

## Cloning of the Endoglucanase Gene from a *Bacillus amyloliquefaciens* PSM 3.1 in *Escherichia coli* Revealed Catalytic Triad Residues Thr-His-Glu

<sup>1</sup>Zeily Nurachman, <sup>1</sup>Sari Dewi Kurniasih, <sup>1</sup>Ferra Puspitawati, <sup>1</sup>Sarwono Hadi,  
<sup>2</sup>Ocky Karna Radjasa and <sup>1</sup>Dessy Natalia

<sup>1</sup>Biochemistry Division, Faculty of Mathematics and Natural Sciences,  
Institute Technology Bandung, Jl. Ganesha No 10, Bandung 40132, Indonesia

<sup>2</sup>Department of Marine Science, Faculty of Fishery and Marine Sciences,  
Diponegoro University, Jl. Prof. Soedharto, S.H. Tembalang, Semarang 50275, Indonesia

---

**Abstract: Problem statement:** An Indonesian marine bacterial isolate, *Bacillus amyloliquefaciens* PSM 3.1 was isolated for hydrolyzing cellulose. A 1500-bp nucleotide fragment was amplified from the chromosomal DNA by the use of primers directed against the conserved sequence of *Bacilli* endoglucanase genes obtained from GenBank. **Approach:** The fragment was cloned and expressed in *Escherichia coli*. **Results:** The endoglucanase gene (*eglIII* gene) had an open reading frame of 1500 nucleotides encoding a protein of 499 amino acids. The EglIII protein belonged to Glycosyl Hydrolase family 5 (GH5) with a Cellulose Binding Module 3 (CBM 3). The structure model of the EglIII protein revealed that the catalytic residues seemed to be Glu169 (as proton donor) and Glu257 (as nucleophile) and the catalytic triad residues were Thr256, His229 and Glu169. The EglIII endoglucanase exhibited an optimum pH of 6.0 and temperature of 50°C and the enzyme tolerated to high salt concentration. **Conclusion/Recommendations:** This EglIII endoglucanase is a promising candidate for many applications in biomass degradation.

**Key words:** *Bacillus amyloliquefaciens*, *Escherichia coli*, catalytic triad residues, endoglucanase, cellobiohydrolases, cellulolytic, *Galaxea* sp, nucleotides, hydrolysis, agarose gel electrophoresis, synthesized, polysaccharide, DNA polymerase, carboxymethyl cellulose

---

### INTRODUCTION

Cellulose is completely hydrolyzed by means of three synergistic enzymes: endoglucanases (endo-1,4- $\beta$ -D-glucanases, EC 3.2.1.4, hydrolyze internal  $\beta$ -1,4-glycosidic bond randomly in the cellulose chain), cellobiohydrolases (exo-1,4- $\beta$ -D-glucanases, EC 3.2.1.91, cleave off cellobiose units from the ends of cellulose chains) and  $\beta$ -glucosidases (1,4- $\beta$ -D-glucosidases, EC 3.2.1.21, hydrolyze cellobiose and cellooligosaccharides to glucose units) (Henrissat, 1994; Wood and Garica-Campayo, 1990; Zhang and Lynd, 2004). Of these complex enzymes, the role of endoglucanase is very important for initiating cellulose hydrolyses. The looking for new endoglucanases and its correlated genes are still continuing since most of these enzymes have not met the requirements in industrial yet. Hence, worldwide efforts for searching microorganisms producing endoglucanases with higher activities and greater efficiency are being undertaken.

Endoglucanases from a wide range of terrestrial organisms have been reported, however, few

endoglucanase reports from marine sources are known. In order to explore Indonesian marine microbial collections, gene encoding enzyme for hydrolyzing cellulose is investigated. Of 32 isolates we have, *Bacillus amyloliquefaciens* PSM 3.1, which isolated from the surface of the hard coral *Galaxea* sp.-collected from Pantai Merah, Komodo Island, Flores, Indonesia by scuba diving at depth of approximately 3 m, shows the highest cellulolytic activity. Here, we report the cloning of endoglucanase gene of *B. amyloliquefaciens* PSM 3.1 in *E. coli*. Protein structure model of the endoglucanase from amino acid residues deduced its gene shows novel catalytic triad residues.

### MATERIALS AND METHODS

**Bacterial strains, vector and culture conditions:** The bacterial strains and plasmids that used in this study are listed in Table 1. The *B. amyloliquefaciens* PSM 3.1 was cultivated in marine broth medium (0.25% yeast extract and 0.5% bacto peptone in seawater-water 3:1).

---

**Corresponding Author:** Zeily Nurachman, Biochemistry Division, Faculty of Mathematics and Natural Sciences, Institute Technology Bandung, Jl. Ganesha No 10, Bandung 40132, Indonesia  
Tel: +62 22 250 2103 Fax: +62 22 250 4154

Table 1: Strains and plasmids used in this study

Strain and plasmids	Description	Source
<b>Strains</b>		
<i>B. amylolyquefaciens</i> PSM 3.1	Wild-type, can degrade cellulose	Isolated by our research group
<i>E. coli</i> TOP 10	General cloning host strain	Purchased from Invitrogen
<i>E. coli</i> BL21 (DE3)	Expression host strain	Purchased from Promega
<b>Plasmids</b>		
pGEM-T	General cloning vector, Amp <sup>r</sup>	Purchased from Promega
pET-20b	Expression vector, Amp <sup>r</sup>	Purchased from Novagen
pG-Cel	<i>eglIII</i> gene clone into pGEM-T	Constructed in this study
pET-Egl	<i>eglIII</i> gene clone into pET-20b	Constructed in this study

The *E. coli* strains were cultivated in LB broth medium. For plating, 1% bacto agar was added to LB broth medium. For screening of recombinants, 100 µg mL<sup>-1</sup> of ampicillin, 100 mM Isopropyl-β-Thiogalactopyranoside (IPTG) and 50 mg mL<sup>-1</sup> 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) were added to agar plate.

**DNA manipulations:** The molecular cloning procedure was performed mainly using the procedure of Sambrook *et al.* (1989). Genomic DNA was isolated from *B. amylolyquefaciens* PSM 3.1 using a Wizard Genomic Purification Kit (Promega). Digestion of DNA with restriction endonuclease, separation of fragment by agarose gel electrophoresis, ligation of DNA fragments, transformation of *E. coli* with plasmid DNA and extraction of recombinant DNA were all performed as described by Sambrook *et al.* (1989). DNA fragments were recovered from agarose gel using Gel Extraction Kit (QiaGen).

**Primers and PCR amplification conditions of the endoglucanase gene:** According to the published sequences of endoglucanase genes in the GenBank database (accession nos. AY044252, Z29076, DQ116829, DQ82954 and EU022560), two DNA oligonucleotide primers were designed and synthesized to allow the PCR amplification of the entire gene from genomic DNA of *B. amylolyquefaciens* PSM 3.1. The primers were as follows:

- CelF: 5'-ATGAAACGGTCAATCTC-3' (deduced from conserved sequence in the genus of *Bacillus*)
- CelR: 5'-CTAATTTGGTTCTGTTCCC-3' (deduced from conserved sequence in the genus of *Bacillus*)

The PCR reaction mixtures for 25 µL volume reaction system contained 1× Dream *Taq* reaction buffer (Fermentas), 0.2 mM forward primer and reverse primer, 0.2 mM dNTP, 2mM MgCl<sub>2</sub>, 2.5 U Dream *Taq* DNA polymerase (Fermentas) and 2 ng of DNA

template. The PCR started with a 2 min denaturation at 94°C, followed by 30 cycles of 1 min at 94°C, 1.5 min at 51°C and 2 min at 72°C. A final extension of 5 min at 72°C was performed. The amplified PCR products were analyzed by agarose gel electrophoresis.

**Cloning and expression of endoglucanase gene:** The PCR products with size of 1500 bp which is predicted as an endoglucanase gene (Fig. 1A) was cloned into a pGEM-T vector, yielding pG-Cel and then transformed into *E. coli* TOP10 for sequencing and assaying. According to the result of sequences (Fig. 1B), two new primers were designed and synthesized to allow the PCR amplification of the endoglucanase gene in pG-Cel without signal peptide sequence (1-87) and addition of restriction sites for *EcoRI* and *XhoI*. The primers were as follows:

- Egl-F: 5'-AGAATTCGGCAGGGACAAGAACGCCAGTAGC-3' (underline, *EcoRI* restriction site)
- Egl-R: 5'-ACTCGAGATTTGGTTCTGTTCCCCAAATCATGTTTCC-3' (underline, *XhoI* restriction site)

Each of PCR products were cloned into a pET-20b vector, yielding the expression vector pET-Egl. The recombinant plasmids were then transformed into *E. coli* BL21 (DE3) for further analysis and expression experiments.

**late assays for endoglucanase activity:** The transformed *E. coli* BL21 (DE3) cells harbouring pET-Egl were screened for the expression of endoglucanase activity by the Congo-red assay (Wood *et al.*, 1988). The transformed cells were overlaid on LB agar medium containing 1% Carboxymethyl Cellulose (CMC) and then incubated at 37°C overnight. Following incubation, the plates were stained with 1% Congo-red for 5 min and destained with 1M NaCl for 5 min. Positive clones were identified by a clear zone around endoglucanase-expressing colonies.

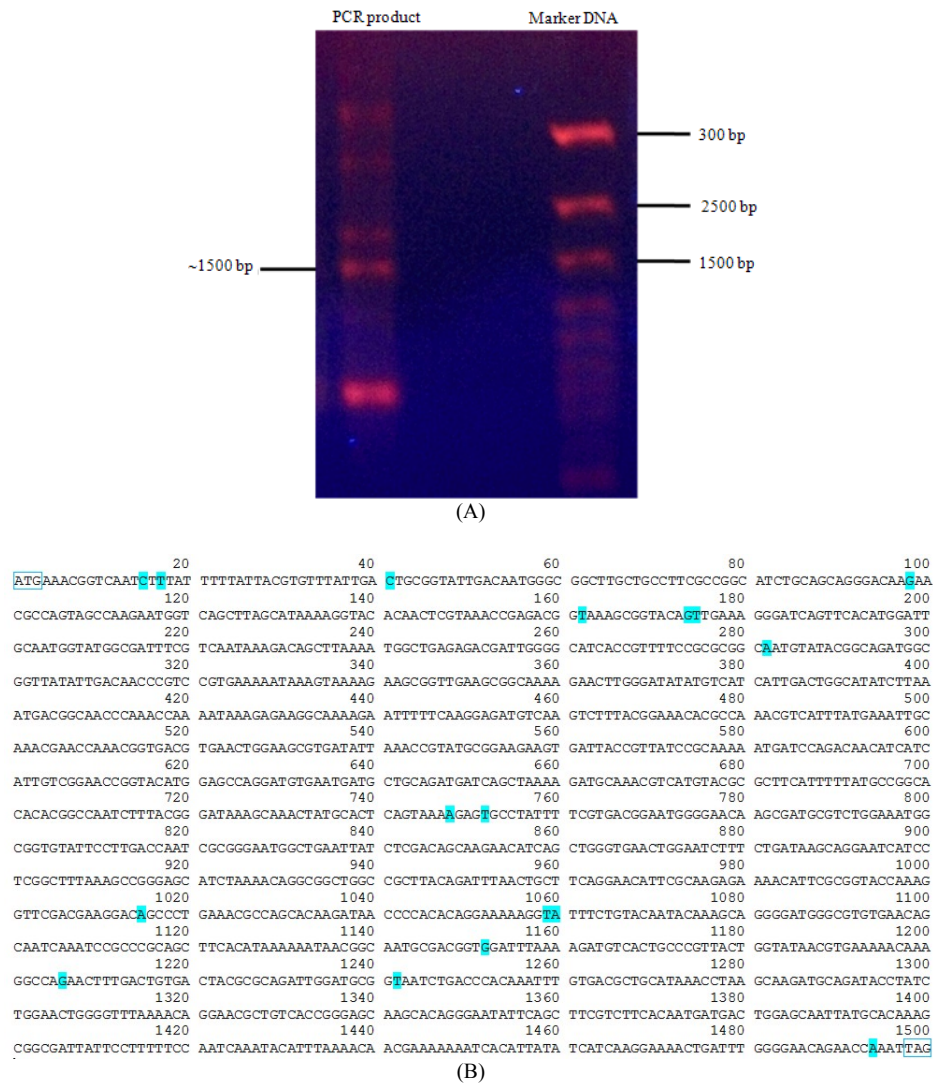


Fig. 1: A 1500 bp nucleotide fragment and its sequence. (A). Agarose gel electrophoresis of PCR product of *B. amyloliquefaciens* PSM 3.1 chromosomal DNA shows 1500 bp nucleotide fragment. (B). Nucleotide sequence of 1500 bp band of PCR product starts with ATG (in box) and ends with TAG (in box). The differences of bases that are compared with its counterpart (GenBank no AF363635.1 and M16185.1B) are highlighted

**DNA sequencing and protein structure modeling:**

The resulting DNAs were sequenced using dye-end terminator (Macrogen, Korea). A similarity analysis of the nucleotide sequence arising from PCR amplification was carried out using the BLAST(N) program (<http://www.ncbi.nlm.nih.gov>). A similarity analysis of the deduced amino acid sequence of the cloned gene was performed using the BLAST(P) program (<http://www.ncbi.nlm.nih.gov>). Protein structure modeling was carried out by using the SWISS-MODEL (Arnold *et al.*, 2006).

**Nucleotide sequence number:**

The DNA sequence of the endoglucanase gene from *B. amyloliquefaciens* PSM 3.1 has been deposited at GenBank under accession no. GU390463.

**Enzyme assays:**

The positive clone producing endoglucanase was cultured in LB broth medium and induced by IPTG. The cells were harvested by centrifugation (7000×g for 20 min at 4°C), suspended in 50 mM universal buffer (succinic acid, Na<sub>2</sub>HPO<sub>4</sub>, glycine), pH 6.0 in a ratio of 1 g cell: 2 mL buffer and

then sonicated (50% pulse for 15 second at 4°C, repeat 10 times). The endoglucanase was obtained from centrifugation (11000×g for 20 min at 4°C) of the solution and its activity was measured by measuring the release of reducing sugar, as given by the Dinitrosalicylic acid (DNS) method using cellobiosa as a standard (Miller, 1959). One unit of endoglucanase activity was defined as the amount of enzyme producing 1 μmol of reducing equivalents per minute under the assay condition. All assays were conducted in triplicates.

**Effect of pH, temperature and NaCl on endoglucanase activity:** The optimal pH of endoglucanase activity was determined by evaluating the hydrolysis reaction in the range of pH 4.0-10.0 using 25 mM universal buffers at 50°C. The optimum temperature of endoglucanase activity determined by evaluating the hydrolysis reaction in the temperature range of 26-80°C at pH 6.0. The influence of NaCl on the endoglucanase activity was determined by incubation of reaction mixture with different NaCl concentration at 50°C and pH 6.0.

**Protein determination:** Total protein concentration was determined by the Bradford method using bovine serum albumin as a standard (Bradford, 1976).

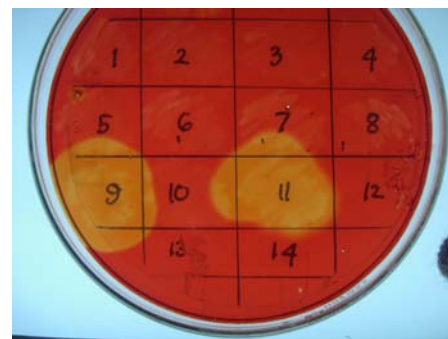
## RESULTS AND DISCUSSION

**Analysis of the endoglucanase gene:** Endoglucanases are the most diverse class of enzymes amongst cellulolytic enzymes (<http://afmb.cnrs-mrs.fr/CAZY/index.html>). Most of endoglucanase genes are obtained by PCR or by constructing DNA libraries. Here, the PCR products obtained using *B. amyloliquefaciens* PSM 3.1 chromosomal DNA as template were about 1500 bp and other non-specific bands (Fig. 1A). A 1500 bp band of the PCR products was recovered from the agarose gel and ligated to pGEM-T vector and then recombinant plasmids (pG-Cel) were transformed into *E. coli* TOP10.

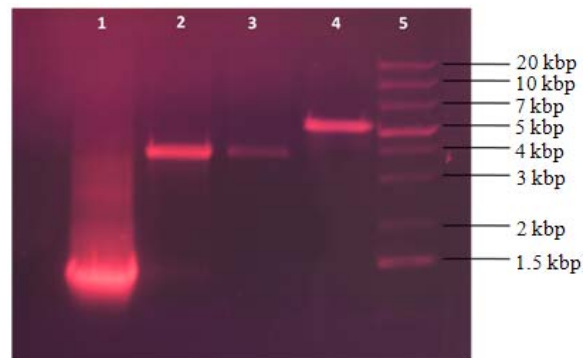
To analyze whether pG-Cel containing the targeted endoglucanase gene, the nucleotide sequence was determined (Fig 1B). The nucleotide sequence of endoglucanase gene *B. amyloliquefaciens* PSM 3.1 was generally 97% identical to that of endoglucanase (*engA*) gene of *B. amyloliquefaciens* strain UMAS1002 (GenBank no AF363635.1) and 97% identical to that of endoglucanase (*bglC*) gene of *B. subtilis* DLG (GenBank no M16185.1B) and there are 17 base differences. Thus, with respect to the sequence identity, the 1500 bp PCR product of the genome DNA is

considered to be an endoglucanase gene (designated as *eglIII* gene).

To examine expression of endoglucanase, the *eglIII* gene of pG-Cel was PCR amplified using primers Egl-F and Egl-R, cloned to pET-20b vector and then recombinant plasmids (pET-Egl) were transformed into *E. coli* BL21 (DE3). The transformed cells were screened and positive clones were identified by the presence of a clear zone surrounding colonies suggesting CMC hydrolysis (Fig. 2A). A positive clone contained pET-Egl harboring about 1423 bp of *eglIII* gene (Fig. 2B). Further resequencing result of the *eglIII* gene in pET-Egl showed no significant nucleotide sequence difference from that of *eglIII* gene in pG-Cel, except signal peptide sequence. So, the *eglIII* gene contained an Open Reading Frame (ORF) of 1500 nucleotides that started with ATG start codon and terminated with a non coding TAG as stop codon (Fig. 1B).



(A)



(B)

Fig. 2: Screening of transformant cells and plasmid analysis. (A) A clear zone around a clone indicates endoglucanase activity of a positive clone. (B) pET-Egl harboring *eglIII* gene is digested by nucleases: (1) 1423 bp PCR product, (2) pET-20b/*EcoRI*, (3) pET-20b/*XhoI*, (4) pET-Egl/*EcoRI*, and (5) Marker DNA



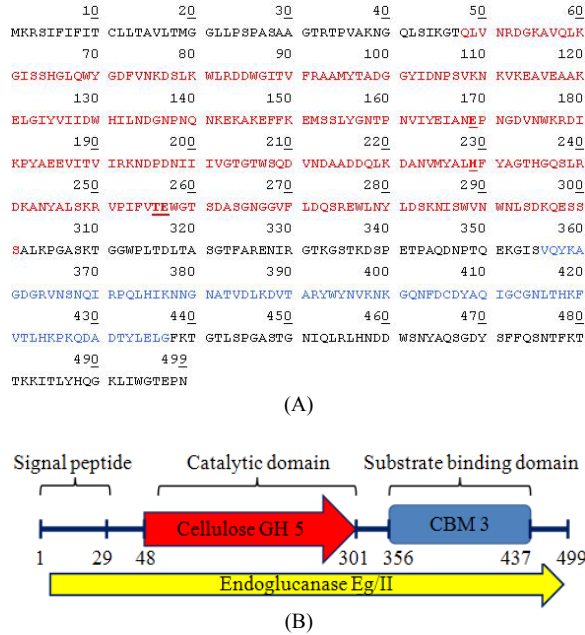


Fig. 3: Deduced amino acid sequence of EglIII endoglucanase of *B. amyloliquefaciens* PSM 3.1 (A), and its topology (B). The enzyme has two domains: catalytic domain (Glycosyl Hydrolase 5 (GH5); residues 48-301, red) and substrate binding domain (CBM 3; residues 356-437, blue). Four residues for conserved catalytic residues were underlined

**Structural analysis of the endoglucanase model:** Endoglucanases are modular enzymes consisting of two or more functional module, such as catalytic modules and carbohydrate binding modules (Ohmiya *et al.*, 1977). The amino acid sequence deduced from *eglIII* gene is shown in Fig. 3A. The ORF of *eglIII* gene consisted of 1500 nucleotides encoding a protein 499 amino acids, called EglIII endoglucanase. BLAST(P) result of EglIII endoglucanase revealed a modular enzyme composed of a signal peptide (1-29), a catalytic domain (48-301) of Glycosyl Hydrolase family 5 (GH5) and a substrate binding domain (356-437) of Cellulose Binding Module 3 (CBM 3) (Fig 3B). The modular organization (GH5-CBM 3) of endoglucanase are also found in many *Bacillus* genus such as in *B. licheniformis* strain B-41362 (Bischoff *et al.*, 2007), *B. amyloliquefaciens* DL-3 (Lee *et al.*, 2008) and *B. subtilis* (Li *et al.*, 2009).

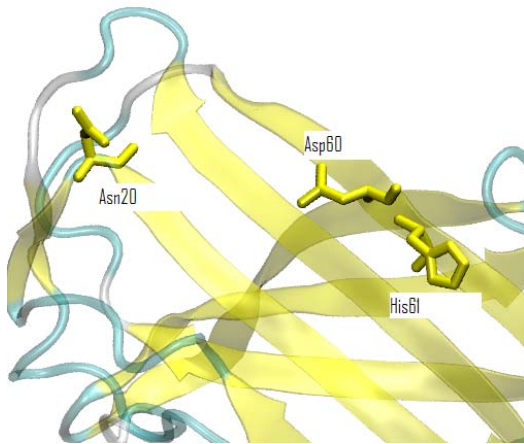
To identify a signal peptide of the EglIII endoglucanase, the Computer Program Signal IP 3.0 (<http://www.cbs.dtu.dk/services/SignalP>) was employed. The result showed that a signal peptide of

the EglIII endoglucanase was located at the first twenty nine residues (Fig. 3B). This sequence is conserved among *Bacillus* genus such as endoglucanase *B. amyloliquefaciens* strain UMASI002 (GenBank no AF363635.1) located at residues 1-29 and endoglucanase (BglC) *B. subtilis* DLG (GenBank no M16185.1B) located at residues 1-38.

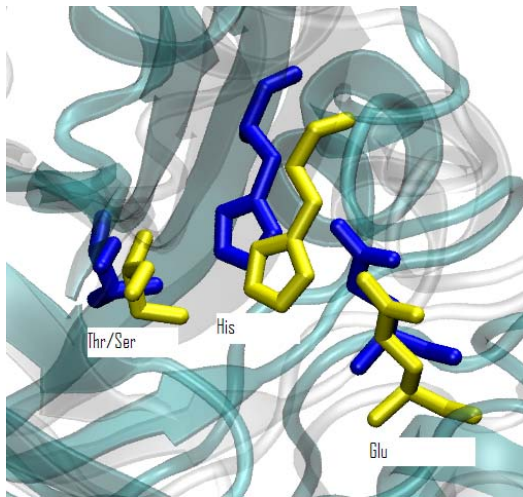
About 150 amino acid residues forming  $\beta$ -sheet fold structure construct CBM 3 of EglIII endoglucanase (Fig. 4A). Conserved residues Asn13-Asp52-Tyr53 for cellulose binding of EglIII endoglucanase were identical to residues Asn20-Asp60-His61 of *Clostridium cellulolyticum* (PDB no. 1G43). In addition, conserved amino acid residues forming the crevice on the surface of CBM 3 of *Clostridium cellulolyticum* were Arg44-Tyr46-Tyr93-Glu96. These residues were identical to Arg37-Trp39-Tyr78-Glu80 of CBM 3 of EglIII endoglucanase. The CBM 3 acts to bind a cellulose surface and acts as an anchor so that it will lead to strong adsorption of the enzyme on the cellulose surface. CBM 3 helps cellulose hydrolysis by selecting a single cellulose chain directly into the active site of the enzyme (Shimon *et al.*, 2000).

To detect conserved catalytic domain and substrate binding domain, structural model of EglIII endoglucanase was built by the SWISS-MODEL (Miller, 1959). The catalytic domain model of EglIII endoglucanase was a core domain of  $(\beta/\alpha)_8$  barrel structure. Its sequence had 69% identity with the reported Cel5A *B. agarethans* (PDB/Protein Data Bank no. 7A3H) and 69% identity with Cel5 *Bacillus* sp. (PDB no. 1LF1). Multiple alignment of those catalytic domain sequences indicated conserved catalytic residues of EglIII endoglucanase were Glu169 (predicted as a proton donor) and Glu257 (predicted as a nucleophile). Furthermore, close to Glu169 was located residues His229 and Thr256 (Fig. 4B). These residues (Thr-His-Glu) were a triad that to be different from a triad in their counterpart from *Bacilli* genus, that commonly Ser-His-Glu (Davies *et al.*, 1998; Shaw *et al.*, 2002). The position of residue Thr256 together His229 in the active site of EglIII endoglucanase may control the protonation of the acid/base Glu160. Protonated Glu may facilitate the first step of the catalytic reaction through protonation of the substrate (Shaw *et al.*, 2002).

**Activity of the endoglucanase:** Recombinant EglIII endoglucanase was obtained from *E. coli* BL21 (DE3) harboring pET-Egl. The profile of activity of EglIII endoglucanase versus pH seemed to be a saddle-like curve (Fig. 5A). The optimal pH of EglIII endoglucanase was determined to be 6.0. The enzyme retained more than 50% its activity in a wide range of pH (4-10).



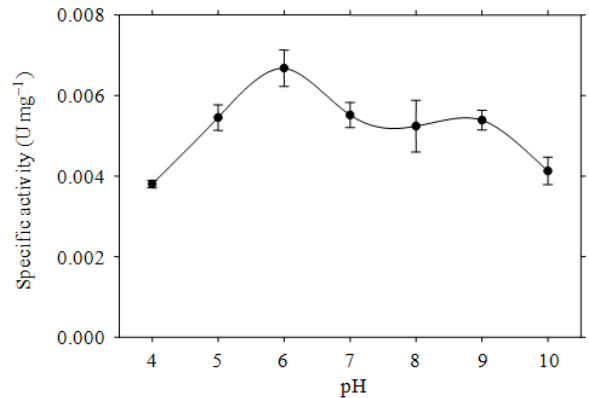
(A)



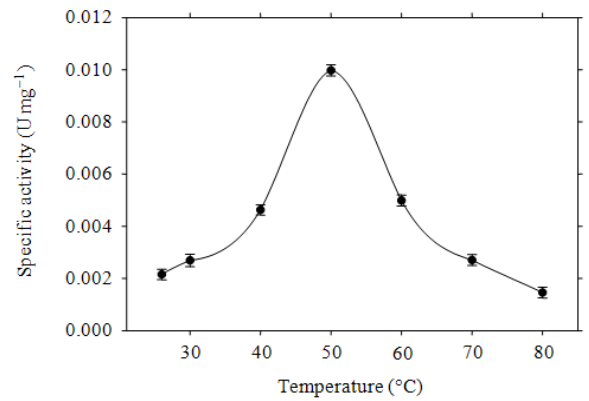
(B)

Fig. 4: Model structure of catalytic domain and cellulose binding module of EglIII endoglucanase. A. The cellulose binding module of EglIII endoglucanase had a  $\beta$ -sheet fold structure. B. The catalytic triad residues of EglIII endoglucanase *B. amyloliquefaciens* PSM 3.1 (blue) were superimposed to that of Cel5A *B. agardhans* (yellow)

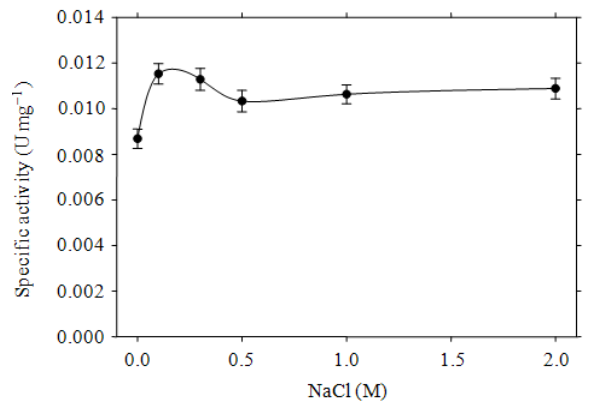
The apparent optimal temperature of EglIII endoglucanase at pH 6.0 was 50°C, with 50% of maximum activity being retained between 40 and 60°C (Fig. 5B). EglIII endoglucanase was found to be salt tolerant (Fig. 5C). The enzyme retained more than 90% of its activity in the presence of up to 2M NaCl. The overall characteristics of recombinant endoglucanase *EglIII* activity were similar to those of wild type EglIII endoglucanase.



(A)



(B)



(C)

Fig. 5: Properties of EglIII endoglucanase of *B. amyloliquefaciens* PSM 3.1. Effects of pH, temperature, and NaCl on EglIII endoglucanase activity are showed in A, B, and C, respectively

## CONCLUSION

The EglIII endoglucanase features described here show cellulolytic activity across a broad pH range,

moderate thermostable and high salt tolerant. On the basis of this characteristics, EglII endoglucanase is a promising candidate for application in biomass degradation, cellulose processing industry and biotechnological processes. In the future, it will be essential to perform further study of structure-function relationship of the enzyme.

### ACKNOWLEDGEMENT

This study was financially supported by a grant from the Riset Unggulan, Institute Technology Bandung, Indonesia. We thank Mr. Joshua Christian Tumboimbela for molecular artworks.

### REFERENCES

- Arnold, K., L. Bordoli, J. Kopp and T. Schwede, 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics*, 22: 195-201. DOI: 10.1093/bioinformatics/bti770
- Bischoff, K., S. Liu and S.R. Hughes, 2007. Cloning and characterization of a recombinant family 5 endoglucanase from *Bacillus licheniformis* strain B-41361. *Process Biochem.*, 42: 1150-1154. DOI: 10.1016/j.procbio.2007.05.001
- Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 72: 248-54. DOI: 10.1016/0003-2697(76)90527-3
- Davies, G.J., L. Mackenzie, A. Varrot, A.M. Brzozowski and M. Schülein *et al.*, 1998. Snapshots along an enzymatic reaction coordinate: Analysis of retaining  $\beta$ -glycoside hydrolase. *Biochem.*, 37: 11707-11713. DOI: 10.1021/bi981315i
- Henrissat, B., 1994. Cellulases and their interaction with cellulose. *Cellulose*, 1:1 69-196. DOI: 10.1007/BF00813506
- Lee, Y.J., B.K. Kim, B.H. Lee, K.I. Jo and N.K. Lee *et al.*, 2008. Purification and characterization of cellulase produced by *Bacillus amyloliquefaciens* DL-3 utilizing rice hull. *Bioresour. Technol.*, 99: 378-386. DOI: 10.1016/j.biortech.2006.12.013
- Li, W., X. Huan, Y. Zhou, Q. Ma and Y. Chen, 2009. Simultaneous cloning and expression of two cellulase gene from *Bacillus subtilis* newly isolated from Golden Takin (*Budorcas taxicolor Bedfordi*). *Biochem. Biophys. Res. Commun.*, 383: 397-400. DOI: 10.1016/j.bbrc.2009.04.027
- Miller, G.L., 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.*, 31: 426-428. DOI: 10.1021/ac60147a030
- Ohmiya, K., K. Sakka, S. Shikata and T. Kimura, 1977. Structure of cellulase and their applications. *Biotechnol. Gen. Eng. Rev.*, 14: 365-414. ISSN: 0264-8725
- Sambrook, J., E.F. Fritsch and T. Maniatis, 1989. *Molecular Cloning: A Laboratory Manual*. 2nd Edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Shaw, A., R. Bott, C. Vonrhein, G. Bricogne and S. Power *et al.*, 2002. A novel combination of two classic catalytic shemes. *J. Mol. Biol.*, 320: 303-309. DOI: 10.1016/S0022-2836(02)00387-X
- Shimon, L.J., W. Shimon, S. Pagès, A. Belaich and J.P. Belaich *et al.*, 2000. Structure of a family IIIa scaffolding CBD from the Cellulosome of *Clostridium cellulolyticum* at 2.2 Å resolution. *Acta Cryst.*, 56: 1560-1568. DOI: 10.1107/S0907444900012889
- Wood, P.J., J.D. Erfle and R.M. Teather, 1988. Use of Congo red polysaccharide interaction complex formation between Congo red and polysaccharide in detection and assay of polysaccharide hydrolases. *Methods Enzymol.*, 160: 59-74. DOI: 10.1016/0076-6879(88)60107-8
- Wood, T.M. and V. Garica-Campayo, 1990. Enzymology of cellulose degradation. *Biodegradation*, 1: 147-161. DOI: 10.1007/BF00058833
- Zhang, Y.H.P. and L.R. Lynd, 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnol. Bioeng.*, 88: 797-824. DOI: 10.1002/bit.20282