio-Rad) from 20 to 90°C in steps of 1°C, during which changes in fluorescence were monitored with a charge-coupled device (CCD) camera. The wavelengths for excitation and emission were 490 and 575 nm, respectively.

2.3. Screening of aryl-glycoside substrates and thin-layer chromatography

Enzyme activity was initially screened in a high-throughput format at 50°C in 50 mM bis-tris buffer pH 6.5 using a 96-well deep-well plate format with the following compounds (all from Sigma-Aldrich): *p*-nitrophenyl- α -D-glucopyranoside (*p*NP- α Glc), *p*-nitrophenyl- β -D-glucopyranoside (*p*NP- β Glc), *o*-nitrophenyl- β -D-glucopyranoside (*o*NP- β Glc), *p*-nitrophenyl- α -D-galactopyranoside (*p*NP- α Gal), *p*-nitrophenyl- β -D-galactopyranoside (*p*NP- β Gal), *o*-nitrophenyl- β -D-galactopyranoside (*o*NP- β Gal), *p*-nitrophenyl- α -L-arabinofuranoside (*p*NP- α Araf), *p*-nitrophenyl- α -L-arabinopyranoside (*p*NP- α Arap), *p*-nitrophenyl- β -L-arabinopyranoside (*p*NP- β Arap), *p*-nitrophenyl- α -D-xylopyranoside (*p*NP- α Xyl) and *p*-nitrophenyl- β -D-xylopyranoside (*p*NP- β Xyl). Activity was tested in the absence and presence of 2 mM metal chloride salts (Ca²⁺, Mg²⁺ and Mn²⁺) and at different NaCl concentrations (0, 1, 3 and 4 M NaCl).

The enzyme reactions were performed in volumes of 100 μ l, including 90 μ l buffer solution (50 mM bis-tris pH 6.5) with the appropriate concentrations of salt and metal and 8 μ l 19 mM substrate. The buffer mixtures were placed at 50°C for 30 min to equilibrate before the enzyme was added. The reactions were started by adding 2 μ l 12 mg ml⁻¹ enzyme. After 2, 4 and 24 h, 100 μ l sodium carbonate was added to inactivate the enzyme and to stabilize the chromophore in its anionic 4-nitrophenolate form. Samples of 150 μ l were

withdrawn and the absorbance was measured at 420 nm (*o*NP) or 405 nm (*p*NP). Calibration curves were prepared for *o*NP and *p*NP as standards at two selected NaCl concentrations (0 and 4 M) using the aforementioned reaction conditions but without adding enzyme. The *p*NP concentration range used was 0–80 mM. For *o*NP, a molar extinction coefficient of 4600 M⁻¹ cm⁻¹ was used to calculate the amount of generated *o*NP.

For a repeated activity assay, the same protocol was used as above but only for those conditions that showed the highest enzyme activity in the initial screen. This time, both EDTA/EGTA-treated and untreated enzyme were used. To address the problems associated with inflated absorbance values owing to oxidation and secondary reactions of Mn²⁺ under alkaline conditions, the absorbance values for the reactions were taken before and after adding sodium carbonate. Graphs were prepared using *GraphPad Prism*.

For thin-layer chromatography (TLC) analysis, 2 µl 12 mg ml⁻¹ *H. orenii* glycosidase was mixed with 8 µl 10 mM substrate (α -L-arabinofuranopentaose or lactose) in 50 mM bis-tris pH 6.5 to a final volume of 100 µl and incubated at 40°C for 4 h. The composition of the reactions (substrates and products) was analyzed using TLC. The samples from the reaction with lactose as the substrate were run on Silica gel 60 plates (Merck Millipore), while for the α -L-arabinofuranopentaose samples HPTLC plates with Silica gel 60 matrix (Merck Millipore) were used. A mixture of *n*-butanol:isopropanol:ethanol:water (2:3:3:2) was used as the mobile phase. The plates were developed with a thymol solution (0.5 g thymol dissolved in 95 ml ethanol and 5 ml sulfuric acid) and heated at 120°C for 15 min.

2.4. Crystallization, data collection and structure determination

Crystallization conditions for the *H. orenii* glycosidase were screened in 96-well Corning 3550 plates using the sitting-drop vapour-diffusion method and Mosquito Crystal robotics (TTP Labtech). Initial hits were optimized using 24-well Intelli-Plate 24 plates (Art Robbins Instrument). Well diffracting crystals grew at 298 K from 1 µl protein solution (178 mg ml⁻¹ in 20 mM HEPES pH 7.0, 150 mM NaCl) mixed with 1 µl reservoir solution consisting of 0.1 M HEPES pH 6.2, 0.16 M potassium thiocyanate, 25%(w/v) polyethylene glycol 3350. The crystals belonged to space group *P*2₁, with unit-cell parameters *a* = 44.13, *b* = 73.88, *c* = 87.52 Å, β = 94.26° and two molecules per asymmetric unit.

Prior to data acquisition, the crystal was flash-cooled in liquid nitrogen. Diffraction intensities to 1.10 Å resolution were collected on the microfocuss beamline I24 at Diamond Light Source, UK equipped with a Dectris PILATUS 6M pixel detector and Oxford Danfysik/SESO bimorph-type mirrors. The data were indexed, processed and scaled using the *XDS* package (Kabsch, 2010). An initial *Phyre2* analysis (Kelley & Sternberg, 2009) returned endo-1,4- β -xylanase from *Bacteroides thetaiotaomicron* VPI-5482 (PDB entry 3kst; Joint Center for Structural Genomics, unpublished work) as the

Table 1
Data and refinement statistics for *HoAraf43*.

Values in parentheses are for the outer shell.

Data statistics	
Space group	<i>P</i> 2 ₁
Molecules per asymmetric unit	2
Wavelength (Å)	0.9795
Unit-cell parameters	
<i>a</i> (Å)	44.13
<i>b</i> (Å)	73.88
<i>c</i> (Å)	87.52
α (°)	90
β (°)	94.26
γ (°)	90
Resolution range (Å)	44.01–1.10 (1.20–1.10)
<i>R</i> _{merge} (%)	2.9 (22.4)
<i>R</i> _{meas} (%)	3.2 (24.6)
No. of measured reflections	1404460 (273638)
No. of unique reflections	213964 (46299)
Mean <i>I</i> σ (<i>I</i>)	30.2 (7.7)
Completeness (%)	94.5 (89.3)
Multiplicity	6.6 (5.9)
CC _{1/2} [†] (%)	100 (96.9)
Wilson <i>B</i> factor (Å ²)	10.0
Crystallographic refinement	
<i>R</i> factor/ <i>R</i> _{free} (%)	12.2/14.5
Total No. of non-H atoms	6121
No. of protein atoms	5026
Other molecules/ions	3 HEPES, 2 Na
No. of waters	1032
R.m.s. deviation from standard geometry	
Bond lengths (Å)	0.009
Bond angles (°)	1.33
Average <i>B</i> factors (Å ²)	
Protein	12.0
Water	26.0
Ramachandran statistics	
Allowed region (%)	96.2
Favoured region (%)	99.8
Outliers (%)	0.2

[†] CC_{1/2} is the percentage correlation between intensities from random half data sets. The values given represent correlations significant at the 0.1% level (Karplus & Diederichs, 2012). Shells with CC_{1/2} exceeding 50% have been included.

best candidate to use as a search probe for molecular-replacement calculations. Phasing was performed by molecular replacement using *AutoMR* as implemented in the *PHENIX* suite (Adams *et al.*, 2010) with one monomer of the 3kst model. A clear solution was obtained and model building was performed using *Coot* (Emsley & Cowtan, 2004) guided by 2*F*_o – *F*_c electron-density maps. *PHENIX* was used for refinement against the maximum-likelihood target, including *xyz* refinement, real-space refinement, riding H atoms and refinement of individual anisotropic temperature factors. Data-collection and refinement statistics are given in Table 1. All figures showing structural details were generated with *PyMOL* (DeLano, 2002). Coordinates and structure factors have been deposited in the PDB as entry 4qqs.

2.5. Analysis of amino-acid composition

To determine the distribution of each amino acid (on the surface or buried), the surface accessibility in Å³ was calculated using *DSSP* (Kabsch & Sander, 1983). The relative surface accessibility (RelAcc) was calculated by comparing the surface accessibility of each amino acid with the highest

value observed. According to the definition of Fukuchi *et al.* (2003), an amino acid was considered to be surface-exposed if it showed a RelAcc value of greater than 25%. The percentage of each amino-acid type in the interior or on the surface was calculated with respect to the total number of interior or surface-exposed amino acids. To allow a direct comparison with the results of Fukuchi and coworkers, the amino acids were classified as apolar (Val, Ile, Leu, Met, Phe, Trp and Tyr), polar (Asn, Gln, Ser and Thr), basic (Lys and Arg), acidic (Asp and Glu) or other (Ala, Cys, Gly, Pro and His).

3. Results and discussion

3.1. Overall structure and the active site

The final model was refined with riding H atoms at 1.10 Å resolution with R and R_{free} values of 0.122 and 0.145, respec-

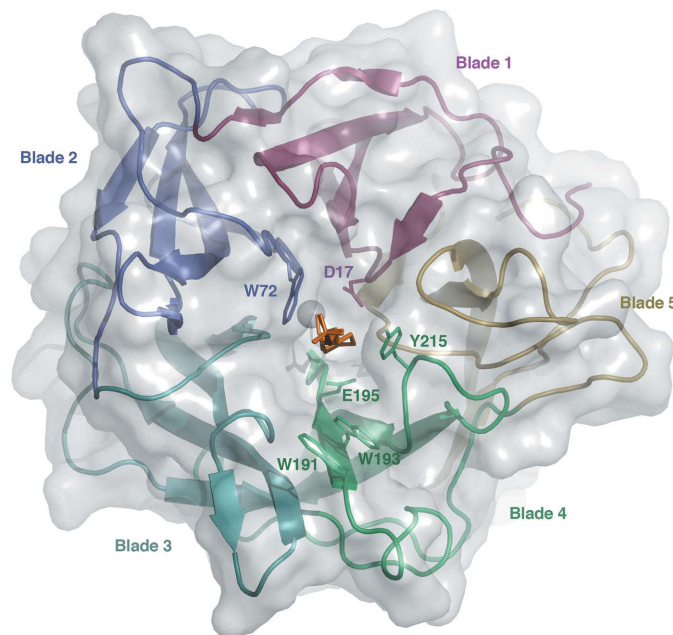


Figure 1
Ribbon representation showing the five-bladed β -propeller of *HoAraf43* with the bound HEPES molecule and the sodium ion at the central axis of the propeller.

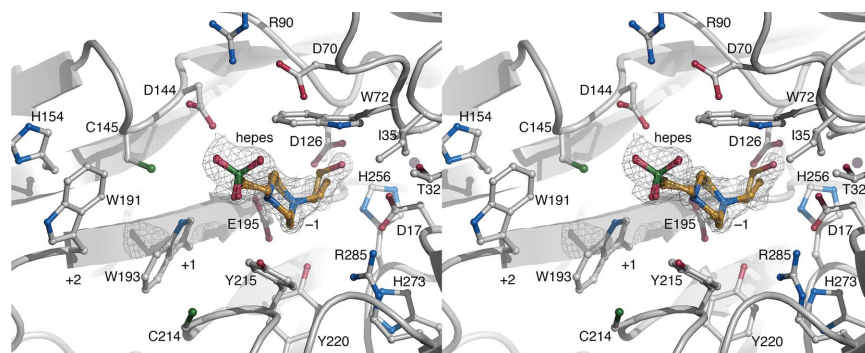


Figure 2
Wall-eyed stereo image showing the binding of a disordered HEPES molecule in subsite -1 superimposed on the σ_A -weighted $2|F_o| - |F_c|$ map contoured at 0.7σ .

tively. The deposited model includes riding H atoms, and the asymmetric unit contains two protein molecules (molecule *A*, residues 4–315; molecule *B*, residues 3–315), three HEPES molecules (one associated with molecule *A* and two with molecule *B*), one sodium ion per protein molecule and 1032 water molecules.

Based on primary and tertiary structure, the *H. orenii* glycosidase is a member of family 43 of inverting glycoside hydrolases, with accession No. c114647 and sequence cluster cd08772 (superfamily GH43_62_32_68) in the Conserved Domains Database (CDD). Specifically, the *H. orenii* GH43 glycosidase belongs to subfamily cd08991 (GH43_bXyl_2) of the GH43 family. The members of this subfamily mainly include uncharacterized GH43 glycosidases and putative β -xylosidases.

The enzyme folds as a five-bladed β -propeller similar to those described for inverting GH43 enzymes of clan GH-F (Fig. 1). Each blade is composed of a four-stranded anti-parallel β -sheet repeated five times from the N-terminus to the C-terminus: blade 1, residues 14–59; blade 2, residues 69–120; blade 3, residues 125–179; blade 4, residues 190–244; blade 5, residues 255–306. As inferred from homology to related GH43 enzymes, the catalytic -1 subsite is situated at the bottom of a pocket formed at the ‘top’ of the β -propeller (Fig. 1). The conserved catalytic nucleophile is placed on blade 4 (Glu195), the catalytic base (Asp17) on blade 1 and Asp126 on blade 3 (Fig. 2). In homologous GH43 enzymes, the carboxylate group of Asp126, which is positioned between the nucleophile and the base, has been suggested to help modulate the pK_a of the nucleophile and participate in substrate binding (Nurizzo *et al.*, 2002). On one side of the -1 subsite, Trp72 (blade 2) forms a conserved sugar-binding platform. In the present structure, the piperazine ring of the HEPES buffer molecule is wedged between Trp72 and Tyr215 in subsite -1 , but does not form a planar stacking interaction with the aromatic rings (Fig. 2). Moreover, Arg285 and Tyr215 should be able to contribute to binding in the -1 subsite. Additional subsites are present but display fundamental differences to other GH43 enzymes of known structure (see §3.5 below). Importantly, Trp193 and Trp191 (blade 4) may be able to offer interactions with a polysaccharide backbone or branching sugar units in subsites $+1$ and $+2$, respectively (Fig. 2). Besides the two tryptophan

residues, Cys214 (+1), Asp189 (+2) and Ser217 (+2) may contribute to substrate binding.

3.2. Metal-binding preference

A metal-binding site is situated at the central axis of the propeller, immediately below the -1 subsite. Two types of cation, Na^+ and K^+ , were present during crystallization of the *H. orenii* glycosidase, which are also the relevant cytoplasmic cations responsible for counterbalancing the high osmotic pressure exerted by the naturally saline medium. The metal ion present in the structure is represented by a 20σ peak in the σ_A -weighted difference Fourier map with coordination to five ligands (Fig. 3): Glu75 $\text{O}^{\epsilon 2}$ (2.1 Å) and four water O atoms. The imidazole group of the nearby His256 (blade 5) is positioned as a possible sixth ligand, but is 2.8 Å from the metal ion. The coordination distances for Na^+ -carboxylate O and Na^+ -imidazole N (target distances 2.42 and 2.38 Å, respectively; Harding, 2002) are considerably shorter than those for K^+ -carboxylate O and K^+ -imidazole N (target distances 2.84 and 2.80 Å, respectively; Harding, 2002). Additionally, Na^+ and K^+ prefer different coordination numbers, *i.e.* six and eight, respectively. Thus, based on the composition of the crystallization medium, the observed coordination number and the metal-ligand coordination distances for the bound metal, we modelled the metal ion as Na^+ . It should be emphasized that K^+ is still possible. Neither of these metals may, however, represent the biologically relevant metal for the enzyme. We also investigated the effect of various cations on the relative thermal stability against unfolding by means of Thermofluor screening (Ericsson *et al.*, 2006). The results show that the enzyme is indeed stabilized in the presence of divalent cations (Supplementary Fig. S1). At pH 7.0, adding 2 mM Mn^{2+} increases the melting temperature (T_m) of the enzyme by 5.5°C from 63 to 68.5°C (Supplementary Fig. S1a).

The metal-binding site in the *H. orenii* GH43 glycosidase is strategically positioned at the propeller axis to provide stabilizing interactions between blade 2 (Glu75) and blade 5

(His256) across the axis, which probably explains the increase in melting temperature in the presence of divalent cations. As discussed above, the metal ion bound in the crystal structure is most likely to be Na^+ or K^+ , but the binding site would be able to favourably accommodate any of the stabilizing metal ions, *i.e.* Mn^{2+} , Mg^{2+} or Ca^{2+} . The metal observed in the structure is six-coordinated with one elongated interaction between the metal and His256 $\text{N}^{\epsilon 2}$ (Fig. 3), which suggests that the physiologically and functionally relevant metal may be any of Mn^{2+} , Mg^{2+} or Ca^{2+} . Coordination number six is common for chelated Mn^{2+} , Mg^{2+} and Ca^{2+} (Harding, 2001). Ca^{2+} tends to prefer a longer coordination distance to imidazole N atoms than Mn^{2+} or Mg^{2+} , and imidazole nitrogen ligands are generally rare for Mn^{2+} (Harding, 2001). Additionally, the Glu75 $\text{O}^{\epsilon 2}$ atom coordinates the metal monodentately (specifically anti-monodentate coordination), which is especially common for Mg^{2+} , but is also common for protein ligands that coordinate Mn^{2+} and Ca^{2+} (Harding, 2001).

Some GH43 members have been confirmed to depend on divalent cations for activity. For instance, the two-domain endo-acting arabinanase from *Thermotoga petrophila* (*Tp*ABN; PDB entries 4kc7 and 4kc8; Santos *et al.*, 2014) is Ca^{2+} -dependent, but with only backbone carbonyl groups and water molecules as metal ligands. Of the available GH43 structures, only the *B. thetaiotaomicron* endo-1,4- β -xylanase with bound Na^+ (PDB entry 3kst; Joint Center for Structural Genomics, unpublished work), the exo- α -1,5-L-arabinofuranosidase *Sa*Araf43A with bound Ca^{2+} (PDB entry 3akh; Fujimoto *et al.*, 2010) and the *Clostridium acetobutylicum* β -xylosidase Abf2 with bound Mg^{2+} (PDB entry 3k1u; Midwest Center for Structural Genomics, unpublished work) show the same coordination geometry as the *H. orenii* GH43 enzyme with Glu $\text{O}^{\epsilon 2}$ and His $\text{N}^{\epsilon 2}$ as metal ligands. However, the dependence of these enzymes on various metal ions for structural stability or catalytic activity has not been reported.

Our results show that divalent cations improve the structural stability of the *H. orenii* glycosidase, especially in the presence of high salt, which agrees well with the hot, saline

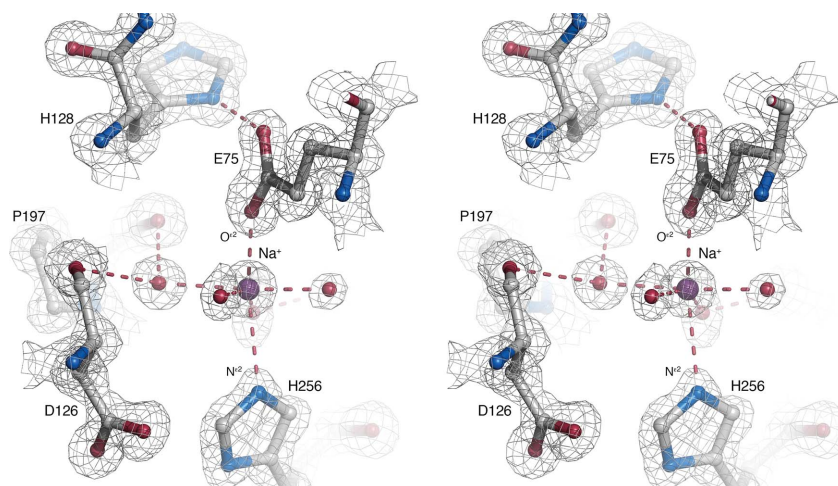


Figure 3

Wall-eyed stereo image showing the metal-binding site with a modelled sodium ion (purple) octahedrally coordinated by Glu75, His256 and four water molecules. The σ_A -weighted $2|F_o| - |F_c|$ map contoured at 2.0σ is shown.

Table 2
Specific activity of *HoAraf43* on *pNP-αAraf*.

Reaction conditions	Specific activity† (μmol min ⁻¹ mg ⁻¹)	
	Before adding Na ₂ CO ₃	After adding Na ₂ CO ₃
4 M NaCl, 2 mM Mg ²⁺ , EDTA/EGTA-treated <i>HoAraf43</i>	6.68 ± 0.32	17.38 ± 0.20
4 M NaCl, 2 mM Mg ²⁺ , untreated <i>HoAraf43</i>	6.69 ± 1.2	15.86 ± 0.56
4 M NaCl, 2 mM Mn ²⁺ , EDTA/EGTA-treated <i>HoAraf43</i>	20.32 ± 0.49	36.20 ± 5.6
4 M NaCl, 2 mM Mn ²⁺ , untreated <i>HoAraf43</i>	20.78 ± 1.7	22.89 ± 7.0

† Expressed as the arithmetic mean and the standard error of the mean.

conditions under which this enzyme operates. Taken together, the structural observations regarding the metal-binding site suggest that it is well designed for binding any of the three stabilizing divalent metal ions Mn²⁺, Mg²⁺ or Ca²⁺.

3.3. Salt tolerance

Since the *H. orenii* glycosidase is produced by a halophile, it is expected that the thermal stability against unfolding will depend on the salt concentration (Supplementary Fig. S1b). A pronounced ΔT_m value of 10°C, from 63 to 73°C, was observed at 4.0 M NaCl and pH 6.0–6.5. The T_m value is increased further by the cumulative stabilizing effects of divalent cations and NaCl (Supplementary Fig. S1c). At pH 6.0–6.5, the T_m value increases to 75–76°C in the presence of 2 mM MnCl₂ with 4.0 M NaCl. For CaCl₂ the most stabilizing condition is observed at pH 7.0 and 4.0 M NaCl, resulting in a T_m value of 74.5°C. In agreement with the optimal pH for the growth of *H. orenii*, the enzyme displays optimal stability between pH 6.5 and 7.0, which is the pH range where Mn²⁺ provides the highest degree of structural stabilization. The *H. orenii* GH43 glycosidase is an intracellular enzyme (no export signal is present), and bacteria of the order Halobacteriales are known to manage osmotic stress using the ‘salt-in’ strategy, where the molar concentrations of K⁺ and Na⁺ are maintained in the cytoplasm to match those in the surrounding medium (Oren, 2006). Thus, the increased thermal stability to unfolding at higher salt concentration agrees well with the strategy employed by *H. orenii*.

3.4. Substrate screening

The initial activity screen using various *o*- and *p*-nitrophenyl glycosides did not provide definitive answers regarding substrate specificity. Low activity was observed after an extended incubation time (24 h) with the substrates *pNP-αAraf*, *pNP-αArap*, *pNP-βArap*, *pNP-βGal*, *pNPαGlu*, *pNPβGlu* and *oNPβGlu* (Supplementary Fig. S2). These substrates were selected for a second activity screen using the conditions that showed the highest activity for each substrate: *pNP-αAraf* (4 M NaCl, Mn²⁺; 4 M NaCl, Mg²⁺), *pNP-αArap* (3 M NaCl, Mg²⁺; 0 M NaCl, Mn²⁺), *pNP-βArap* (0 M NaCl, Mg²⁺), *pNP-βGal* (1 M NaCl, Mg²⁺), *pNP-αGlu* (1 M NaCl, no metal; 0 M NaCl, Mn²⁺), *pNP-βGlu* (0 M NaCl, Mn²⁺) and

oNP-βGlu (0 M NaCl, Mn²⁺). Both EDTA/EGTA-treated and untreated enzyme were used. The results from the repeated experiment confirmed and strengthened the assumption that *pNP-αAraf* is a substrate of the *H. orenii* GH43 glycosidase (Supplementary Fig. S3).

The highest activity was obtained for *pNP-αArap* in the presence of 4 M NaCl and Mn²⁺, and activity was lost in the absence of metal. For this condition, a slightly higher activity was observed for enzyme that had been pretreated with EDTA and EGTA compared with untreated enzyme. By taking absorbance readings before and after the addition of sodium carbonate, lower and upper estimates of the specific activity for the arabinofuranosidase activity were obtained (Table 2). The specific activity is within the range ~7–17 μmol min⁻¹ mg⁻¹ for 4 M NaCl/Mg²⁺ and within the range ~20–36 μmol min⁻¹ mg⁻¹ for 4 M NaCl/Mn²⁺. These values can be compared with a commercial source of α-L-arabinofuranosidase (Megazyme), which has a specific activity of ~32 μmol min⁻¹ mg⁻¹ for *pNP-αAraf*. TLC analysis of reactions run during 3 h incubation of the enzyme with either α-L-arabinofuranopentose or lactose as a substrate did not reveal any hydrolysis products.

3.5. Structural comparison to other GH43 enzymes

Structure-based similarity analyses were performed using the refined model of the *H. orenii* glycosidase and the *DaliLite* (http://ekhidna.biocenter.helsinki.fi/dali_server; Holm & Rosenström, 2010) and *PDBFold* (<http://www.ebi.ac.uk/msd-srv/ssm>; Krissinel & Henrick, 2004) servers. In agreement with the initial *Phyre2* analysis, both *PDBFold* and *DaliLite* returned PDB entry 3kst (Joint Center for Structural Genomics, unpublished work) as the best structural match.

To further analyze the possible function and substrate specificity, available crystal structures of GH43 enzymes were examined. Although the overall fold is indeed similar, the precise details of the substrate-binding region differ significantly between the enzymes. These differences are exemplified in Supplementary Fig. S4, where the *H. orenii* glycosidase structure has been superimposed with the closest structural homologue *B. thetaiotaomicron* endo-1,4-β-xylanase (PDB entry 3kst; Joint Center for Structural Genomics, unpublished work), as well as with representative GH43 enzymes in complex with carbohydrate ligands: *C. japonicus* exo-α-1,5-L-arabinanase *CjArb43A* in complex with arabinohexose (PDB entry 1gye; Nurizzo *et al.*, 2002), *S. avermitilis* exo-α-1,5-L-arabinofuranosidase *SaAraf43A* in complex with α-L-arabinofuranotriose (PDB entry 3akh; Fujimoto *et al.*, 2010), *Bacillus subtilis* arabinoxylan α-1,3-L-arabinofuranohydrolase in complex with xylotriase (PDB entry 3c7f; Vandermarliere *et al.*, 2009), *B. subtilis* endo-α-1,5-L-arabinanase *Abn2* in complex with arabinotriose (PDB entry 2x8s; de Sanctis *et al.*, 2010), *C. japonicus* α-1,2-L-arabinofuranosidase *Abf43A* in complex with α-1,3-arabinofuranose-substituted α-1,5-L-arabinotriose (PDB entry 3qef; Cartmell *et al.*, 2011), *Humicola insolens* *HiAXHd3* in complex with xylotriase (PDB entry 3zxx; McKee *et al.*, 2012) and *Geobacillus*

Table 3
Comparison of amino-acid composition.

Amino-acid class§ (%)	Mesophiles and nonhalophiles†		Thermophiles‡		Halophiles‡		Halothermophile (<i>H. orenii</i> GH43 glycosidase‡)	
	Interior	Surface	Interior	Surface	Interior	Surface	Interior	Surface
Apolar	45.6	18.1	46.9	20.2	41.5	15.1	39.2	25.0
Polar	15.3	21.0	13.4	15.6	14.9	17.5	19.8	21.0
Basic	6.0	16.7	7.1	20.1	5.8	10.9	3.3	20.0
Acidic	7.2	21.0	7.8	23.0	8.6	28.6	6.6	23.0
Other	25.9	3.1	24.8	21.1	29.1	28.0	31.1	11.0

† Values from Fukuchi *et al.* (2003). ‡ Values that are different in the *H. orenii* GH43 enzyme with respect to the other groups are highlighted in bold. § Classification: apolar, Val, Ile, Leu, Met, Phe, Trp, Tyr; polar, Asn, Gln, Ser, Thr; basic, Lys, Arg; acidic, Asp, Glu; other, Ala, Cys, Gly, Pro, His.

stearothermophilus β -xylosidase XynB3 in complex with xylobiose (PDB entry 2exj; Br ux *et al.*, 2006).

Only four residues appear to be well conserved among the 48 GH43 structures examined (*H. orenii* GH43 sequence numbers): Glu195 (the catalytic nucleophile), Asp17 (the catalytic base), Trp72 and Asp126 (Fig. 2). Beyond these side chains, larger variations occur both in the sequence and in the conformations of the loops forming the substrate-binding region (Supplementary Fig. S4). The lack of structural consensus prevents the assignment of substrate specificity to the *H. orenii* GH43 glycosidase based on homologous structures; however, a slightly higher level of active-site structure similarity exists for the *B. thetaiotaomicron* endo-1,4- β -xylanase (PDB entry 3kst; Joint Center for Structural Genomics, unpublished work), the exo- α -1,5-L-arabinofuranosidase SaAraf43A (PDB entry 3akh; Fujimoto *et al.*, 2010) and the *C. acetobutylicum* β -xylosidase Abf2 (PDB entry 3k1u; Midwest Center for Structural Genomics, unpublished work). To conclude, we find that the only GH43 glycosidase produced by *H. orenii* shares a common catalytic mechanism with other known GH43 members, but that the substrate-binding region is distinct. The enzyme displays activity with pNP- α Araf as a substrate, but no detectable activity on pNP- α Xyl or pNP- β Xyl, which suggests that it is an α -1,5-L-arabinofuranosidase and not an α - or β -xylosidase. We therefore propose that the *H. orenii* GH43 glycosidase is referred to as HoAraf43. However, the lack of degradation products from α -L-arabinofuranopentaose and the low sequence and structural similarity of the substrate-binding region in HoAraf43 to other known GH43 enzymes, including arabinofuranosidases, indicate that although the natural substrate for HoAraf43 is likely to contain arabinofuranose units, the natural substrate is probably more complex.

3.6. Amino-acid composition analysis

The interior and exterior amino-acid distributions have been analyzed for proteins produced by bacteria and archaea that thrive in different environments (Fukuchi & Nishikawa, 2001; Fukuchi *et al.*, 2003). The polyextremophilic HoAraf43 shares features with mesophiles, thermophiles and halophiles (Table 3): (i) the percentage of hydrophobic core amino acids is similar to that of halophiles, (ii) the percentage of polar amino acids on the surface correlates with that of the average mesophilic/nonhalophilic protein and (iii) the percentages of

surface-exposed positively and negatively charged residues are identical to those of thermophilic proteins. While these similarities are noted, there are some differences (Table 3): (i) the percentage of hydrophobic residues on the surface is higher in HoAraf43, (ii) the percentage of buried polar residues is higher, (iii) the percentage of buried basic residues is lower and (iv) the distribution of other amino acids (Ala, Cys, Gly, Pro and His) in the interior is similar to halophiles, although slightly higher in HoAraf43.

4. Conclusions

Thermostable and robust glycoside hydrolases are highly relevant to a number of biotechnological applications, including lignocellulose biomass processing and bioconversion to produce value-added sugars from inexpensive bulk resources. In this work, we have shown that the only GH43 enzyme produced by *H. orenii* is unique within this family in that (i) it is polyextreme with an unusual distribution of amino acids, (ii) it displays pronounced stabilization in the presence of high salt and divalent cations, (iii) the active site shows unique details that are distinct from other hemicellulases of known structure in this enzyme family and (iv) the enzyme is able to hydrolyze pNP- α Araf. The weak but distinct activity with pNP- α Araf justifies the classification of this *H. orenii* glycosidase as an α -1,5-L-arabinofuranosidase and its naming HoAraf43. Information about the composition of the bottom sediment of Chott El Guettar is lacking, which complicates the search for the true substrate of the enzyme. However, further and wider screening of substrates may offer additional clues to the role of this intriguing enzyme and provide information regarding the potential usefulness of this enzyme in biotechnological applications.

Acknowledgements

CD acknowledges support from the Swedish Research Council Formas (Grant No. 2013-1741) and the Swedish Research Council VR (Grant No. 2013-5717). LDK acknowledges the receipt of a Griffith University Postgraduate Research Scholarship (GUPRS) and a Griffith University International Postgraduate Research Scholarship (GUIPRS) award. The authors thank the beamline staff at Diamond Light Source beamline I24 for support during data collection. This work was facilitated by the Protein Science Facility at

Karolinska Institutet/SciLifeLab (<http://psf.ki.se>). The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007–2013) under BioStruct-X (grant agreement No. 283570).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Brüx, C., Ben-David, A., Shallom-Shezifi, D., Leon, M., Niefind, K., Shoham, G., Shoham, Y. & Schomburg, D. (2006). *J. Mol. Biol.* **359**, 97–109.
- Cartmell, A., McKee, L. S., Peña, M. J., Larsbrink, J., Brumer, H., Kaneko, S., Ichinose, H., Lewis, R. J., Viksø-Nielsen, A., Gilbert, H. J. & Marles-Wright, J. (2011). *J. Biol. Chem.* **286**, 15483–15495.
- Cayol, I., Ollivier, B., Patel, B. K. C., Prensier, G., Guezennec, J. & Garcia, J. L. (1994). *Int. J. Syst. Bacteriol.* **44**, 534–540.
- DeBoy, R. T., Mongodin, E. F., Fouts, D. E., Tailford, L. E., Khouri, H., Emerson, J. B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H. J. & Nelson, K. E. (2008). *J. Bacteriol.* **190**, 5455–5463.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Doyle, S. A. (2005). *Methods Mol. Biol.* **310**, 107–113.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.
- Fujimoto, Z., Ichinose, H., Maehara, T., Honda, M., Kitaoka, M. & Kaneko, S. (2010). *J. Biol. Chem.* **285**, 34134–34143.
- Fukuchi, S. & Nishikawa, K. (2001). *J. Mol. Biol.* **309**, 835–843.
- Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M. & Nishikawa, K. (2003). *J. Mol. Biol.* **327**, 347–357.
- Harding, M. M. (2001). *Acta Cryst.* **D57**, 401–411.
- Harding, M. M. (2002). *Acta Cryst.* **D58**, 872–874.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Kelley, L. A. & Sternberg, M. J. E. (2009). *Nature Protoc.* **4**, 363–371.
- Kori, L. D. (2012). PhD thesis. Griffith University, Nathan, Queensland, Australia.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Lombard, V., Golaconda, R. H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.
- Mavromatis, K., Ivanova, N., Anderson, I., Lykidis, A., Hooper, S. D., Sun, H., Kunin, V., Lapidus, A., Hugenholtz, P., Patel, B. & Kyrpides, N. C. (2009). *PLoS One*, **4**, e4192.
- McKee, L., Pena, M. J., Rogowski, A., Jackson, A., Lewis, R. J., York, W. S., Krogh, K. B. R. M., Viksø-Nielsen, A., Skjot, M., Gilbert, H. J. & Marles-Wright, J. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 6537–6542.
- Mijts, B. & Patel, B. K. C. (2001). *Extremophiles*, **5**, 61–69.
- Nurizzo, D., Turkenburg, J. P., Charnock, S. J., Roberts, S. M., Dodson, E. J., McKie, V. A., Taylor, E. J., Gilbert, H. J. & Davies, G. J. (2002). *Nature Struct. Biol.* **9**, 665–668.
- Oren, A. (2006). *The Prokaryotes*, 3rd ed., edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt, Vol. 2, pp. 263–282. New York: Springer. doi: 10.1007/0-387-30742-7_9.
- Sanctis, D. de, Inácio, J. M., Lindley, P. F., de Sá-Nogueira, I. & Bento, I. (2010). *FEBS J.* **277**, 4562–4574.
- Santos, C. R. *et al.* (2014). *J. Biol. Chem.* **289**, 7362–7373.
- Savitsky, P., Bray, J., Cooper, C. D. O., Marsden, B. D., Mahajan, P., Burgess-Brown, N. A. & Gileadi, O. (2010). *J. Struct. Biol.* **172**, 3–13.
- Vandermarliere, E., Bourgois, T. M., Winn, M. D., van Campenhout, S., Volckaert, G., Delcour, J. A., Strelkov, S. V., Rabijns, A. & Courtin, C. M. (2009). *Biochem. J.* **418**, 39–47.