Cloning and in silico study of an endoglucanase from a thermophilic bacterium isolated from a hydrothermal vent of West Kawio, Sangihe-Talaud waters, North Sulawesi, Indonesia

Edvan Arifsaputra Suherman¹, Maelita Ramdani Moeis¹, and Elvi Restiawaty²

¹School of Life Sciences and Technology, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia
²Faculty of Industrial Technology, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia

*Corresponding author: edvan.suherman@gmail.com

SUBMITTED 29 July 2019  REVISED 16 September 2019  ACCEPTED 17 December 2019

ABSTRACT Endoglucanase is used in industries that apply high temperatures, such as bioethanol, detergent, paper, and animal feed industries. Most available endoglucanases have very low stability at high temperatures. Therefore, this study aimed to identify a new thermostable endoglucanase that is able to maintain its activity at high temperatures. Five isolates of thermophilic bacteria were previously isolated from the hydrothermal vent of West Kawio, Indonesia. Among them, the DS12 isolate showed the highest endoglucanase activity, and was identified and named as Bacillus safensis DS12. The EgDS12 gene was cloned from B. safensis DS12. EgDS12 is 1851 bp long encoding a protein of 616 amino acids. The encoded protein, EgDS12, has high sequence identity to other B. safensis endoglucanases and was predicted with the Compute pI/Mw tool to be 69.41 kDa. EgDS12 was high in hydrophobic amino acids. The enzyme had higher percentage of Ala and Pro, and lower percentage of Gly compared to thermolabile endoglucanases from two Bacillus species. EgDS12 harbored a catalytic domain belonging to glycosyl hydrolase family 9 (GH9) and a type 3 cellulose-binding domain (CBM3). Properties of endoglucanases with GH9-CBM3 modular organization include activity over a wide pH range, high optimum temperature, and thermostability. Therefore, EgDS12 has potential applications in the industries.

KEYWORDS cloning; endoglucanase; thermostable; hydrothermal vent; in silico

1. Introduction

Cellulose is the most abundant, renewable biopolymer on earth and the dominating waste material from agriculture (Patel et al. 2014). A promising strategy for efficient utilization of this renewable resource is hydrolysis with cellulase to release glucose, which in turn can be fermented to produce ethanol (Sukumaran et al. 2005). Cellulase is an enzyme system that consists of three types of enzymes that work synergistically to hydrolyze cellulose into glucose, i.e. endoglucanase (EC 3.2.1.4), exoglucanase (EC 3.2.1.74), and β-glucosidase (EC 3.2.1.21) (Lynd et al. 2002). Endoglucanase (Eg) hydrolyzes cellulose by randomly cleaving glycosidic bonds resulting in oligosaccharides and thus new chain ends. Exoglucanase or cellobiohydrolase hydrolyzes the ends of the oligosaccharide chains, releasing cellobiose. β-glucosidase finally hydrolyzes cellobiose to glucose. Thus, among the three cellulases, Eg, also called carboxymethyl cellulase (CMCase), is a key enzyme that plays an important role for initiating cellulase hydrolysis (Wood and Bhat 1988; Zhang et al. 2006; Gupta and Verma 2015). If Eg biosynthesis is increased, more cellulose will be hydrolyzed to glucose which will be finally converted to ethanol through fermentation (Srinivas and Panda 1998).

In addition to the bioethanol industry, cellulases are also used in detergent, paper, food, and animal feeds industries (Sukumaran et al. 2005). These industries apply high temperatures for easy mixing of materials, increased substrate solubility, and decreased risk of contamination (Turner et al. 2007). However, high temperature is a factor limiting the wide industrial use of cellulases. High temperature may cause protein denaturation and loss of catalytic function. Most of the available cellulases have very low activity at high temperatures. Therefore, new cellulases with thermostable property need to be identified (Zarafeta et al. 2016).

New Egs were reported from various thermophilic bacteria. Li et al. (2008) obtained thermostable Eg which had an optimum temperature at 50°C from Bacillus subtilis DR which was isolated from a hot spring. Yang et al. (2010) obtained Eg with maximum activity at 60°C from B. subtilis strain I15 which was isolated from compost. Moeis et al. (2014) obtained Eg which had an optimum activity at 50°C from Bacillus sp. RP1 isolated from a
2. Materials and Methods

2.1. Growing DSI thermophilic bacterial isolates

The thermophilic bacterial isolates DSI1, DSI2, DSI3, DSI4, and DSI5 were grown by taking 30 µL bacterial suspensions from the glycerol stocks, then inoculating them onto agar media (90% of Luria-Bertani medium [LB: tryptone 10 g/L, yeast extract 5 g/L, NaCl 10 g/L], 10% of Bushnell-Haas marine salts broth [BHMS: MgSO$_4$ 0.2 g/L, CaCl$_2$ 0.02 g/L, KH$_2$PO$_4$ 1 g/L, (NH$_4$)$_2$HPO$_4$ 1 g/L, KNO$_3$ 1 g/L, FeCl$_3$ 0.05 g/L], MgSO$_4$ 0.8 g/L, agar 15 g/L) and into 20 mL broth media (90% of LB, 10% BHMS, MgSO$_4$ 0.8 g/L) in 100 mL flasks. The cultures were incubated at 50°C until growth was detected. The liquid cultures were shaken at 150 rpm during incubation.

2.2. Screening for Eg activity

Eg activity of the DSI bacterial isolates was screened using the Congo red plate assay based on Teather and Wood (1982) and Sheng et al. (2012) with modifications. A single colony of each isolate was taken with a sterile toothpick and spotted in the middle of screening medium in the Petri dish. The screening medium used was a mixture of 90% of LB, 10% of BHMS, MgSO$_4$ 0.8 g/L, agar 15 g/L and 1% (m/v) carboxymethyl cellulose (CMC) as endoglucanase substrate. The screening was carried out in triplicate for each isolate. The media were incubated at 50°C for 48 h. The diameters of bacterial colonies were measured.

The media were then dripped with 0.1% (m/v) Congo red solution until the entire surface of the media was submerged, allowed to stand for 15 min to stain the media, then the Congo red solution was removed. Next, the media were dripped with 1M NaCl solution until the entire surface was submerged, allowed to stand for 15 min, then the NaCl solution was removed. The diameters of the clear zones around the colonies were measured. The isolate that had the largest cellulolytic index (ratio of the clear zone diameter to the colony diameter) was selected to be further analyzed.

2.3. DNA isolation

Isolation of genomic DNA of the DS12 isolate was carried out by the chloroform-isoamyl alcohol method. A single colony of DS12 bacterial isolate was inoculated into 15 ml LB broth in a 50 mL centrifuge tube, incubated at 50°C, 150 rpm, 18 h. One mL of the culture was loaded into a 1.5 mL microcentrifuge tube, centrifuged at 18,800g for 1 min, the supernatant was discarded. This step was performed six times. The final cell pellet was resuspended with 750 µL lysis buffer (25 mM EDTA, 50 mM Tris-Cl, 0.5% SDS), vortexed, then 750 µL chloroform-isoamyl alcohol (24:1) was added. The mixture was incubated at -80°C for 10 min. The tube was centrifuged at 18,800 g, 3 min. The aqueous phase was transferred into a new 1.5 mL microcentrifuge tube, 750 µL chloroform-isoamyl alcohol (24:1) was added and the tube was centrifuged at 18,800g for 3 min. The aqueous phase (500 µL) was transferred into a new 1.5 mL microcentrifuge tube. This chloroform-isoamyl alcohol extraction was repeated once again, 400 µL aqueous phase was transferred into a new 1.5 mL microcentrifuge tube and 40 µL of 0.8 M LiCl and 1 mL absolute ethanol were added to the tube. The mixture was incubated at -20°C for 30 min. Afterwards, the tube was centrifuged at 18,800 g for 3 min. The supernatant was discarded, 200 µL of 70% ethanol was added to the tube, centrifuged at 18,800 g for 3 min. After the supernatant was discarded, the dried DNA pellet was dissolved in 50 µL TE-RNase buffer pH 8.0. The DNA solution was stored in -20°C. Plasmid DNA isolation was conducted us-
2.4. Gel electrophoresis

Electrophoresis was performed at 100 V for 30 min, using 1% agarose gel, 1X TAE buffer solution, 10 µL DNA, and 3 µL GeneRuler™ 1 kb DNA Ladder (Thermo Scientific, Vilnius, Lithuania). The gel was then immersed in ethidium bromide solution (2 µg/mL) for 5-10 s and rinsed in distilled water for 5-10 min. DNA bands were observed with a UV transilluminator.

2.5. Polymerase chain reaction (PCR)

The 16S rRNA gene of DSI2 isolate was PCR amplified at Macrogen, Inc., South Korea, using universal primers 27F (5’-AGAGTTTGATCMTGGCTCAG-3’) and 1492R (5’-TACGGYTACCTTGTTACGACTT-3’). Endoglucanase gene EgDSI2 from DSI2 isolate was PCR amplified in our lab. with thermocycler (Applied Biosystems 2720) using custom-designed forward and reverse primers (see Materials and methods 2.9). The PCR was conducted following the protocol for Q5® High-Fidelity Master Mix (New England Biolabs, Ipswich, USA). The reaction consisted of 1X Q5® High-Fidelity Master Mix, 0.5 µM forward primer, 0.5 µM reverse primer, 560 ng template DNA and nuclease-free water (Promega, Madison, USA) to a total volume of 50 µL. The thermocycling conditions were initial denaturation at 98°C for 30 s followed by 35 cycles of denaturation at 98°C for 10 s, annealing at 70°C for 30 s, elongation at 72°C for 1 min and final elongation at 72°C for 2 min. The annealing temperature was determined by the Tm Calculator program version 1.11.0 (https://tmcalculator.neb.com). The PCR product was confirmed by agarose gel electrophoresis.

2.6. DNA purification

Purifications of DNA from agarose gels were performed using GenepHlow™ Gel/PCR Kit DFH100 (Geneaid, New Taipei City, Taiwan) according to the kit’s manual.

2.7. DNA sequencing

Sequencing of the 16S rRNA and EgDSI2 genes were carried out at Macrogen, Inc., South Korea. The sequencing reactions were performed by the Sanger big dye terminator method using PRISM 3730XL Analyzer ABI sequencer. Primers for sequencing of the 16S rRNA gene were the universal 27F and 1492R. Primers for sequencing of the EgDSI2 were T7 promoter (5’-TAATACGACTCACTATAGGG-3’) and T7 terminator (5’-GCTAGTTATGGCTACGCGG-3’).

2.8. Phylogenetic analysis

Ten sequences of 16S rRNA genes that had the highest homologies with the DS12 isolate were taken as ingroup sequences and one 16S rRNA sequence from Geobacillus stearothermophilus strain BCRC 10285 was used as out-group. Multiple sequence alignment (MSA) analysis was performed with the ClustalW version 2.1 program (Larkin et al. 2007). The MSA result in FASTA format was opened with the BioEdit version 7.2.5 program (Hall 1999) to carry out trimming of ambiguous bases at the beginning and the end of MSA until all sequences started and ended at the same site. Phylogenetic analysis was conducted with MEGA version 5 program (Kumar et al. 2018) using the Neighbor-Joining tree construction method. The level of confidence in the phylogenetic tree was evaluated by the Bootstrap method with 1000 repetitions.

2.9. EgDSI2 primer design

Primers for EgDSI2 were designed based on ten Eg sequences of the same length (616 amino acids) from Bacillus safensis strains (Table 1) using the CEMAsuite version 2.0.9 program (Lane et al. 2015). The ends of the consensus coding sequence (CDS) were chosen as the primer sequences to isolate the whole gene (start codon - CDS - stop codon) with a length of 1851 bp. The 5’ end of the forward primer was given extra bases (GCAATAGA) and a NdeI restriction site (CA↓TATG). The 5’ end of the reverse primer was given extra bases (TCGT) and a BamHI restriction site (G↓GATCC). The forward and reverse primers were FEG 5’-GCAATAGACATATGCATTGCTACGCGGAG and R EG 5’-GCTAGTTATGGCTACGCGGAG.
Specificity of the primers was tested with the Primer-BLAST program (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Ye et al. 2012). Quality of the primers was evaluated with the Clone Manager 9 program.

2.10. Construction of recombinant plasmid

Construction of the recombinant plasmid was carried out using the restriction-ligation method. The EgDSI2 and pET-32b vector (Novagen, United States) were double-digested with the NdeI and BamHI restriction enzymes (Thermo Scientific, Vilnius, Lithuania). The optimal condition for the double digest reactions was determined using the DoubleDigest Calculator-Thermo Scientific program (https://www.thermofisher.com/id/en/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-scientific/double-digest-calculator-thermo-scientific.html). The pET-32b and the PCR product of EgDSI2 were double digested with 2X final concentration of buffer Tango (Thermo Scientific, Waltham, USA), 20 U NdeI and 20 U BamHI. The reactions were incubated at 37°C for 15 min. The whole double digests were loaded into wells of a 0.8% agarose gel for electrophoresis. DNA bands with the desired sizes were extracted from the gel.

The double-digested EgDSI2 (12 µL) and pET-32b (5 µL) were ligated with T4 DNA ligase buffer (Thermo Scientific, Waltham, USA) and 1 Weiss U T4 DNA ligase (Thermo Scientific, Waltham, USA). The ligation reaction mixture was incubated at 4°C in a water bath for 16 h. The ligation product was named pET-EgDSI2.

2.11. Transformation of Escherichia coli BL21 (DE3) competent cells with pET-EgDSI2

A tube of competent E. coli BL21 (DE3) cells was taken from −80°C and thawed on ice for 5 min. The ligation reaction mixture was centrifuged briefly, then 5 µL ligation reaction mixture and 50 µL competent cells were transferred aseptically into a cooled 1.5 mL microcentrifuge tube. The tube was flicked with fingers five times to mix the competent cells and the ligation reaction mixture. The tube was incubated on ice for 30 min, heated at 42°C for 90 s, then re-incubated on ice for 5 min. LB broth 300 µL was added.

### TABLE 2 Bioinformatics programs used for in silico analysis of the nucleotide and amino acid sequences of endoglucanase EgDSI2.

| No | Program | Function | Source |
|----|---------|----------|--------|
| 1  | BioEdit ver. 7.2.5 (Hall 1999) | Editing of biological sequences and alignment. | www.mbio.ncsu.edu/BioEdit/bioedit.html |
| 2  | BLAST (Altschul et al. 1990) | Comparing nucleotide/protein sequences to sequence databases and calculating the statistical significance of matches. | https://blast.ncbi.nlm.nih.gov/Blast.cgi |
| 3  | ClustalX ver. 2.1 (Larkin et al. 2007) | Performing multiple sequence alignment (MSA) to find conserved/unconserved regions. | www.clustal.org |
| 4  | Clustal Omega (Sievers et al. 2011) | Performing MSA with numbers indicating positions of residues. | www.ebi.ac.uk/Tools/msa/clustalo/ |
| 5  | MEGA X (Kumar et al. 2018) | Inferring phylogenetic trees. | www.megasoftware.net |
| 6  | SnapGene ver. 2.3.2 | Simulating gene cloning, PCR, and gel electrophoresis. | GSL Biotech, snapgene.com |
| 7  | GC Content (%) Calculator (Mc Gowan 2015) | Calculating GC content (%) of a nucleotide sequence. | https://jamiemcg.github.io/bioinf/gc.html |
| 8  | EMBOSSTranseq (Rice et al. 2000) | Translating nucleotide sequence to peptide sequence. | https://www.ebi.ac.uk/Tools/st/emboss_transeq/ |
| 9  | InterProScan (Jones et al. 2014) | Scanning a sequence for matches against the InterPro protein signature databases. | www.ebi.ac.uk/interpro/search/sequence-search |
| 10 | ScanProsite (De Castro et al. 2006) | Scanning protein sequences for the occurrence of patterns, profiles and motifs against the PROSITE database. | https://prosite.expasy.org/ |
| 11 | Compute pi/Mw (Gasteiger et al. 2005) | Predicting protein isoelectric point and molecular weight. | web.expasy.org/compute_pi/ |
| 12 | ProtParam (Gasteiger et al. 2005) | Computing amino acid composition. | web.expasy.org/protparam/ |
| 13 | MyDomains (Hulo et al. 2007) | Creating custom domain figures. | https://prosite.expasy.org/mydomains/ |
| 14 | PredictProtein (Yachdav et al. 2014) | Predicting protein secondary structures composition. | www.predictprotein.org |
| 15 | Sequence Annotated by Structure (SAS) (Milburn et al. 1998) | Annotating protein sequences. | www.ebi.ac.uk/thorntonsrv/databases/sas/ |
| 16 | SWISS-MODEL (Waterhouse et al. 2018) | Modeling of 3D protein structures based on homology. | www.swissmodel.expasy.org/interactive |
to the tube and incubated at 37°C, 200 rpm, 2 h. The culture was spread onto LB agar + ampicillin (100 µg/mL) and incubated upside down at 37°C for 16 h. Growing colonies were cultured for plasmid isolation.

2.12. In silico analysis of nucleotide and amino acid sequences of EgDSI2

Bioinformatics programs used for in silico analysis are listed in Table 2. The EgDSI2 sequence was analyzed in silico to determine its homology and GC percentage. The deduced amino acid sequence, EgDSI2, was analyzed in silico to predict its size and molecular mass, to classify its family, to detect its domains, to determine the amino acid composition, to predict its secondary structures composition, and to model its 3D structure based on homology. Further analysis included a comparison of amino acid compositions of thermostable EgDSI2, thermostable Eg from Bacillus sp. KSM-635 (PDB ID 1G0C), thermostable Eg from mesophilic Bacillus agaradhaerens (PDB ID 7A3H) and thermostable Eg from mesophilic B. licheniformis (PDB ID 2JEN) to find the abundance of certain amino acids in the thermostable Egs.

3. Results and Discussion

3.1. Growing DSI thermophilic bacterial isolates

Thermophilic bacterial isolates DSI2 and DSI5 showed growth in broth and agar media after overnight incubation. DSI3 isolate showed growth in broth medium on the second day, and on agar medium on the fourth day. Meanwhile, DSI1 and DSI4 isolates did not show any growth until the seventh day.

3.2. Ability of DSI bacterial isolates in producing Eg

The Congo red plate assay revealed that isolates DSI2, DSI3, and DSI5 had the ability to produce Eg as indicated by the presence of clear zones around the colonies. The formation of clear zones indicated that Eg was secreted by the bacteria (Sheng et al. 2012). The three isolates formed clear zones with different diameters indicating that they had different capability in degrading the CMC substrate, thus show different Eg activities. The cellulolytic indexes of the three isolates are summarized in Table 3. DSI2 isolate had the highest cellulolytic index (about 2 times greater) compared to the other two isolates (Table 3). According to Jang and Chen (2003), the diameter of the clear zone is generally proportional to Eg activity. Therefore, DSI2 isolate was selected to be further analyzed.

3.3. Identity of DSI2 isolate based on 16S rRNA gene and phylogenetic analyses

Phylogenetic analysis was performed to determine the phylogenetic relationship of DSI2 isolate with other bacterial species based on the similarity of the 16S rRNA gene sequences. The contig sequence obtained from the sequencing of the 16S rRNA gene of the DSI2 isolate had a size of 1340 bp (GenBank accession no. MN726487). Identity search with BLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al. 1990) showed that the 16S rRNA gene sequence of DSI2 isolate had 99.78% identity with Bacillus safensis strain NBRC 100820 and B. safensis FO-36b. The phylogenetic tree showed that the DSI2 isolate formed a cluster with Bacillus safensis strain FO-36b and B. safensis strain NBRC 100820 (Figure 1). This showed that the DSI2 isolate had the closest relationship with B. safensis. These results confirmed that the DSI2 isolate is B. safensis and it is named as B. safensis DS12.

The genus Bacillus is a group of aerobic/facultative aerobic, rod-shaped, endospore-forming bacteria (Slepecky and Hemphill 2006). The species B. safensis was first isolated from spacecraft and assembly-facility (SAF) surfaces at the Jet Propulsion Laboratory, USA. The species name safensis was derived from SAF abbreviation (Satomi et al. 2006). This species lives in a variety of habitats, such as spacecraft and related environments, deserts, industrial waste, oil-polluted environments, compost, pond water, marine sediments, rhizosphere, insect gut, plants, fermented foods, human and animal feces, and soil (Lateef et al. 2015). Several studies have reported the ability of B. safensis in producing Eg. Khianggam et al. (2014) isolated Eg-producing B. safensis PJ1-24S from palm oil by-product. Kanchanadumkerng et al. (2017) isolated Eg-producing B. safensis M3 from freshwater swamp forest soil.

3.4. Isolation of the EgDSI2 gene

Isolation of EgDSI2 was carried out by PCR method. The PCR resulted in a single DNA band with a size of around 1.8 kb (Figure 2) which corresponded to the expected size of the gene, i.e., 1851 bp.

3.5. Cloning of EgDSI2

The double-digested EgDSI2 resulted in a single DNA band with a size of around 1.8 kb. The double-digested pET-32b vector resulted in a single DNA band with a size of around 5.5 kb. The intact (not digested) pET-32b vector has a size of 5,899 bp. The double-digest reaction of pET-32b caused a fragment of 492 bp to be cut out off the backbone, forming a linear DNA with a size of 5407 bp. The single DNA band of the double-digested pET-32b had the expected size of more than 5.0 kb.

The double-digested products were ligated, and the ligation product (pET-EgDSI2) was used to transform E. coli BL21 (DE3) competent cells. Plasmid DNA was iso-

| TABLE 3 Cellulolytic index of DSI isolates. |
|-----------------|------------------|
| Isolates       | Cellulolytic Index (*) |
| DSI2           | 2.42 ± 0.25       |
| DSI3           | 1.11 ± 0.10       |
| DSI5           | 1.32 ± 0.37       |

*) Cellulolytic index = Ratio of clear zone diameter to colony diameter
lated from the transformants to confirm the presence of the pET-EgDSI2 construct. Confirmation of the construct was carried out by restriction-digestion and PCR methods. Digestion with BamHI resulted in a single DNA band with a size of around 7 kb (Figure 3a) which corresponded to the expected size of pET-EgDSI2. PCR was conducted with the same primers to isolate EgDSI2 and resulted in a single DNA band with a size of around 1.8 kb which was the size of the target gene (Figure 3b).

The EgDSI2 sequence was obtained using T7 promoter and T7 terminator as sequencing primers, i.e. sequences that flank the inserted EgDSI2 in pET-EgDSI2. The gene was 1851 bp and contained the ATG start codon and the TAA stop codon at the ends of the sequence (GenBank accession no. MN709889). The result of BLASTN alignment showed that EgDSI2 had 99.19% identity with the Eg gene from Bacillus sp. WP8 (CP010075.1) and 97.73% identity with Eg genes from Bacillus safensis strain U17-1 (CP015611.1) and B. safensis strain U41 (CP015610.1). These results confirmed that the DNA fragment isolated from B. safensis DS12 was an Eg gene. The EgDSI2 had a low GC content (44%). According to Grosjean and Oshima (2007), mRNAs from thermophilic organisms have a low GC content. A too high GC content may reduce the informational coding potential of mRNAs. High GC content may also favor the formation of highly stable stem-loop of mRNA molecule which can be a problem during translation. This showed that there was no correlation between the GC content and the optimal growth temperature of the cell.

3.6. In silico structure analysis of the thermostable EgDSI2

Secondary structure analysis with the PredictProtein program (https://www.predictprotein.org/) (Yachdav et al. 2008) showed that EgDSI2 had high stability as a stem-loop structure with a stability of 99.19%. This stability is important for the thermostability of the enzyme.

FIGURE 1 Phylogenetic tree of DS12 isolate based on 16S rRNA gene sequence.

FIGURE 2 PCR product of endoglucanase gene EgDSI2. M = 1 Kb DNA marker, S = sample (EgDSI2), C = control without template.

FIGURE 3 pET-EgDSI2 construct confirmation by digestion with BamHI (a) and PCR (b). M = 1 Kb DNA marker, RP = pET-EgDSI2 recombinant plasmid, P = pET-32b plasmid (without gene insertion), C = control without template.
Suherman et al. *Indonesian Journal of Biotechnology* 24(2), 2019, 106-118

![Endoglucanase EgDSI2 sequence annotation. Red helix = α-helix, red arrow = β-strand, red square = residue binds to ligand, blue square = residue binds to metal, letter with red frame = catalytic residue (Asp55, Asp58, Glu423).](image)

2014 showed that thermostable EgDSI2 has 31.3% α-helix, 12.3% β-strand, and 56.3% loop. The EgDSI2 sequence was annotated with the SAS program (https://www.ebi.ac.uk/thornton-srv/databases/sas/) (Milburn et al. 1998) (Figure 4) based on the annotated sequence of endo/exocellulase E4 (glycosyl hydrolase family 9) from *Thermomonospora fusca* (PDB ID 1JS4). The amino acid sequence of IJS4 had 53.3% identity with EgDSI2, which was the highest value among other proteins in the PDB database.

The three dimensional (3D) structure of the thermostable EgDSI2 was predicted using the SWISS-MODEL program (https://swissmodel.expasy.org/) (Waterhouse et al. 2018) (Figure 5). The 3D structure modeling was based on the 3D structure of endo/exocellulase E4 (glycosyl hydrolase family 9) from *T. fusca* (PDB ID 1JS4) as template. The EgDSI2 sequence had 54.87% identity with this template, the highest compared to other proteins in the PDB database. The estimated value of the constructed model quality or the Global Model Quality Estimation (GMQE) was 0.8 (maximum 1) which indicated that the resulting model had a fairly high level of reliability. The active site residues in EgDSI2 were predicted with the ScanProsite program (https://prosite.expasy.org/scanprosite/) (De Castro et al. 2006) as Asp58, Asp414, and Glu423. The active sites could not be shown on the 3D model with the SWISS-MODEL program, but analysis with the program showed that the three active site residues were all located in the catalytic domain.

### 3.7. Amino acid composition of EgDSI2

The amino acid sequence of EgDSI2 was obtained from translation of its gene sequence with the Transseq EMBOSS program (https://www.ebi.ac.uk/Tools/st/emboss_transseq/) (Rice et al. 2000). EgDSI2 consisted of 616 amino acids. Alignment using BLASTP program showed that EgDSI2 had 99.35% identity with Eg from *B. safensis* strain SCAL1 (KMK71141.1). Results from BLASTP and InterProScan (http://www.ebi.ac.uk/interpro/search/sequence-search) (Jones et al. 2014) showed that EgDSI2 had a catalytic domain which belonged to the glycoside hydrolase family 9 (GH9) and a type 3 cellulose-binding domain (CBM3). Glycoside hydrolases (EC 3.2.1.-) are a widespread group of enzymes that hydrolyze glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate part (Henrissat et al. 1995). The catalytic domain of EgDSI2 was located in residues 6 - 434. The CBM3 domain was located in residues 465 - 546, at the C-terminus of the protein. CBM3 is composed of nine beta-
FIGURE 5 Endoglucanase EgDSI2 3D structure predicted by SWISS-MODEL. The protein is composed of two domains: the GH9 domain, shown in left, is rich in α-helices; and the CBM3 domain shown in right, is formed by β-strands arranged in a β-sandwich fashion.

strands that form a compact domain. The domain is arranged in two antiparallel beta-sheets stacking together to form a beta-sandwich. (Tormo et al. 1996; Shimon et al. 2000). The domain architecture of EgDSI2 is shown in Figure 6.

The molecular weight of EgDSI2 was predicted with the Compute pI/Mw tool (https://web.expasy.org/compute_pi/) (Gasteiger et al. 2005) to be 69.41 kDa. According to Kanchanadumkerng et al. (2017), Egs from Bacillus sp. had molecular weights in the range of 30 - 97 kDa. The EgDSI2 had similarity, in terms of modular structure and protein size of about 70 kDa, with Egs from Bacillus sp. AC-1 isolated from Mollusca (Li et al. 2006), endophytic B. pumilus strain CL16 (Lima et al. 2005), and EG-IV from Bacillus sp. KSM-522 isolated from soil (Hitomi et al. 1997). Enzymes with a modular organization of GH9-CBM3 always have the same properties, such as active in a wide pH range, high optimum temperature and thermostable (Kanchanadumkerng et al. 2017).

The amino acid composition of the EgDSI2 was determined by the ProtParam program (https://web.expasy.org/protparam/) (Gasteiger et al. 2005) and is displayed in a pie chart (Figure 7). The amino acid composition of the thermostable EgDSI2 was compared with thermostable Eg from Bacillus sp. strain KSM-635 (PDB ID 1G0C), thermolabile Eg from mesophilic bacteria B. agaradhaerens (PDB ID 7A3H) and thermolabile Eg from mesophilic bacteria B. licheniformis (PDB ID 2JEN). The comparison was performed to find the abundance of certain amino acids.
acids in the thermostable Egs. The comparison of amino acid compositions of the four Egs is shown in Table 4.

The thermostable EgDSI2 had a higher percentage of charged amino acids (Arg, Lys, Asp, Glu, His) than the thermolabile Eg 2JEN, but the percentage was lower than the thermolabile Eg 7A3H. The thermostable EgDSI2 had a lower percentage of charged amino acids than the thermostable Eg 1G0C, which also had a lower percentage of charged amino acids than Eg 7A3H. In general, thermostable proteins have more charged amino acids capable of forming ion pairs that function as weak bonds to maintain molecular conformation of the proteins (Kumar et al. 2007). However, the role of ion pairs in protein thermostability is controversial. Most experiments and comparisons between thermostable and thermolabile proteins had shown that ion pairs had a role in increasing thermostability. On the other hand, some studies had indicated that ion pairs/salt bridges usually destabilized the native conformation of proteins. The reason for the destabilization effect of an ion pair is that the formation of the ion pair in the protein core has a desolvation penalty, which is energetically unfavorable and was not fully compensated (Sadeghi et al. 2006). Formation of ion pairs and their networks did not completely account for all the charged amino acids found in a protein (Kajander et al. 2000). The high number of ion pairs in thermostable proteins was not due to the high number of charged residues, but was due to the high number of positively charged residues that bound to negatively charged residues (Sadeghi et al. 2006). The lower charged amino acids content in the thermostable EgDSI2 than the thermolabile Eg 7A3H could be an indirect consequence of the high percentage of hydrophobic amino acids. Glu percentage in the thermostable EgDSI2 was lower than the thermolabile Eg 7A3H but was higher than the thermolabile Eg 2JEN. Glu percentage in the thermostable Eg 1G0C was higher than both 7A3H & 2JEN. Nevertheless, Sadeghi et al. (2006) found that both thermostable proteins and thermolabile proteins from mesophilic organisms have a high content of Glu. This eliminated Glu from having a role in increasing protein thermostability.

The percentage of polar amino acids (Gln, Asn, Ser, Thr, Tyr, Cys) in the thermostable EgDSI2 was higher than the thermostable Eg 1G0C. The percentage of polar amino acids in the thermostable EgDSI2 was higher than the thermolabile Eg 7A3H but lower than the thermolabile Eg 2JEN. Due to this discrepancy, it was difficult to

| No. | Amino Acid (AA)          | AA Group          | Percentage in |
|-----|--------------------------|-------------------|---------------|
|     |                          |                   | EgDSI2 | 1G0C | 7A3H | 2JEN* |
| 1   | Arginine (Arg/R)         | Charged           | 3.4    | 3.2  | 2.6  | 1.7   |
| 2   | Lysine (Lys/K)           |                   | 5.0    | 3.6  | 3.3  | 6.0   |
| 3   | Aspartic acid (Asp/D)    |                   | 6.5    | 7.5  | 8.9  | 5.2   |
| 4   | Glutamic acid (Glu/E)    |                   | 6.2    | 9.2  | 8.3  | 2.2   |
| 5   | Histidine (His/H)        |                   | 2.4    | 1.8  | 2.6  | 1.7   |
|     | Total                    |                   | 23.5   | 25.3 | 25.7 | 16.8  |
| 6   | Glutamine (Gln/Q)        | Polar / Hydrophilic | 4.5    | 2.5  | 4.0  | 2.6   |
| 7   | Asparagine (Asn/N)       |                   | 5.7    | 6.8  | 6.9  | 9.9   |
| 8   | Serine (Ser/S)           |                   | 6.2    | 7.5  | 7.3  | 12.1  |
| 9   | Threonine (Thr/T)        |                   | 6.2    | 7.3  | 3.6  | 8.2   |
| 10  | Tyrosine (Tyr/Y)         |                   | 6.0    | 4.1  | 4.3  | 6.0   |
| 11  | Cysteine (Cys/C)         |                   | 0.3    | 0    | 0    | 0     |
|     | Total                    |                   | 28.9   | 28.2 | 26.1 | 38.8  |
| 12  | Alanine (Ala/A)          |                   | 10.6   | 8.4  | 7.9  | 6.5   |
| 13  | Glycine (Gly/G)          |                   | 7.1    | 7.0  | 8.3  | 9.1   |
| 14  | Isoleucine (Ile/I)       |                   | 3.6    | 5.9  | 6.9  | 6.0   |
| 15  | Leucine (Leu/L)          |                   | 7.1    | 6.5  | 5.3  | 3.0   |
| 16  | Methionine (Met/M)       | Nonpolar / Hydrophobic | 1.8    | 1.4  | 2.3  | 0.9   |
| 17  | Phenylalanine (Phe/F)    |                   | 4.5    | 3.9  | 3.3  | 3.4   |
| 18  | Valine (Val/V)           |                   | 5.5    | 6.2  | 6.9  | 6.5   |
| 19  | Proline (Pro/P)          |                   | 4.2    | 4.5  | 3.6  | 2.6   |
| 20  | Tryptophan (Trp/W)       |                   | 3.1    | 2.6  | 3.6  | 6.5   |
|     | Total                    |                   | 47.5   | 46.4 | 48.1 | 44.5  |

*) EgDSI2 = thermostable endoglucanase from Bacillus safensis DSI2; 1G0C = thermostable endoglucanase from Bacillus sp. KSM-635; 7A3H = thermolabile endoglucanase from B. agaradhaerens; 2JEN = thermolabile endoglucanase from B. licheniformis.
draw a conclusion from this comparison. However, in the thermostable EgDSI2 itself, polar amino acids were less frequent compared to hydrophobic amino acids. The low content of polar amino acids was in agreement with the study by Sadeghi et al. (2006). The reason behind this low content of polar amino acids could be to avoid deamination and backbone cleavage in proteins (Sadeghi et al. 2006). Polar amino acids could form hydrogen bonds. So it was also assumed that a high percentage of polar amino acids would probably increase the hydrogen bonding capability. A hydrogen bond is formed by three atoms, i.e. one H atom and two electronegative atoms (often N or O) (Jeong et al. 2003). Hydrogen bonds are important to stabilize protein structure (thus retain its activity) under extreme temperature (Ishak et al. 2019). The number and types of hydrogen bond in a protein are related to its thermostability (Kar and Scheiner 2004; Ragone 2001). The study by Dalhus et al. (2002) found that malate dehydrogenase (MDH) from thermophiliic bacteria had the highest number of hydrogen bonds compared to MDH from moderate thermophiliic bacteria and mesophilic bacteria. In addition, the study by Sadeghi et al. (2006) showed similar results that thermostable proteins had a higher number of hydrogen bonds than thermolabile proteins from mesophilic organisms.

The percentage of the hydrophobic amino acids (Ala, Gly, Ile, Leu, Met, Phe, Val, Pro, Trp) in the thermostable EgDSI2 and the thermostable Eg 1G0C were higher than the thermolabile Eg 2JEN but lower than the thermolabile Eg 7A3H. The study by Sadeghi et al. (2006) showed that thermostable proteins contained a high percentage of hydrophobic amino acids. Ala in the thermostable EgDSI2 and the thermostable Eg 1G0C were present in a higher percentage compared to the thermolabile Egs 7A3H & 2JEN. This was consistent with the results of the study by Kumwenda et al. (2013) which reported that proteins from extreme thermophilic bacterium *Thermus thermophilus* HB27 had a greater frequency of Ala than proteins from thermotolerant bacteria *T. scotoductus* SA-01. The percentage of Pro in the thermostable EgDSI2 and the thermostable Eg 1G0C were higher than both 7A3H and 2JEN. This was in accordance with the results of other studies that Pro was more frequent in thermostable proteins than in thermolabile proteins (Watanabe et al. 1991; Sadeghi et al. 2006). Pro had been used in mutational studies to improve protein stability (Velman et al. 1996; Van den Burg et al. 1998). Pro is the only amino acid whose side chain binds to the central α carbon at the carboxyl group (COOH) and to the nitrogen atoms in the amino group (NH2) forming a ring structure (Koenig et al. 2018). This ring structure restricted conformational freedom of Co-N rotation (Yu et al. 2015). Pro also restricted the conformation of the preceding amino acid in a protein sequence (Bajaj et al. 2007). Pro was known to rigidify flexible regions and enhanced thermostability (Yu et al. 2015). These results indicated that the thermostable EgDSI2 was a more rigid enzyme than the thermolabile Egs 7A3H & 2JEN.

However, the percentage of Trp and Gly in the thermostable EgDSI2 and the thermostable Eg 1G0C were lower than the two thermolabile Egs. (Sadeghi et al. 2006) also reported that thermostable proteins had a high content of Trp. This finding, along with the result by (Sadeghi et al. 2006) indicated that Trp did not account for thermostability. Gly has a simple side chain, a hydrogen atom, and has greater conformational freedom so that it can provide flexibility for adjacent residues (Yan and Sun 1997). Reduction in the Gly content in thermostable proteins reduced flexibility and thus enhanced rigidity, preventing the protein from unfolding at high temperature (Ladenstein and Antranikian 1998). The lower Gly percentage in thermostable EgDSI2 indicated that it had greater rigidity than the thermolabile Egs 7A3H & 2JEN.

4. Conclusions

This study described the cloning and molecular characterization of endoglucanase (Eg) from *B. safensis* DS12. Amino acid composition analysis showed that EgDSI2 had the characteristics of thermostable proteins, i.e. high percentage of hydrophobic residues, high Pro content, and low Gly content. The modular organization of GH9-CBM3 indicated that EgDSI2 had properties of being active in a wide pH range, a high optimum temperature, and thermostable. These properties are advantageous to be applied in industry.

Acknowledgments

This study was supported by the Indonesia Endowment Fund for Education (LPDP) Scholarship, Batch IV 2015.

Authors’ contributions

EAS and MRM designed the study. EAS carried out the laboratory work. EAS and MRM analyzed the data. EAS wrote the manuscript, MRM reviewed and edited the manuscript. ER gave the laboratory work. EAS and MRM designed the study. EAS carried out the laboratory work. EAS and MRM analyzed the data. EAS and MRM wrote the manuscript, MRM reviewed and edited the manuscript. ER gave the laboratory work.

Competing interests

The authors declare no competing interest.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2.

Bajaj K, Madhusudhan MS, Adkar BV, Chakraborti P, Ramakrishnan C, Sali A, Varadarajan R. 2007. Stereoenhanced criteria for prediction of the effects of proline mutations on protein stability. PLoS Comput Biol. 3(12):e241. doi:10.1371/journal.pcbi.0030241.

Sadeghi et al. (2006) showed that thermostable proteins contained a high percentage of hydrophobic amino acids. Ala in the thermostable EgDSI2 and the thermostable Eg 1G0C were present in a higher percentage compared to the thermolabile Egs 7A3H & 2JEN. This was consistent with the results of the study by Kumwenda et al. (2013) which reported that proteins from extreme thermophilic bacterium *Thermus thermophilus* HB27 had a greater frequency of Ala than proteins from thermostolerant bacteria *T. scotoductus* SA-01. The percentage of Pro in the thermostable EgDSI2 and the thermostable Eg 1G0C were higher than both 7A3H and 2JEN. This was in accordance with the results of other studies that Pro was more frequent in thermostable proteins than in thermolabile proteins (Watanabe et al. 1991; Sadeghi et al. 2006). Pro had been used in mutational studies to improve protein stability (Velman et al. 1996; Van den Burg et al. 1998). Pro is the only amino acid whose side chain binds to the central α carbon at the carboxyl group (COOH) and to the nitrogen atoms in the amino group (NH2) forming a ring structure (Koenig et al. 2018). This ring structure restricted conformational freedom of Co-N rotation (Yu et al. 2015). Pro also restricted the conformation of the preceding amino acid in a protein sequence (Bajaj et al. 2007). Pro was known to rigidify flexible regions and enhanced thermostability (Yu et al. 2015). These results indicated that the thermostable EgDSI2 was a more rigid enzyme than the thermolabile Egs 7A3H & 2JEN.

However, the percentage of Trp and Gly in the thermostable EgDSI2 and the thermostable Eg 1G0C were lower than the two thermolabile Egs. (Sadeghi et al. 2006) also reported that thermostable proteins had a high content of Trp. This finding, along with the result by (Sadeghi et al. 2006) indicated that Trp did not account for thermostability. Gly has a simple side chain, a hydrogen atom, and has greater conformational freedom so that it can provide flexibility for adjacent residues (Yan and Sun 1997). Reduction in the Gly content in thermostable proteins reduced flexibility and thus enhanced rigidity, preventing the protein from unfolding at high temperature (Ladenstein and Antranikian 1998). The lower Gly percentage in thermostable EgDSI2 indicated that it had greater rigidity than the thermolabile Egs 7A3H & 2JEN.

4. Conclusions

This study described the cloning and molecular characterization of endoglucanase (Eg) from *B. safensis* DS12. Amino acid composition analysis showed that EgDSI2 had the characteristics of thermostable proteins, i.e. high percentage of hydrophobic residues, high Pro content, and low Gly content. The modular organization of GH9-CBM3 indicated that EgDSI2 had properties of being active in a wide pH range, a high optimum temperature, and thermostable. These properties are advantageous to be applied in industry.

Acknowledgments

This study was supported by the Indonesia Endowment Fund for Education (LPDP) Scholarship, Batch IV 2015.

Authors’ contributions

EAS and MRM designed the study. EAS carried out the laboratory work. EAS and MRM analyzed the data. EAS wrote the manuscript, MRM reviewed and edited the manuscript. ER gave the laboratory work. EAS and MRM designed the study. EAS carried out the laboratory work. EAS and MRM analyzed the data. EAS and MRM wrote the manuscript, MRM reviewed and edited the manuscript. ER gave the laboratory work.

Competing interests

The authors declare no competing interest.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2.

Bajaj K, Madhusudhan MS, Adkar BV, Chakraborti P, Ramakrishnan C, Sali A, Varadarajan R. 2007. Stereoenhanced criteria for prediction of the effects of proline mutations on protein stability. PLoS Comput Biol. 3(12):e241. doi:10.1371/journal.pcbi.0030241.
Dalhus B, Saarinen M, Sauer UH, Eklund P, Johansson K, Karlsson A, Ramaswamy S, Bjork A, Synstad B, Naterstad K, et al. 2002. Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases. J Mol Biol. 318(3):707–721. doi:10.1016/S0022-2836(02)00050-5.

De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34(suppl_2):W362–W365. doi:10.1093/nar/gkl124.

De Marco EG, Heck K, Martos ET, Van Der Sand ST. 2017. Purification and characterization of a thermostable alkaline cellulase produced by Bacillus licheniformis. BioResources 380 isolated from compost. An Acad Bras Cienc. 89(3):2359–2370. doi:10.1590/0001-7509-Ambiv795.

Grosjean H, Oshima T. 2007. How Nucleic Acids Lies of glycosyl hydrolases. Proc Natl Acad Sci. 94(7):2413–2419. doi:10.1073/pnas.061039194.

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. In: JM Walker, editor, The Proteomics Protocols Handbook. Totowa: Humana Press. p. 571–607. doi:10.1385/1-5925-890-0:571.

Grosjean H, Oshima T. 2007. How Nucleic Acids Cope with High Temperature. In: C Gerday, N Glansdorff, editors, Physiol Biochem Extreme. Washington DC: ASM Press. p. 75–85.

Kajander T, Kahn PC, Passila SH, Cohen DC, Lehtio L, Adolfsen W, Schell U, Goldman A. 2000. Buried charged surface in proteins. J Mol Biol. 318(3):707–721. doi:10.1006/jmbi.2001.4979.

Koenig JJ, Neudörfl JM, Hansen A, Breugst M. 1999. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30(9):1236–1240. doi:10.1093/bioinformatics/btu031.

Kanchanadumkerpeng P, Rakkana K, Sakka K, Wiwat C. 2017. Characterization and secretive expression in Bacillus subtilis of endoglucanase from Bacillus safensis isolated from freshwater swamp forest. Walailak J Sci & Tech. 14(3):199–213.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biology Evol. 35(6):1547–1549. doi:10.1093/molbev/msy096.
Kumwenda B, Litthauer D, Bishop OT, Reva O. 2013. Analysis of protein thermostability enhancing factors in industrially important Thermus bacteria species. Evol Bioinf Online. 9:EBO–S12539. doi:10.4137/EBO.S12539.

Ladenstein R, Antranikian G. 1998. Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. In: G Antranikian, editor, Biotechnology of Extremophiles, volume 61. Berlin: Springer. p. 37–85. doi:10.1007/BFb0102289.

Lane CE, Hulgan D, O’Quinn K, Benton MG. 2015. CE-MASuite: open source degenerate PCR primer design. Bioinformatics. 31(22):3688–3690. doi:10.1093/bioinformatics/btv420.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23(21):2947–2948. doi:10.1093/bioinformatics/btm404.

Lateef A, Adelere IA, Gueguim-Kana EB. 2015. The biology and potential biotechnological applications of Bacillus safensis. Biologia. 70(4):411–419. doi:10.1515/biolog-2015-0062.

Li W, Zhang WW, Yang MM, Chen YL. 2008. Cloning of the thermostable cellulase gene from newly isolated Bacillus subtilis and its expression in Escherichia coli. Mol Biotechnol. 40(2):195–201. doi:10.1007/s12033-008-9079-y.

Li YH, Ding M, Wang J, Xu GJ, Zhao F. 2006. A novel thermoacidophilic endoglucanase, Ba-EGA, from a new cellulose-degrading bacterium, Bacillus sp. AC-1. Appl Microbiol Biotechnol. 70(4):430–436. doi:10.1007/s00253-005-0075-x.

Lima AOS, Quecine MC, Fungaro MHP, Andreote FD, Maccheroni W, Araújo WL, Silva-Filho MC, Pizzirani-Kleiner AA, Azevedo J. 2005. Molecular characterization of a β-l, 4-endoglucanase from an endophytic Bacillus pumilus strain. Appl Microbiol Biotechnol. 68(1):57–65. doi:10.1007/s00253-004-1740-1.

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev. 66(3):506–577. doi:10.1128/MMBR.66.3.506-577.2002.

Mc Gowan J. 2015. Bioinformatics Tools. URL https://jamiemcg.github.io/bioinf/gc.html.

Milburn D, Laskowski RA, Thornton JM. 1998. Sequences annotated by structure: a tool to facilitate the use of structural information in sequence analysis. Protein Eng. 11(10):855–859. doi:10.1093/protein/11.10.855.

Moes IR, Natalia D, Ningrum RW, Dwijayanti A. 2014. Cloning and expression of endoglucanase gene from thermophilic bacteria Bacillus sp. RP1. Microbiol Indonesia. 8(4):4. doi:10.5454/mi.8.4.4.

Nakashima H, Fukuchi S, Nishikawa K. 2003. Compositional changes in RNA, DNA and proteins for bacterial adaptation to higher and lower temperatures. J Biochem. 133(4):507–513. doi:10.1093/jb/mvg067.

Patel K, Patel H, Shah G. 2014. Production and optimization of cellulase enzymes from newly isolated fungi. J Pure Appl Microbiol. 8(5):4163–4169.

Ragone R. 2001. Hydrogen-bonding classes in proteins and their contribution to the unfolding reaction. Protein Sci. 10(10):2075–2082. doi:10.1101/plos.09201.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16(6):276–277. doi:10.1016/S0168-9525(00)02024-2.

Sadeghi M, Naderi-Manesh H, Zarrabi M, Ranjarb B. 2006. Effective factors in thermostability of thermophilic proteins. Biophys Chem. 119(3):256–270. doi:10.1016/j.biopha.2005.09.018.

Satomi M, La Duc M, Venkateswaran K. 2006. Bacillus safensis sp. nov., isolated from spacecraft and assembly-facility surfaces. Int J Syst Evol Microbiol. 56(8):1735–1740. doi:10.1099/ijs.0.64189-0.

Sheng P, Huang S, Wang Q, Wang A, Zhang H. 2012. Isolation, screening, and optimization of the fermentation conditions of highly cellulolytic bacteria from the hindgut of Holotrichia parallela larvae (Coleoptera: Scarabaeidae). Appl Biochem Biotechnol. 167(2):270–284. doi:10.1007/s12100-012-9670-3.

Shimon L, Belaich A, Belaich JP, Bayer E, Lamed R, Shoham Y, Frolow F, et al. 2007. Clustal W and Clustal X version 2.0, an improved version of the Clustal alignment algorithms. Bioinformatics. 23(21):2947–2948. doi:10.1093/bioinformatics/btm404.

Sievers F, Wilm A, Dineen D, Gibson T, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson J, Higgins D. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:539. doi:10.1038/msb.2011.75.

Slepecky R, Hemphill H. 2006. The Genus Bacillus—Nonmedical. New York: Springer US. p. 530–562. doi:10.1007/0-387-30744-3_16.

Srinivas R, Panda T. 1998. pH and thermal stability studies of carboxymethyl cellulase from intergeneric fusants of Trichoderma reesei/Saccharomyces cerevisiae. J Ind Microbiol Biotechnol. 21(4-5):178–183. doi:10.1007/s10295-000-0566-1.

Sukumaran R, Singhania RR, Pandey A. 2005. Microbial cellulases-production, applications and challenges. J Sci Ind Res 64:832–844.

Teather R, Wood P. 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol. 43(4):777–780.

Tormo J, Lamed R, Chirino A, Morag E, Bayer E, Shoham Y, Steitz T. 1996. Crystal structure of a bacterial family-III cellulose-binding domain:
a general mechanism for attachment to cellulose. EMBO J. 15(21):5739–5751. doi:10.1002/j.1460-2075.1996.tb00960.x.

Turner P, Mamo G, Karlsson E. 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining. Microb Cell Fact. 6(1):9. doi:10.1186/1475-2859-6-9.

Van den Burg B, Vriend G, Veitman OR, Venema G, Eijsink VGH. 1998. Engineering an enzyme to resist boiling. Proc Natl Acad Sci. 95(5):2056–2060. doi:10.1073/pnas.95.5.2056.

Veltman OR, Vriend G, Middelhooven PJ, van den Burg B, Venema G, Eijsink VGH. 1996. Analysis of structural determinants of the stability of thermolysinlike proteases by molecular modelling and site-directed mutagenesis. Protein Eng, Des Sel. 9(12):1181–1189. doi:10.1093/protein/9.12.1181.

Watanabe K, Chishiro K, Kitamura K, Suzuki Y. 1991. Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1, 6-glucosidase from Bacillus thermoglucosidasius KP1006. J Biol Chem. 266(36):24287–24294.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucl Acids Res. 46(W1):W296–W303. doi:10.1093/nar/gky427.

Wood TM, Bhat KM. 1988. Biomass, Part A: Cellulose and hemicellulose. In: WA Wood, ST Kellog, editors, Methods in Enzymology, volume 160. New York: Academic Press. p. 87–91.

Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Höngschmid P, Schafferhans A, Roos M, Bernhofer M, et al. 2014. PredictProtein—an open resource for online prediction of protein structural and functional features. Nucl Acids Res. 42(W1):W337–W343. doi:10.1093/nar/gku366.

Zhang YPH, Himmel ME, Mielenz JR. 2006. Outlook for cellulase improvement: screening and selection strategies. Biotechnol Adv. 24(5):452–481. doi:10.1016/j.biotechadv.2006.03.003.