Appl Biochem Biotechnol (2011) 164:1323–1338
DOI 10.1007/s12010-011-9215-1

Identification and Characterization of a Novel Thermostable *gh-57* Gene from Metagenomic Fosmid Library of the Juan De Fuca Ridge Hydrothemal Vent

Hui Wang · Yingxue Gong · Wei Xie · Wenjuan Xiao · Junmei Wang · Yangyang Zheng · Jia Hu · Zehuan Liu

Received: 10 November 2010 / Accepted: 1 March 2011 /
Published online: 1 April 2011

Abstract A novel glycoside hydrolases family 57 gene (*gh-57*) was found from a metagenomic fosmid library constructed from a black smoker chimney sample 4143-1 from the Mothra hydrothermal vent at the Juan de Fuca Ridge. Sequence and homology analysis using BLAST revealed that it had high similarity to *gh-57* family. Conserved domain research revealed that the novel *gh-57* contained a Glyco-hydro-57 domain and five conserved regions, including two putative catalytic residues Glu^{154} and Asp^{263}. The three-dimensional features of the protein and its homologue from *Pyrococcus horikoshii* OT3 known as α-amylase were generated by homology modeling. The *gh-57* gene was cloned, expressed, and purified in *Escherichia coli* using pQE system. Enzyme activity revealed that the recombinant protein could hydrolyze soluble starch and demonstrated amylase activity. It showed an optimal pH of 7.5, an optimal temperature of 90°C, and its thermostability at 90°C could remain over 50% enzyme activity for 4 h. The enzyme activity could be increased by DTT and Mg^{2+} while an inhibitory effect was observed with EDTA, ATP, and Ca^{2+}. These results showed that the *gh-57* gene was a novel thermostable amylase from oceanic microorganisms.

Keywords Metagenomics · Thermophile · Glycosidase · Cloning · Gene expression · Enzyme activity · Amylase

Introduction

Deep-sea hydrothermal vent chimneys were formed by chemical interactions between the hot fluids and cold sea water and have steep chemical and thermal gradients [16]. These unique

Hui Wang and Yingxue Gong contributed equally to this work.

H. Wang · Y. Gong · W. Xiao · J. Wang · Y. Zheng · J. Hu · Z. Liu (✉)
Research Center for Molecular Biology, Institutes of Life and Health Engineering,
National Engineering and Research Center for Genetic Medicine, College of Life Science and Technology,
Jinan University, Guangzhou 510632, People's Republic of China
e-mail: zhliu@jnu.edu.cn

W. Xie
State Key Laboratory of Biocontrol, Sun Yat-sen University, Guangzhou 510275,
People's Republic of China

environments harbor a wild range of unknown microorganisms, especially extreme microorganisms, and show extremely high diversity. Recent researches applied both cultivation and cultivation-independent methods including high-throughput genomics technologies to study the microbial diversity in these environments [7], all of which have greatly expanded the knowledge of life forms and living adaption strategies in extreme environments on Earth.

In previous studies, a sample from the outer part of a mature sulfide chimney (4143–1) was collected at the Mothra hydrothermal field of the Juan de Fuca Ridge and the phylogenetic diversity and metabolic diversity of microbial populations were examined using GeoChip hybridization, clone libraries, and quantitative PCR [29]. The Geochip data indicated a rapid succession of microbial communities during chimney development, and also indicated that a mature chimney (sample 4143–1) harbored a metabolically highly diverse community [29]. Then a novel method combining a fosmid library construction with pyrosequencing was applied to obtain an integrated view of the metabolic potential of the complex microbial community inhabiting chimney 4143–1 [30]. This metagenomics analysis revealed that the microbial community contained lots of genes for mismatch repair, homologous recombination, carbon fixation, and nitrogen metabolism.

Among many new genes in the metagenomic library, we focused on a new member of glycoside hydrolases family 57 gene. The family 57 of glycoside hydrolases (GH-57 family) was first identified as a novel α-amylase family in 1996 [6], and it is obviously dissimilar from the main α-amylases of the family GH-13 [10]. The GH-57 family proteins originate exclusively from prokaryotic sources of extremophilic environments including many archaeal hyperthermophiles [32]. Members of GH-57 family are distinct in catalytic domain of a $(\beta/\alpha)_7$-barrel, which is remarkably different from GH-13 α-amylases family of a $(\beta/\alpha)_8$-barrel [12]. They do not exhibit any of the conserved sequence regions characteristic of the known α-amylase family [11]. Instead, bioinformatics analysis showed that five conserved sequence regions were required as well as Glu^{123} in region III and Asp^{214} in region IV of a GH-57 member from *Thermococcus litoralis* were identified as nucleophile and proton donor, respectively [8].

In addition, the GH-57 family proteins vary from less than 400 to more than 1,500 amino acid (aa) residues and show relatively high level of sequence and length diversity [32]. During the last few years, the complete genomes of many microorganisms were sequenced and the GH-57 family has expanded consequently to over 100 members. At present, the GH-57 family members exhibit the enzyme activity of α-amylase, amylopullulanase, 4-α-glucanotranserase [12], α-galactosidase [28], and branching enzyme [21]. However, most members (90%) are putative proteins and their functions have not been identified by experiments yet. As a result, this new family of glycoside hydrolases has attracted much research interests in the past decade.

In this study, a novel thermostable *gh-57* gene was identified from the metagenomic fosmid library of the Juan de Fuca Ridge hydrothemal vent. Sequence analysis revealed that the gene did not belong to the previously characterized GH-13 family, but showed high sequence and structural similarity to the less well characterized GH-57 family. This novel *gh-57* was overexpressed in *Escherichia coli* and the enzyme activity was biochemically characterized as a novel amylase.

Materials and Methods

Stains, Vectors, and Media

E. coli strain EPI300 was used for fosmid library conservation and was supplied by State Key Laboratory of Biocontrol, Sun Yat-sen University. *E. coli* JM109 and M15 were used for DNA

manipulation and gene expression, respectively. The cloning plasmid pMD18-T was purchased from Takara (Otsu, Japan) and the 6× His-tagged protein expression plasmid pQE30 was purchased from Qiagen (Hilden, Germany). The *E. coli* strains were cultivated in Luria Bertani (LB) medium (10 $g l^{-1}$ of tryptone, 5 $g l^{-1}$ of yeast extract, and 10 $g l^{-1}$ of NaCl) at 37°C.

Enzymes, Substrates, and Reagents

Restriction enzymes and T4 DNA ligase were purchased from Takara (Otsu, Japan). The 6× His-tagged protein expression, purification, and detection system QIA*expressionist*™ was purchased from Qiagen (Hilden, Germany). Bradford Protein Assay Kit was purchased from Pierce Chemical Co. Soluble starch and maltose were used as substrates and standard, respectively, for enzyme activity assays and were purchased from Sigma-Aldrich.

gh-57 Gene Fragment Screening and Sequencing in Fosmid Library

The outer layer of a mature actively venting (316°C) black smoker chimney sample 4143-1 was collected from the Mothra hydrothermal vent field during the expedition of the Alvin/Atlantis to the Juan de Fuca Ridge in 2006 [29]. From 4143-1, a fosmid library containing around 2880 clones was constructed and random pyrosequenced by State Key Laboratory of Biocontrol, Sun Yat-sen University with the GenBank SRA accession number SRA009990.1 [30].

The *gh-57* gene fragment screening was carried out by PCR using the fosmid library as a template. Based on DNA sequence analysis of the *gh-57* fragment, two primers were designed to amplify and detect the fragment: P_{fragup} 5′-AATTCTGAATTGTACAGGGT AGCCTTTTGT-3′, $P_{fragdown}$ 5′-GCTATGCCCCAGACGACCTGCGTGATCTTG-3′.

The PCR positive clone was cultivated in LB medium with chloramphenicol at the concentration of 12.5 μg ml^{-1}. Then two oligonucleotide primers as following were designed to sequence the full length of *gh-57* gene: P_{sequp} 5′-TGCAACATCAAGAAGTT AATTGTTTACTGT-3′, $P_{seqdown}$ 5′-GGCAATGCAATAGCGCAGGTGTATAACCA-3′.

Computer-Assisted Sequence Analysis and Three-Dimensional Structure Modeling

The obtained DNA and deduced amino acid sequence were analyzed using the DNASTAR program. The putative protein was compared with a variety of homologous (listed in Table 1) obtained from updated GenBank database using BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/) [4]. The signal sequence were analyzed using program SignalP online (http://www.cbs.dtu.dk/services/SignalP/) and the predictions of domains and motifs was carried out by NCBI Conserved Domain Research (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) [18–20]. The conserved sequence alignments of known GH-57 proteins was carried out using ClustalW [26].

The three-dimensional (3D) structures of the novel GH-57 and its homologues *P. horikoshii* OT3 known as a α-amylase were carried out by homology modeling using Discovery Studio 2.5 (DS2.5, Accelrys Software Inc., San Diego, CA). Briefly, the crystal structure of 4-α-glucanotranserase from *T. litoralis* (PDB code: 1K1X A) in Protein Data Bank (PDB) was selected as a template by BLAST Search module of DS2.5. The sequence alignment between the target GH-57 protein and the template (1K1X A) was performed using the Align Multiple Sequences module of DS2.5. Then the output files were served as

Table 1 The members of GH-57 family used in the present study

Enzyme (hypothetical protein)	EC	Microorganism	Abbreviation (UniProt Accession number_UniProt species code)	GenPept	Length (aa)
4-α-Glucanotransferase	2.4.1.25	*Thermococcus litoralis* (A)	O32462_THELI	BAA22063	659
PAB0118 (amyA)	ND	*Pyrococcus abyssi* GE5 (A)	Q9V298_PYRAB	CAB49100	655
PH0193 (α-amylase)	3.2.1.1	*Pyrococcus horikoshi* OT3 (A)	O57932_PYRHO	BAA29262	633
α-Amylase	3.2.1.1	*Pyrococcus furiosus* DSM3638 (A)	P49067_PYRFU	AAA72035	649
α-Amylase (amyA)	3.2.1.1	*Dictyoglomus thermophilum* (B)	P09961_DICTH	CAA30735	686
Q2RHH9	ND	*Moorella thermoacetica* (B)	Q2RHH9_MOOTA	ABC20110	811
TLR2270	ND	*Synechococcus elongates* BP-1 (B)	Q8DGP5_SYNEL	BAC09822	852
ALR0627	ND	*Anabaena* sp. PCC7120 (B)	Q8YZ60_ANASP	BAB72585	907
PAB0644	ND	*Pyrococcus abyssi* GE5 (A)	Q9V038_PYRAB	CAB49868	597
PH1023	ND	*Pyrococcus horikoshi* OT3 (A)	O58774_PYRHO	BAA30120	598
PF0870	ND	*Pyrococcus furiosus* DSM3638 (A)	Q8U2G5_PYRFU	AAL80994	597

Archaea (A) or Bacteria (B) is given in parentheses under microorganism. The abbreviations consist of the UniProt Accession numbers and UniProt species code (http://www.expasy.org/cgi-bin/speclist) (Apweiler et al. [1]).The GenPept protein identification numbers are obtained from GenBank (Benson et al. [4]).

ND not determined.

an input for the homology model. The modeling using the output alignment files was generated using the MODELER of DS2.5 [27].

Plasmid Construction

Based on the DNA sequence analysis, two primers were designed to amplify the ORF region of *gh-57*: P_{ORFup}: 5′-GCATGCATGAAACAGCATATAT-3′ incorporating a *Sph*I site (underlined) and $P_{ORFdown}$: 5′-AAGCTTCTATAGGGCGAATT-3′ incorporating a *Hin*dIII site (underlined). The amplified region was cloned into pMD18-T vector (Takara) to construct plasmid pMD-GH and further cloned into the expression vector pQE30 (Qiagen) to generate the expression plasmid pQE-GH. The resulting plasmids were identified using restriction enzyme analysis and DNA sequencing (Invitrogen Corp., Shanghai).

Recombinant Expression and Purification of GH-57

The expression plasmid pQE-GH was transformed into *E. coli* M15 and the gene expression under control of Bacteriophage T5 promoter was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) as recommended by the manufacturer (Qiagen). Expressed 6×His-GH-57 protein was purified from *E. coli* M15 by Ni-NTA columns under native condition as described in the handbook (QIAexpressionist™).

The purified protein concentration was determined using Bradford Protein Assay Kit (Pierce Chemical Corp.) and submitted to Western blot analysis and further enzyme activity assay.

Western Blot Analysis

Equal amounts (10 μg) of purified 6× His-GH57 protein were separated in 10% SDS-polyacrylamide (SDS-PAGE) gels according to Sambrook [22]. After transferring to nitrocellulose membrane (Amersham), the protein blots were blocked with 1% blocking solution as described by the manufacturer (Roche). The Anti-His HRP Conjugates was used to detect the recombinant protein following the protocol provided by the manufacturer.

Amylolytic Enzyme Activity Assay

Amylolytic enzyme activity assay was carried out using the 3,5-dinitrosalicylic acid method as described previously [3], and all the assays were carried out in triplicate. Briefly, the 150-μl reaction mixture contained 50 μl 1% soluble starch (wt./vol.) as substrate, 20 μl 250 mM Tris buffer (pH 7.5), 1 μl 1 M dithiothreitol (DTT), and 5 μg appropriately diluted enzyme. After incubation for 1 h at 90°C, 150 μl of 3,5-dinitrosalicylic acid solution contained 1% (wt./vol.) 3,5-dinitrosalicylic acid, 0.2% (vol./vol.) phenol, 0.05% (wt./vol.) Na_2SO_4, 20% KNa-tartrate, 1% (wt./vol.) NaOH was added into the mixture and incubated for 10 min at 100°C. The absorption was then measured at 575 nm, and the specific activity was calculated using maltose as a standard. A negative control sample was prepared in each assay using diluted buffer instead of enzyme in the reaction mixture One unit of amylase was defined as the amount of enzyme that liberates 1 μmol of reducing groups per min.

Effect of pH and Temperature on GH-57 Activity and Stability

The influence of the pH value on enzyme activity was determined at 90°C in Tris buffer with a pH range from 5.5 to 9.5. The influence of the temperature on the enzyme activity was measured at pH 7.5 in Tris buffer with incubation temperatures ranging from 60 to 100°C. All the enzyme activity tests used the standard assay described above.

Thermoinactivation

The recombinant enzyme was diluted at a concentration of 33 μg ml^{-1} in 250 mM Tris buffer at pH 7.5. Thermoinactivation kinetic was measured through preincubation of the enzyme at 90°C for various periods of time from 1 to 5 h, followed by determination of the residual activity with the standard enzyme assay.

Influence of Chelating Agents

The influence of DTT and cations on enzyme activity was tested using enzyme incubated with 10 to 50 mM DTT, 1 to 5 mM $MgCl_2$, 5 to 25 mM $CaCl_2$, and 50 mM NaCl, respectively.

The effect of EDTA and ATP on enzyme activity was analyzed with 1 mM ATP and 0.1 mM EDTA, respectively. In some experiments, 1 mM $MgCl_2$ or $CaCl_2$ (1 or 3 mM) was

added after 5 min. All the reactions were performed in 250 mM Tris buffer at pH 7.5 and 90°C for 1 h.

Determination of Kinetic Parameters

The kinetic experiments were carried out at pH 7.5 and 90°C using soluble starch as a substrate. The Michaelis–Menten constant (K_M) and maximal reaction rates (V_{max}) were calculated by Michaelis–Menten curve.

Results

Identification of the Novel *gh-57* Gene

A metagenomic fosmid library was constructed of a black smoker chimney sample 4143-1 from the Mothra hydrothermal vent at the Juan de Fuca Ridge by State Key Laboratory of Biocontrol, Sun Yat-sen University. The metagenomic fosmid library was sequenced and a total of 308,034 reads containing 71 MB DNA sequences were identified. This fosmid library consisted of 2,880 clones and was conserved in 38 plates [30]. The sequenced microbiome included 31,405 contigs while the *gh-57* fragment was 603 bp long with the contig number of CL2343contig. This *gh-57* fragment was screened using PCR from the fosmid library and detected from Plate 12 (Fig. 1a). Then it was further detected to the single clone from plate 12 (Fig. 1b), and the positive clone was conserved for further DNA sequencing.

The obtained DNA sequence of *gh-57* was 1,632 bp long encoding a predicted protein of 544 aa long with a molecular weight of 62.2 kDa (Fig. 2). The gene sequence has been submitted to GenBank with the accession number of HQ324240.

Computer-Assisted Sequence Analysis

Homology analysis using BLAST revealed that the novel *gh-57* showed the maximum nucleic acid sequence identity of 68% to the *gh-57* from *Geobacter* sp. M21 (GenBank accession number CP001661) and *Geobacter bemidjiensis* Bem (CP001124). The deduced protein showed the maximum amino acid sequence identity of 62% to the GH-57 from

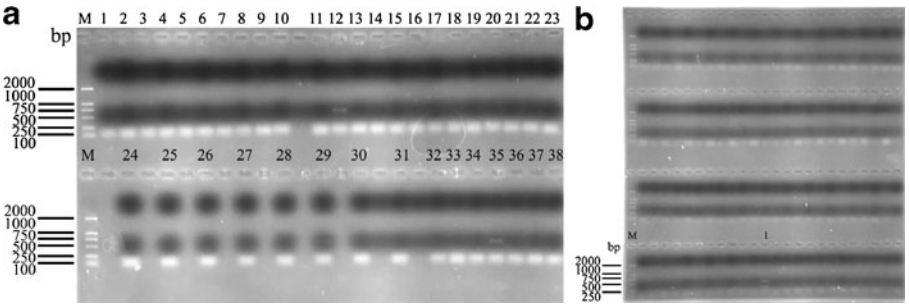

Fig. 1 Screening of *gh-57* from the metagenomic fosmid library. *Lane M*, DNA molecular weight standard DL2000. **a** The *gh-57* gene fragment PCR screening from 38 plates. *Lane 12*, PCR amplification of *gh-57* from plate 12. *Lane 35*, PCR amplification of *gh-57* using the library total DNA as a positive control. **b** The *gh-57* gene fragment PCR screening from 96 clones in plate 12. *Lane 1*, PCR amplification of *gh-57* from the positive clone in plate 12. The other lanes were negative clones in Plate 12

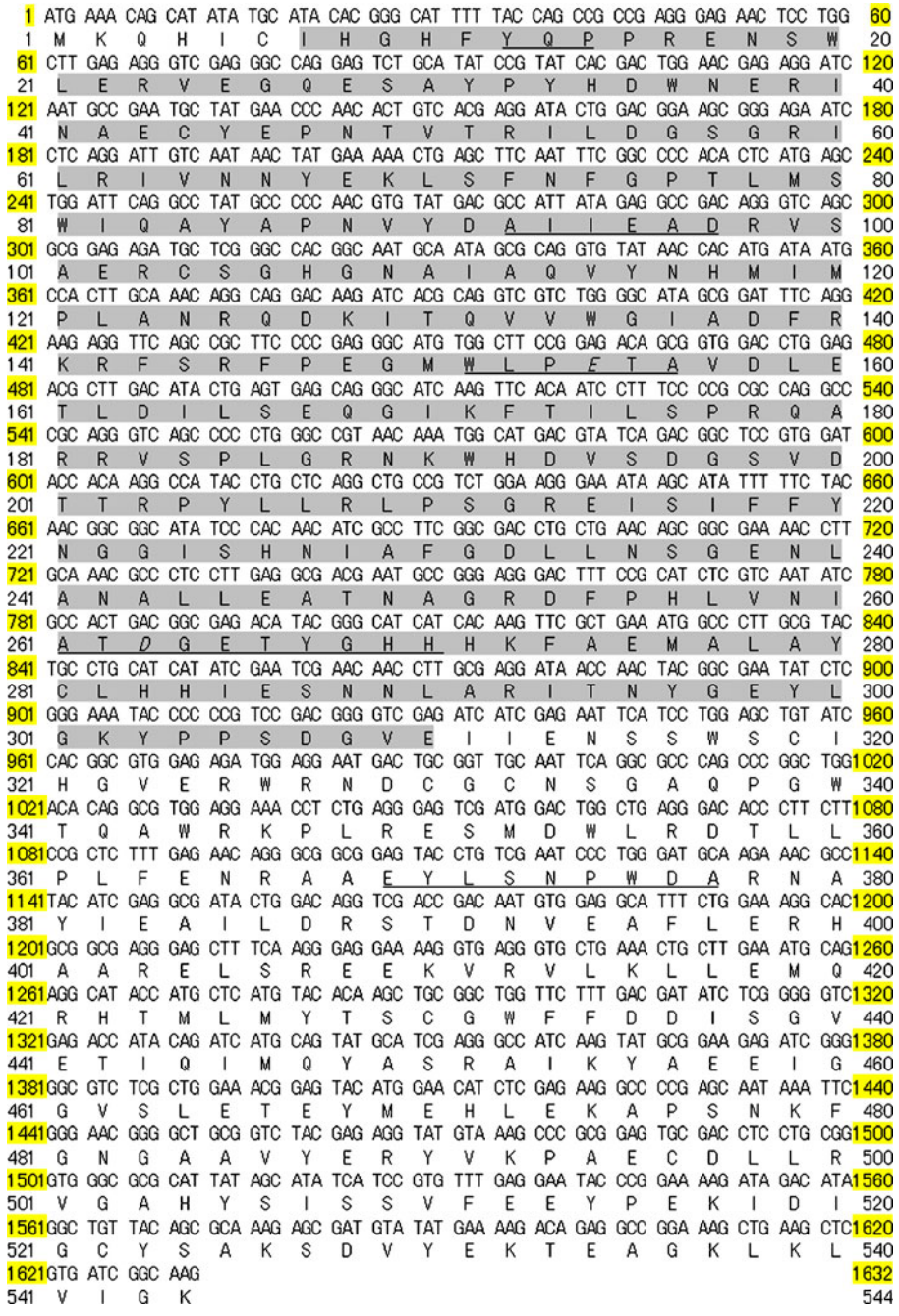

Fig. 2 The nucleotide sequence and deduced amino acid sequence of *gh-57*. The conserved Glyco-hydro-57 domain (Pfam 030065) in the N-terminal was shaded in gray. Five conserved regions were indicated underline. Two putative residues Glu^{154} and Asp^{263} were indicated in *italic*

Moorella thermoacetica (ABC20110), followed by 60% to the GH-57 from *Desulfovibrio fructosovorans* JJ (EFL49755).

Conserved domain research revealed that the novel *gh-57* contained a 304 aa long (from Ile ^{7} to Glu ^{310}) Glyco-hydro-57 domain (Pfam 030065) in its N-terminal region, which was highly conserved and showed similarity to proteins in GH-57 family. Also, a DUF3536 domain (Pfam 12055) appeared in its C-terminal and it was found in bacteria and archaea with unknown function (Fig. 2).

Multiple sequence alignments of GH-57 protein members (Fig. 3) indicated that the novel GH-57 contained five conserved domain regions including 12-YQP (region I), 92-AIIEAD (region II), 151-WLPETA (region III), 261-ATDGETYGHH (region IV), and 369-EYLSNPWDA (region V). These five regions contained functionally important residues identified in other members of GH-57 protein. According to previous studies [13, 32] and the multiple sequence alignments, two catalytic residues of the novel GH-57 are located in region III (Glu^{154}, a nucleophile) and region IV (Asp^{263}, an acid/base catalyst), which might act as nucleophile and proton donor respectively. These two catalytic residues were identified in a series of biochemical and structural studies of *T. litoralis* 4-α-glucanotransferase [8, 9] and were conserved in other GH-57 members [12, 21, 32]. In addition, three residues near the catalytic center (Glu^{95}, Trp^{151}, and Glu^{265}) were also found highly conserved in the novel GH-57 protein and might play an important role in the enzyme function.

In addition, the signal peptide analysis revealed that this protein contained a signal peptide sequence of 23 aa long, and the most likely cleavage site was between positions 23 and 24: ACA-CG (Fig. 4).

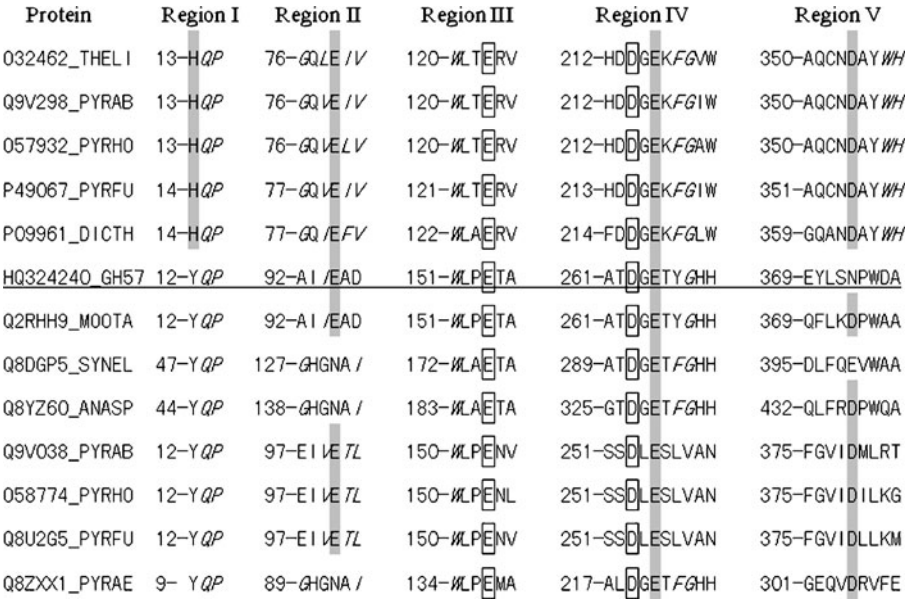

Fig. 3 Multiple sequence alignments of GH-57 protein members. Abbreviations used for the GH-57 members which are similarity with the novel GH-57 are listed in Table 1. The novel GH-57 sequence was *underlined*. Five conserved regions were indicated in regions I–V. Two putative catalytic residues (Glu and Asp) are indicated in *frame*. The potential functionally important residues (His, 2×Glu and Asp) are highlighted in *gray*. The residues conserved at a level of at least 50% are highlighted in *italic*. The enzymes and/or genes are abbreviated according to the UniProt database code (Apweiler et al. [1])

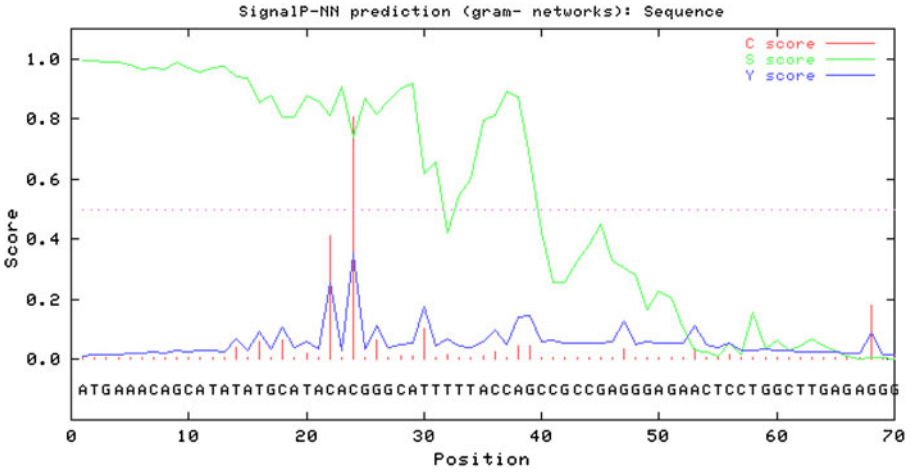

Fig. 4 Signal peptide prediction of GH-57 protein

Three-Dimensional Structure Modeling

The 3D structure of two GH-57 proteins, the novel GH-57, and a GH-57 protein from *P. horikoshii* OT3 known as a α-amylase, was carried out using the crystal structure of GH-57 from *T. litoralies* as a template (GenBank number D88253; PDB code: 1K1X). This template protein was the only GH-57 whose crystal structure was reported at present [9]. The comparison of these two proteins showed that the novel GH-57 shared a long N-terminal catalytic domain which adopted the $(\beta/\alpha)_7$-barrel fold, while the carboxyl groups of the two putative catalytic residues Glu and Asp are positioned face to face (Fig. 5). This structure was obviously different from the main and well known α-amylase family GH-13, the members of which adopt the catalytic domain of $(\beta/\alpha)_8$-barrel [12].

Cloning, Expression, and Purification of the Novel GH-57

The *gh-57* was amplified and cloned into pQE30 with 6× His tag (Fig. 6a). The recombinant plasmid pQE-GH was identified by restriction enzyme analysis (Fig. 6b) and sequencing, and then transformed into *E. coli* M15. The M15 cells were induced with IPTG and the corresponding protein band was detected by Western blot using Anti-His HRP Conjugates (Fig. 7b). This result indicated that the *gh-57* gene was successfully expressed in *E. coli*. The recombinant protein used for function studies was purified on Ni-NTA resin and detected by SDS-PAGE (Fig. 7a).

Physiochemical and Biochemical Characterization

The optimal enzyme activity of the novel GH-57 was tested using soluble starch as substrate. As shown in Fig. 8, the optimal enzyme acting condition was at 90°C and pH 7.5. Under the optimal condition (at 90°C and pH 7.5), the maximum enzymatic activity of GH-57 could reach 5,600 mU mg^{-1} in a 1-h assay. The relative enzyme activity remained above 65% at the pH range of 5.5 to 8.5 (Fig. 8a). And when temperature was lower than 80°C, the relative enzyme activity decreased distinctly to less than 45% (Fig. 8b).

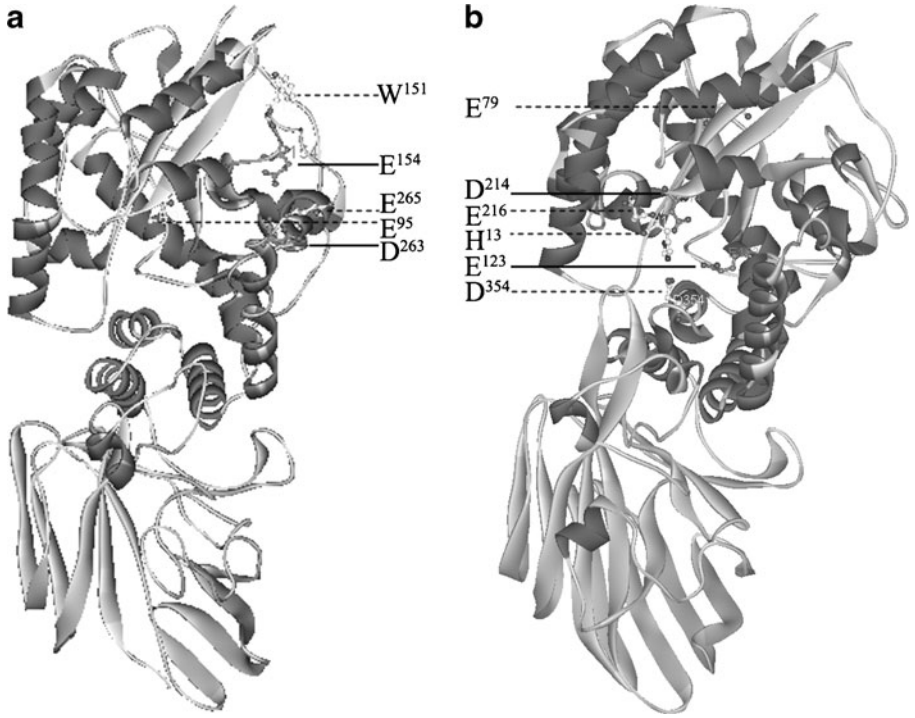

Fig. 5 3D structure of the novel GH-57 protein and the GH-57 from *Pyrococcus horikoshii*. The 3D structure models of the novel GH-57 protein and the GH-57 (α-amlyase) from *P. horikoshii* OT3 (GenBank number NP_142192) were generated by homology modeling in Discovery Studio 2.5 (Accelrys and Symyx Technologies, Inc.), using GH-57 (4-α-glucanotransferase) from *Thermococcus litoralis* (GenBank number D88253; PDB code 1K1X A) as a template. **a** Novel GH-57 protein. Residue Glu^{154} and Asp^{263} (indicated by solid line) are the two catalytic residues conserved in the members of the GH-57 family and the carboxyl groups from the two catalytic residues are positioned face to face. Three important residues (Glu^{95}, Trp^{151}, Glu^{265}) are indicated by dotted line and they locate near the catalytic residues. **b** GH-57 (α-amlyase) from *P. horikoshii*. Residue Glu^{123} and Asp^{214} (indicated by solid line) are the two catalytic residues. Four residues (His^{13}, Glu^{79}, Glu^{216}, and Asp^{354}) (indicated by dotted line) located near the catalytic residues and are considered important for the catalytic

In addition, the thermoinactivation experiments showed that after 4 h of incubation at 90°C, the enzyme retained 72% of its initial activity (Fig. 8c). These results demonstrated that this novel GH-57 protein was a thermostable amylase.

The enzyme's activity could be obviously increased in the presence of reducing agents. DTT with the concentration of 30 mM could lead to a 5.33-fold enhanced activity (Fig. 8d).

The enzyme was very sensitive to chelating agents. The addition of 0.1 mM EDTA could completely abolish the enzyme activity (Fig. 9c). The present of ATP at 1 mM concentration also had a negative effect on the activity and caused the enzyme relative activity drop to 46%. The inhibitory effect of ATP could be reversed by the addition of Mg^{2+}, but not by the Ca^{2+} even if increase the concentration of Ca^{2+} to 3 mM (Fig. 9c). Adding Mg^{2+} alone at the concentration of 5 mM, the enzyme activity slightly increased

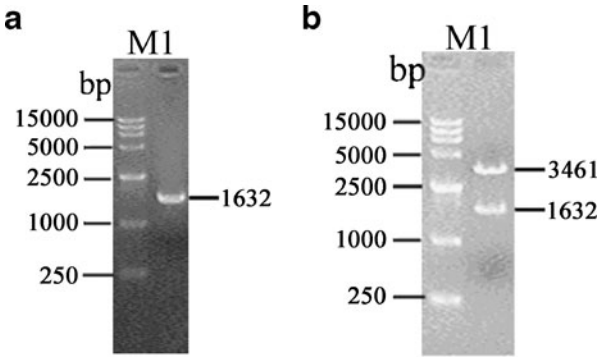

Fig. 6 PCR amplification of *gh-57* and restriction enzymes identification of recombinant expression plasmid pQE-GH. *Lane M*, DNA molecular weight standard DL15000. **a** *Lane 1*, the PCR amplification product of *gh-57* from fosmid library. **b** *Lane 1*, the restriction enzymes identification of expression plasmid pQE-GH digested by *Sph*I and *Hin*dIII

to 120% (Fig. 9a). On the contrary, adding Ca^{2+} alone at the concentration of 5 mM, the enzyme activity decreased to 49% and it decreased consequently as the concentration of Ca^{2+} rose up (Fig. 9b). The enzyme activity slightly decreased to 89% in the presence of 50 mM NaCl (Fig. 9c).

The kinetic values K_M and V_{max} of the enzyme were obtained from Michaelis–Menten curve (Fig. 10). The K_M for hydrolyzing soluble starch was 0.15 mM and the V_{max} was 0.024 μmol min^{-1}.

Discussion

Technological advances in metagenomic library construction and other molecular methods including sequencing, cloning, annotation and comparative sequence analysis have provided the possibility to explore the microbial diversity and new bioactive components [15]. These direct analysis technologies of nucleic acids have greatly improved the characterization studies of environmental samples such as uncultured oceanic microorganisms or extremophiles [14].

It is well known that there are variations in the abundance and diversity of eubacteria and archaea in the chimney of deep-sea hydrothermal vents and thus, there is potentially a wide diversity of biocatalysts from these microorganisms [5]. In a previous study, the metagenomic fosmid library constructed from a black smoker chimney of the Mothra hydrothermal vent field on the Juan de Fuca Ridge was the first report of completely

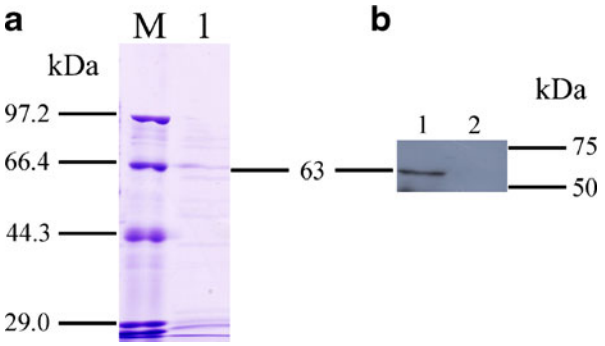

Fig. 7 SDS-PAGE of purified GH-57 and Western blot analysis of the recombinant protein. **a** *Lane M*, protein molecular weight standard; *lane 1*, the recombinant GH-57 after purification through Ni-NTA resin. **b**: *Lane 1*, Western blot analysis of the recombinant *E. coli* M15 cells induced for 3 h. *Lane 2*, the recombinant *E. coli* M15 cells without induction. Western blot was carried out using Anti-His HRP Conjugates (Qiagen)

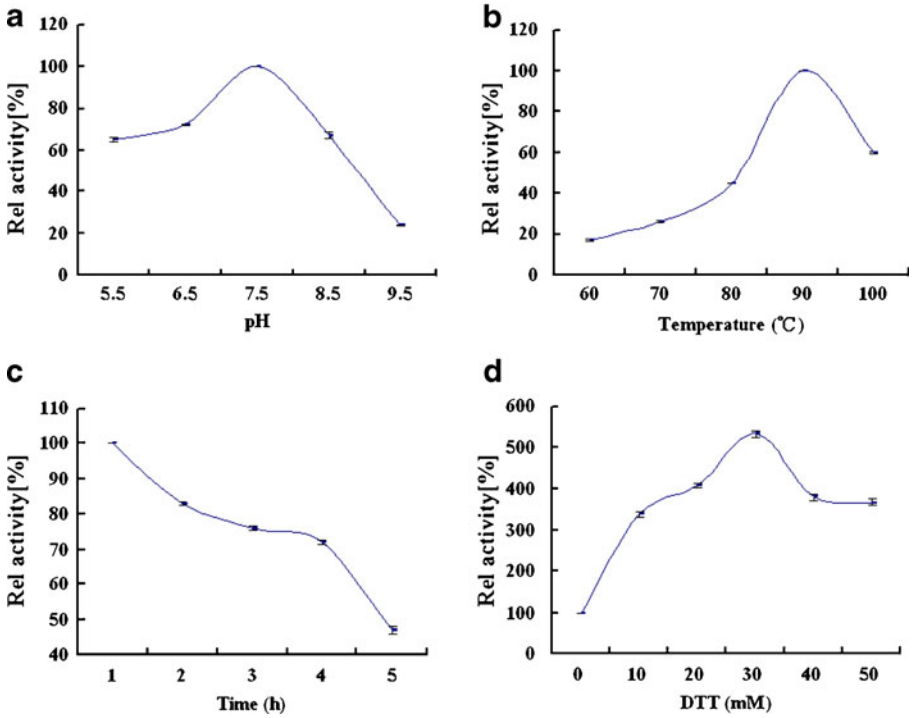

Fig. 8 Effect of pH, temperature, thermoinactivation and DTT on the GH-57 enzyme activity: **a** effect of pH on GH-57; **b** effect of temperature on GH-57; **c** long-term incubation of GH-57 at 90°C; **d** effect of DTT on GH-57 activity

sequencing a large pool of fosmid clones from an environmental sample by pyrosequencing [30]. The screening of metagenomic library for the expression of targeted genes was a culture-independent molecular method to collect genes encoding proteins with potential industrial applications [25].

In this study, it is the first time to directly identify a *gh-57* from an environmental sample using metagenomic method through large pyrosequencing and annotation. Meanwhile, other members of this *gh-57* family were identified from known prokaryotes. Sequence analysis of the nucleic acid and amino acid sequence both revealed that it showed significant similarities to GH-57 family proteins, especially to the distinct subgroup members which are uncharacterized proteins or not annotated enzymes (Fig. 3 and Table 1). Thus this novel GH-57 protein added a new member of GH-57 family, and the function and characterization were explored for the first time. Although this GH-57 demonstrated amylase activity, it showed no sequence similarity to members of the GH-13 family, which is a main and well characterized α-amylase family.

Five conserved sequence regions were found in its N-teminal and two putative catalytic residues (Glu^{154} and Asp^{263}) were indentified which present in all GH-57 proteins, whereas GH-13 proteins contain four to seven conserved sequence regions and three putative catalytic residues including two Asp and one Glu. Therefore, the GH-57 proteins comprise a unique glycoside hydrolases family and share few primary sequences similarity from GH-13 proteins. The carboxyl groups of these two catalytic residues showed a face-to-face

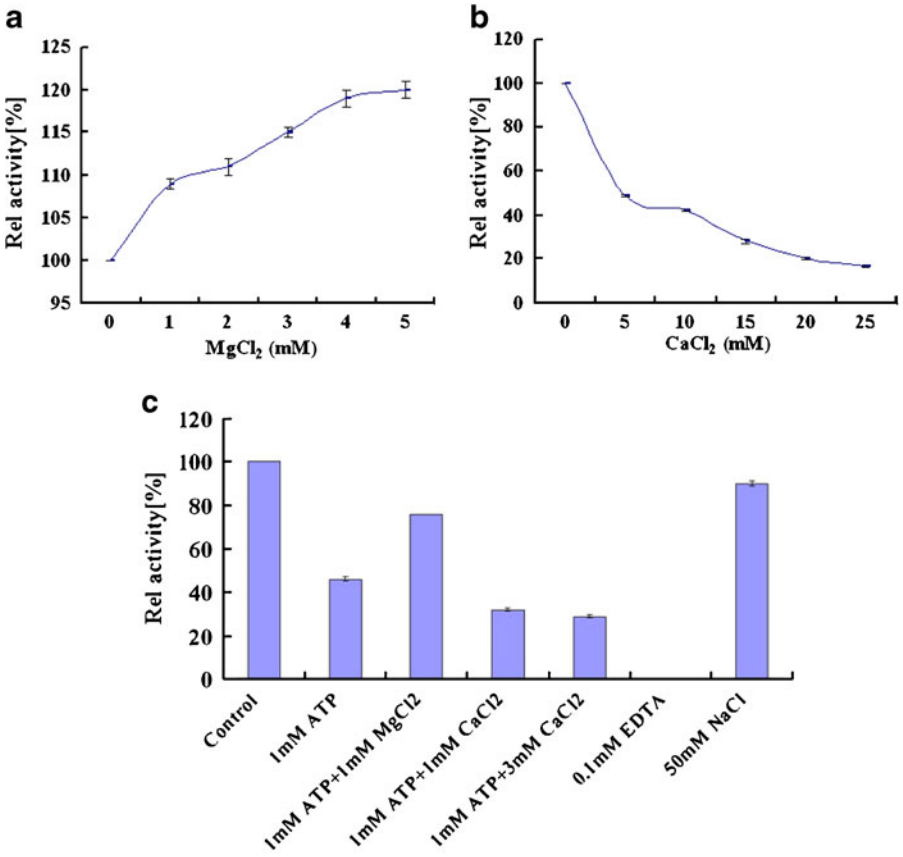

Fig. 9 Effect of EDTA, ATP, and cations on the novel GH-57 enzyme activity. **a** Effect of Mg^{2+} on GH-57; **b** effect of Ca^{2+} on GH-57; **c** effect of EDTA, ATP and cations on GH-57

position and this result was similar to the GH-57 from *Thermococcus kodakaraensis* described as a branch enzyme previously [21]. Both proteins' 3D modeling used *T. litoralis* 4-α-glucanotransferase as a template, which was the only GH-57 family member whose crystal structure has been reported [9]. The obtained 3D structures in the present study

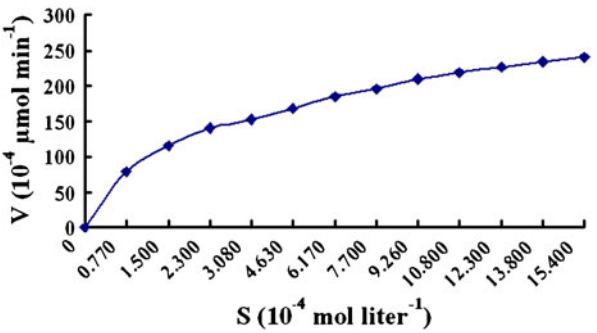

Fig. 10 Michaelis–Menten curve of the novel GH-57 amylase. *S*, substrate concentration, *V*, enzyme reaction rate

resembled the template closely, and all of them showed a $(\beta/\alpha)_7$-barrel and was different from the $(\beta/\alpha)_8$-barrel fold of GH-13 family [12].

The *gh-57* was cloned and expressed in *E. coli* but the expression level appeared not high. It could possibly because the gene contained several rare codons for *E. coli* and the pQE expression system might not be the most appropriate one for this gene expression.

Enzyme activity assay using purified recombinant protein demonstrated that it had the ability to hydrolyze soluble starch and the maximum enzyme activity could reach 5,600 mU mg^{-1} at optimal condition. It was higher than a GH-57 reported in *T. maritima* whose α-amylase activity varied from 300 to 1,600 mU mg^{-1} [3]. The substrate specificity test of the novel GH57 is still under investigation but it seemed that it could not hydrolyze sucrose or malotriose (data not shown).

The temperature effect on this enzyme activity and the thermoinactivation experiments demonstrated that the novel GH-57 was highly thermostable. The result was similar with *T. maritima* GH-57, the optimal temperature of which was 90°C [3]. This thermostability appears to be a general property in GH-57 family because the members originate exclusively from prokaryotic sources of extremophilic nature, especially archaeal hyperthermophiles. Proteins from thermophilic organisms usually exhibit substantially higher intrinsic thermostabilities than their counter parts from mesophilic organisms. Some studies reveal that the primary structure of protein is very important for thermostability and residues Arg and Tyr appear more frequently in thermophilic proteins. There are 37 arginine and 26 tyrosine residues in this novel GH-57, so these residues may affect its thermostability as well as in other GH-57 proteins.

The pH effects on the novel GH-57 and *T. maritima* GH-57 were different. The pH range of *T. maritima* GH-57 was basic from pH 7.5 to 9.0 and it had the optimal pH of 8.5 [3], while the novel GH-57 had a relatively wilder range of pH 5.5 to 8.5 and the optimal pH value was 7.5 near the neutral pH. Both the two amylase activity could be prominently increased by reducing agent DTT but strongly inhibited by EDTA and ATP. This activity decrease by ATP might be caused by the chelating effect of nucleotides with the enzyme [3]. However, the inhibitory effect of ATP could be reversed by Mg^{2+} or Ca^{2+} on *T. maritima* GH-57, while only Mg^{2+} had the same effect on the novel GH-57 (Fig. 9c). Adding Ca^{2+} alone could depress the activity and this result was different from many GH-13 amylases which are known to be Ca^{2+}-dependent enzymes [17]. Relatively few amylases are reported active at both temperatures above 65°C and alkaline pH values of 7.5 or higher. Therefore, this novel GH-57 which demonstrated thermostability and a relatively wide working pH would have considerable potential in industrial processes, e.g., in the paper industry, starch processing, and dishwashing detergents [2].

Amylases have been reported from a number of metagenomic libraries but only a few of them were characterized [23, 24, 31]. In the present study, the identification of a novel GH-57 family member from a deep-sea metagenomic library will contribute to a better understanding of a diversified glycogen biosynthesis in oceanic organisms. Moreover, these hyperthermophilic enzymes could have advantages in industry application not only because of their thermostability but also for their resistance to denaturing agents and solvents.

Acknowledgments This research was supported by the Open Project of the State Key Laboratory of Biocontrol (SKLBC08K09), the Fundamental Research Funds for the Central Universities, the Talents Start-up Foundation of Jinan University (Grant 51208056), and the 211 Grant of MOE.

References

1. Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research, 32*, D115–D119.
2. Ballschmiter, M., Armbrecht, M., Ivanova, K., Antranikian, G., & Liebl, W. (2005). AmyA, an alpha-amylase with beta-cyclodextrin-forming activity, and AmyB from the thermoalkaliphilic organism *Anaerobranca gottschalkii*: two alpha-amylases adapted to their different cellular localizations. *Applied and Environmental Microbiology, 71*, 3709–3715.
3. Ballschmiter, M., Fütterer, O., & Liebl, W. (2006). Identification and characterization of a novel intracellular alkaline alpha-amylase from the hyperthermophilic bacterium *maritima* MSB8. *Applied and Environmental Microbiology, 72*, 2206–2211.
4. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2010). GenBank. *Nucleic Acids Research, 38*, D46–D51.
5. Elend, C., Schmeisser, C., Leggewie, C., Babiak, P., Carballeira, J. D., Steele, H. L., et al. (2006). Isolation and biochemical characterization of two novel metagenome-derived esterases. *Applied and Environmental Microbiology, 72*, 3637–3645.
6. Henrissat, B., & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *The Biochemical Journal, 316*, 695–696.
7. Huber, J., Mark Welch, D., Morrison, H., Huse, S., Neal, P., Butterfield, D., et al. (2007). Microbial population structures in the deep marine biosphere. *Science, 318*, 97–100.
8. Imamura, H., Fushinobu, S., Jeon, B. S., Wakagi, T., & Matsuzawa, H. (2001). Identification of the catalytic residue of *Thermococcus litoralis* 4-alpha-glucanotransferase through mechanism-based labeling. *Biochemistry, 40*, 12400–12406.
9. Imamura, H., Fushinobu, S., Yamamoto, M., Kumasaka, T., Jeon, B. S., Wakagi, T., et al. (2003). Crystal structures of 4-alpha-glucanotransferase from *Thermococcus litoralis* and its complex with an inhibitor. *The Journal of Biological Chemistry, 278*, 19378–19386.
10. Janecek, S. (1998). Sequence of archaeal *Methanococcus jannaschii* alpha-amylase contains features of families 13 and 57 of glycosyl hydrolases: a trace of their common ancestor? *Folia Microbiol. (Praha), 43*, 123–128.
11. Janeček, Š. (2002). How many conserved sequence regions are there in the α-amylase family? *Biologia Bratislava, 57*, 29–41.
12. Janeček, Š. (2005). Amylolytic families of glycoside hydrolases: focus on the family GH-57. *Biologia Bratislava, 60*, 177–184.
13. Jeon, B. S., Taguchi, H., Sakai, H., Ohshima, T., Wakagi, T., & Matsuzawa, H. (1997). 4-Alpha-glucanotransferase from the hyperthermophilic archaeon *Thermococcus litoralis*—enzyme purification and characterization, and gene cloning, sequencing and expression in *Escherichia coli*. *European Journal of Biochemistry, 248*, 171–178.
14. Kerkhof, L. J., & Goodman, R. M. (2009). Ocean microbial metagenomics. *Deep-Sea Research II, 56*, 1824–1829.
15. Kowalchuk, G. A., Speksnijder, A. G., Zhang, K., Goodman, R. M., & van Veen, J. A. (2007). Finding the needles in the metagenome haystack. *Microbial Ecology, 53*, 475–485.
16. Kristall, B., Kelley, D. S., Hannington, M. D., & Delaney, J. R. (2006). Growth history of a diffusely venting sulfide structure from the Juan de Fuca Ridge: a petrological and geochemical study. *Geochemistry, Geophysics, Geosystems, 7*, Q07001. doi:10.1029/2005GC001166.
17. Lim, W. J., Park, S. R., An, C. L., Lee, J. Y., Hong, S. Y., Shin, E. C., et al. (2003). Cloning and characterization of a thermostable intracellular alpha-amylase gene from the hyperthermophilic bacterium *Thermotoga maritima* MSB8. *Research in Microbiology, 154*, 681–687.
18. Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., et al. (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Research, 37*, D205–D210.
19. Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., DeWeese-Scott, C., Gonzales, N. R., Gwadz, M., et al. (2007). CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Research, 35*, D237–D240.
20. Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Research, 32*, W327–W331.
21. Murakami, T., Kanai, T., Takata, H., Kuriki, T., & Imanaka, T. (2006). A novel branching enzyme of the GH-57 family in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *Journal of Bacteriology, 188*, 5915–5924.
22. Sambrook, J., & Russell, D. (2001). *Molecular cloning: a laboratory manual* (3rd ed.). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

23. Schmeisser, C., Steele, H., & Streit, W. R. (2007). Metagenomics, biotechnology with non-culturable microbes. *Applied Microbiology and Biotechnology, 75*, 955–962.
24. Sharma, S., Khan, F. G., & Qazi, G. N. (2010). Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. *Applied Microbiology and Biotechnology, 86*, 1821–1828.
25. Short, J. M. (1997). Recombinant approaches for accessing biodiversity. *Nature Biotechnology, 15*, 1322–1323.
26. Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research, 22*, 4673–4680.
27. Tiwari, M., & Lee, J. K. (2010). Molecular modeling studies of L-arabinitol 4-dehydrogenase of *Hypocrea jecorina*: its binding interactions with substrate and cofactor. *Journal of Molecular Graphics & Modelling, 28*, 707–713.
28. van Lieshout, J. F. T., Verhees, C. H., Ettema, T. J. G., van der Sar, S., Imamura, H., Matsuzawa, H., et al. (2003). Identification and molecular characterization of a novel type of α-galactosidase from *Pyrococcus furiosus*. *Biocatalysis and Biotransformation, 21*, 243–252.
29. Wang, F., Zhou, H., Meng, J., Peng, X., Jiang, L., Sun, P., et al. (2009). GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 4840–4845.
30. Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S. M., Meng, J., et al. (2010). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *The ISME Journal*. doi:10.1038/ismej.2010.1144.
31. Yun, J., Kang, S., Park, S., Yoon, H., Kim, M. J., Heu, S., et al. (2004). Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Applied and Environmental Microbiology, 70*, 7229–7235.
32. Zona, R., Chang-Pi-Hin, F., O'Donohue, M. J., & Janecek, S. (2004). Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. *European Journal of Biochemistry, 271*, 2863–2872.