Cloning, Sequencing and Characterization of The Xylan Degrading Enzymes from *Geobacillus thermoleovorans* IT-08

Ni Nyoman Tri Puspaningsih[1)], Antonius Suwanto[2)], Maggy T. Suhartono[3)], Suminar S. Achmadi[4)], Yogiara[5)] & Tetsuya Kimura[6)]

Department of Chemistry, Faculty of Mathematics and Natural Sciences, Airlangga University, Surabaya, Indonesia[1], Department of Biology[2], Department of Chemistry[4], Faculty of Mathematics and Natural Sciences, Laboratory of Biochemistry and Microbiology, Inter-University Centre[3], Bogor Agriculture University, Bogor,Indonesia, Faculty of Biotechnology,Atma Jaya Catholic University of Indonesia, Jakarta[5], Faculty of Bioresources, Mie University, Tsu, Mie-Japan [6]

ABSTRACT

Geobacillus thermoleovorans IT-08 is a Gram positive, thermophilic bacterium that can utilize xylan as a sole source of carbon. This strain was isolated from Gunung Pancar hot spring, Bogor, West Java, Indonesia. A plasmid genomic library in *Escherichia coli* DH5α was constructed and screened for xylanase activity. One positive clone, namely DH5α (pTP510) has been isolated, sequenced and showed putative exo-xylanase (exo-xyl), β-xylosidase (xyl), and α-L-arabinofuranosidase (abfa) genes (Genebank Accession No.DQ387047, DQ345777 and DQ387046 respectively). Each gene encoded 604, 511 and 502 amino acids, respectively. The BLAST search for protein database revealed that Abfa was high similar with GH51 family Abfa of Geobacillus stearothermophilus T6, but Xyl and Exo-Xyl were slight similar with GH43 family (25-34%) respectively. The deduced protein had a molecular weight of about 70 kDa (Exo-Xyl), and 60 kDa (Xyl and Abfa). These showed good accordance with the calculated molecular weight of each protein (68.64 kDa for Exo-xyl, 57.99 kDa for Xyl and 57.03 kDa for Abfa) from deduced amino acid sequence.

Keywords : Thermophile xylanases, *Geobacillus thermoleovorans* IT-08, pTP510

INTRODUCTION

Hemicellulose is heteroglycans, associated in plant cell walls with cellulose and lignin. Xylan is the major component of hemicellulose and is composed of β-(1,4)-linked D-xylopyranose units, with branches depending on the source (usually acetyl, arabinosyl, and glucuronyl residues). The frequency and composition of the branches are dependent on the source of xylan. Birch wood xylans contain 89.3% xylose, 1% arabinose, 1.4% glucose, and 8.3% anhydrouronic acid. Rice bran neutral xylan contains 46% xylose, 44.9% arabinose, 6.1% galactose, 1.9% glucose, and 1.1% anhydrouronic acid. Wheat arabinoxylan contains 65.8% xylose, 33.5% arabinose, 0.1% mannose, 0.1% galactose, and 0.3% glucose. Corn fiber hemicellulose contains 48-54% xylose, 33-35% arabinose, 5-11% galactose and 3-6% glucuronic acid (Beg *et al*. 2001, Saha 2003). The glycosidic bond between two sugars is one of the most stable bonds in nature, and its enzymatic hydrolysis carried out by glycoside hydrolase provide an acceleration rate which can be as high as 10^{17}-fold (Wolfenden *et al*. 1998). That is why the hemicellulase enzymes have play an important role to convert hemicelluloses biomass to be useful materials.

The total degradation of xylan by enzymes requires endo-β-1,4-xylanase, β-xylosidase, α-L-arabinofuranosidase, α-glucuronidase, acetylxylan esterase, ferulic acid esterase, and *p*-coumaric acid esterase (Subramaniyan & Prema 2002). In recent years, xylan-degrading enzymes have received much attention because of their practical applications in various agro-industrial processes. These include efficient conversion of hemicellulosic biomass to fuels and chemicals, delignification of paper pulp, digestibility enhancement of animal feedstock, clarification of juices, manufacture of bread, food and drink, pharmaceuticals, textiles, and improvement in consistency of beer (Kulkarni *et al*. 1999, Beg *et al*. 2001, Polizeli *et al*. 2005).

Purification and characterization of xylanolytic-thermophilic enzymes from *Bacillus thermoleovorans* IT-08, isolated from Gunung Pancar Hotspring, Bogor-West Java, Indonesia have been done (Puspaningsih *et al*. 2003). Characterization of the enzymes indicated that β-xylosidase showed an optimum

pH of 6.0, pH stability was 5-7, an optimum temperature of 70°C. The α-L-arabinofuranosidase showed optimum pH of 7.0, pH stability was 5-9, and optimum temperature of 70°C. Those results assumed that *Bacillus thermoleovorans* IT-08 produce xylanase complex enzymes. To explore those phenomenons, here we describe cloning, expression and analysis of the gene encoding the xylanase complex enzymes from thermophilic bacteria *Geobacillus thermoleovorans* IT-08 and the characteristics of recombinant enzyme.

MATERIALS AND METHODS

Bacterial strains and plasmid

B. thermoleovorans IT-08 was originally isolated from Gunung Pancar hot spring, Bogor, West Java, Indonesia by Irawan Tan (Tan I *et al*. 2001). It was used as chromosomal DNA source. *E.coli* DH5α was used as the host for DNA manipulation. The plasmid pBluescript II KS(+) was used for cloning, DNA sequencing and expression.

Isolation of chromosomal DNA, construction of the gene library, and transformants selection

(1). Chromosomal DNA was isolated by Freeze thaw method (Sakka *et al*. 1990). This DNA was partially digested with *Sau*3AI and fragments of 4-10 kbp were isolated by agarose gel electrophoresis followed by electroelution. (2). Libraries of these fragments were constructed by ligation into dephosphorilated *Bam*HI site of pBluescript II KS(+) and was transformed into *E.coli* DH5 α. Clones with ampicillin resistance and white colonies phenotypic were examined to xylanolytic enzymes production. (3). Xylanase positive clones were examined by using overlaid method with 0.5% agar and 3% oat spelt xylan in 5 ml PC (phosphate citrate) buffer pH 6.5 and then incubated at 60°C overnight. Xylosidase positive clones were examined by using overlaid method with 0.5% agar and 0.2% 4-methylumbelliferyl-β-D-xylopyranoside in 5 ml phosphate-citrate (PC) buffer pH 6.5 and then incubated at 60°C for 3 hours until overnight (Sakka *et al*. 1990).

Characterization of recombinant DNA

Single and double digestions of recombinant plasmids were carried out using restriction enzymes to construct the restriction maps (Sambrook 2001). The nucleotide sequences were analyzed by the dideoxynucleotide-chain-terminator method, using Automatic DNA Sequencer (Beckman Coulter, USA). Sub-cloning and primer walking were constructed to get full gene sequences. Computer analysis of the DNA and amino acid sequences was carried out using the NCBI.

Characterization of recombinant enzymes

Recombinant xylanolytic enzymes were isolated by ultrasonication and the supernatant was used as crude enzymes. *E.coli* DH5 α [pBKSII(+)] was used as a control. Assay of the recombinant enzymes were done as follow. For xylanase analysis, enzyme solution (0.1 ml) was added to 1% oat spelt xylan suspension (0.1ml) in 50mM PC buffer pH 6.5 and the mixture were incubated at 70°C for 30 minutes. Reducing sugar was determined by DNS method (Miller, 1959). Colour development was measured spectrophotometrically at 550 nm. Furthermore β-xylosidase and α-L-Arabinosidase were also characterized. Enzyme solution (0.1 ml) was mixed with 0.9 M of p-nitrophenyl derivative substrate (1ml) in 50 mM PC buffer pH 6.5 and incubated at 70°C for 30 minutes. Reaction was terminated by adding 0.4 M Na_2CO_3 (500 µl). Colour development was measured spectrophotometrically at 405 nm. The p-nitrophenyl (pNP) derivative specific substrates were pNP- β-D-Xylopyranoside (pNP-X) for β-Xylosidase, pNP-α-L-Arabinofuranoside (pNP-A) for α-L-Arabinofuranosidase. Optimum pH of β-Xylosidase and α-L-Arabinofuranosidase was measured by incubating the assay mixtures with different pHs (50 mM phosphate-citrate buffer for pH 4.0; 5.0 and 6.0; 50 mM phosphate buffer for pH 6.0, 7.0 and 8.0; 50mM glycine-NaOH buffer for pH 8.0, 9.0 and 10.0). pH stability of β-Xylosidase and α-L-Arabinofuranosidase was measured by incubating the enzymes solution in variation buffer (pH 4.0 – 10.0) overnight at 4°C, followed by the measurement of the residual enzymes activities under the standard condition. Estimation of molecular weight of recombinant enzymes was done by SDS-PAGE using Laemmli method (Laemmli 1970).

RESULTS AND DISCUSSION

Isolation of clones containing strain IT-08 xylanolytic genes.

During the cloning study of xylanolytic genes, about 2200 clones of *Bacillus thermoleovorans* IT-08 genomic libraries were screened. It was found one clone that showed weak xylan degradation activity, but high xylosidase and arabinofuranosidase activities. The clone was named pTP510 (Figure 1a.). The sub-clone was constructed to reveal pSA1 and pSA2 (Figure 1 b and 1 c) respectively and used to determine xylanolytic genes sequences.

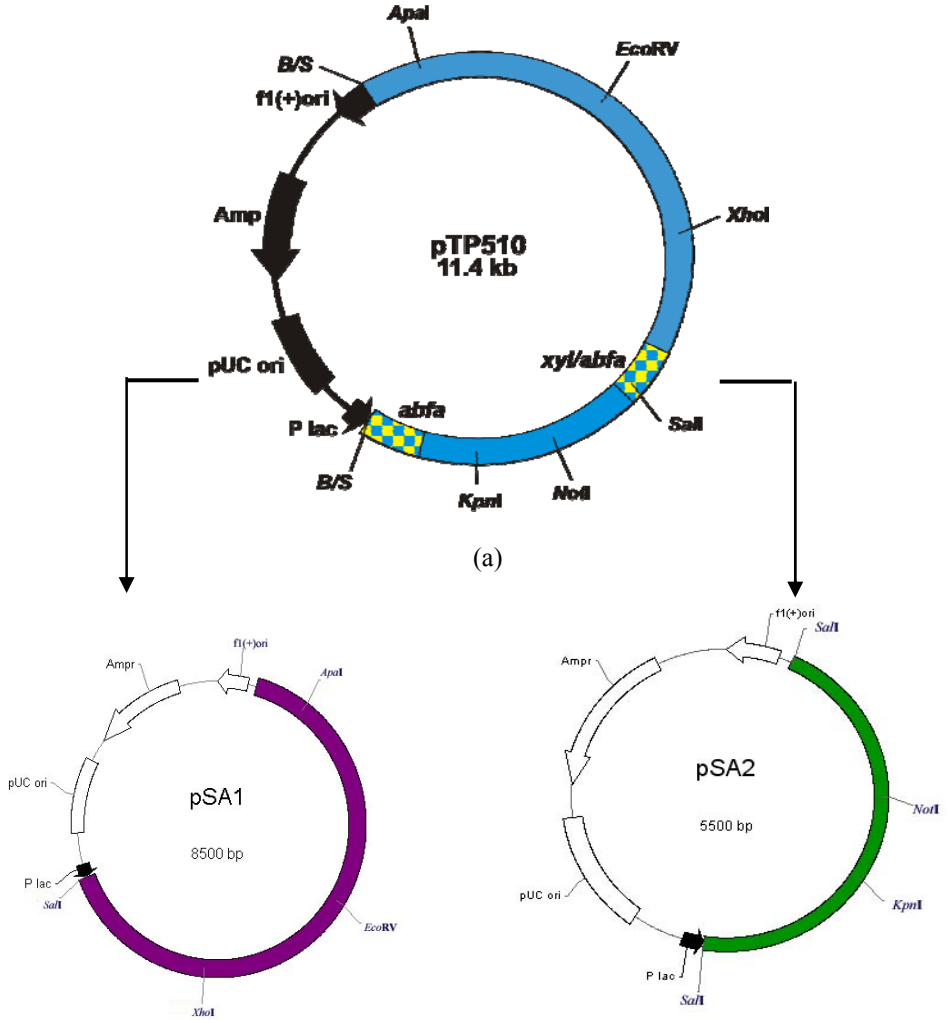

Figure 1. (a) Physical map of pTP510, (b) pSA1 and (c) pSA2

Nucleotide sequences analysis.

Partial sequencing was done on pTP510, pSA1, and pSA2 at 5' and 3'-end. Then the full nucleotide sequences of 8kb inserted gene in pTP510 was continued using 22 primer walking technique. Whole sequences revealed that pTP510 contained 5 inserted genes with nucleotide amount were totally 8444 bp (Figure 2). Those genes were encoding transposase, ABC permease, Exo-xylanase, β-xylosidase, and α-L-arabinofuranosidase respectively.

The three xylanolytic genes in pTP510 were submitted and accepted in Genebank with accession no.DQ387047 for putative exo-xylanase, DQ345777 for β-xylosidase , and DQ387046 for α-L-arabinofuranosidase respectively. The BLAST search for protein database revealed that Abfa was high similar with GH51 family Abfa of *Geobacillus stearothermophilus* T6 (90%), but Xyl and Exo-Xyl were slight similar with GH43 family (25-34%) respectively.

Expression of recombinant clones.

Cell free extracts of the recombinant clones were assayed for various xylanolytic enzyme activities. The result are summarized in Figure 3. The recombinant enzymes produced by DH5α (pTP510) could hydrolyzed oat spelt xylan, birchwood, pNP-X and pNP-A. The recombinant Xyl produced by DH5α (pSA1) could hydrolyzed birchwood xylan and pNP-X, but it could not hydrolyzed oat spelt xylan and pNP-A.

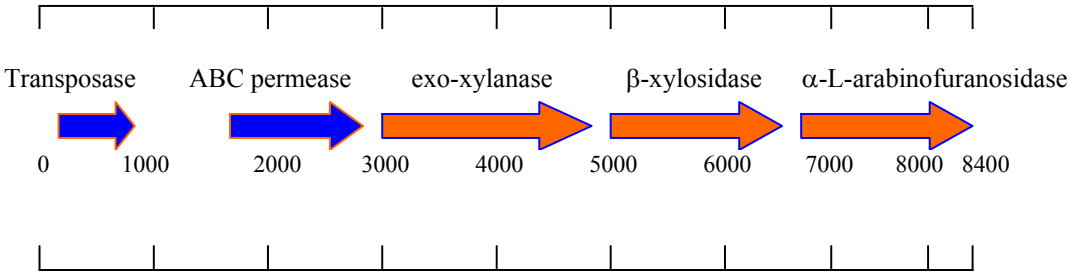

Figure 2. Diagram of whole genes inserted in pTP510. Xylanolytic genes present in pTP510 were exo-xylanase, β-xylosidase, and α-L-arabinofuranosidase

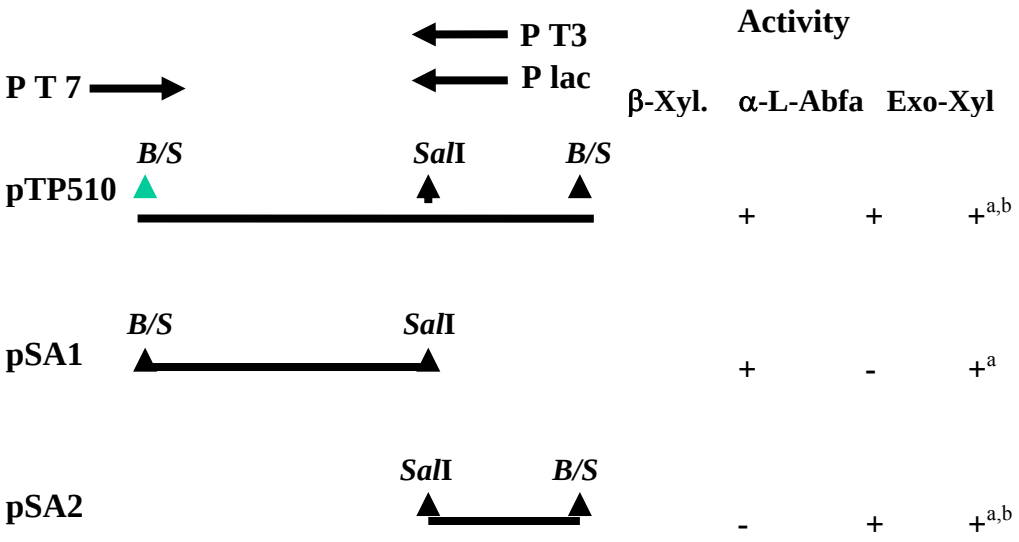

Figure 3. Xylanolytic activities of recombinant clones pTP510 and its derivatives. *B/S* :*Bam*HI/*Sau*3AI, β-Xyl : β-xylosidase, α-L-Abfa :α-L-arabinofuranosidase, Exo-xyl: Exo-xylanase, a: birchwood, and b: oat spelt xylan.

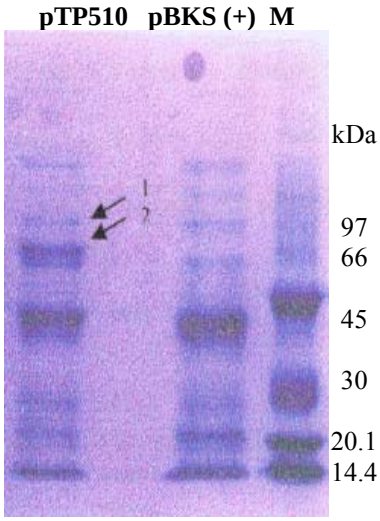

Figure 4. SDS-PAGE result to determine estimated molecular weight of gene encoding xylanolytic enzymes from pTP510. 60 and 70 kDa were estimated of Xyl Abfa and Exo-xyl. Molecular Weight. M : Protein Marker.

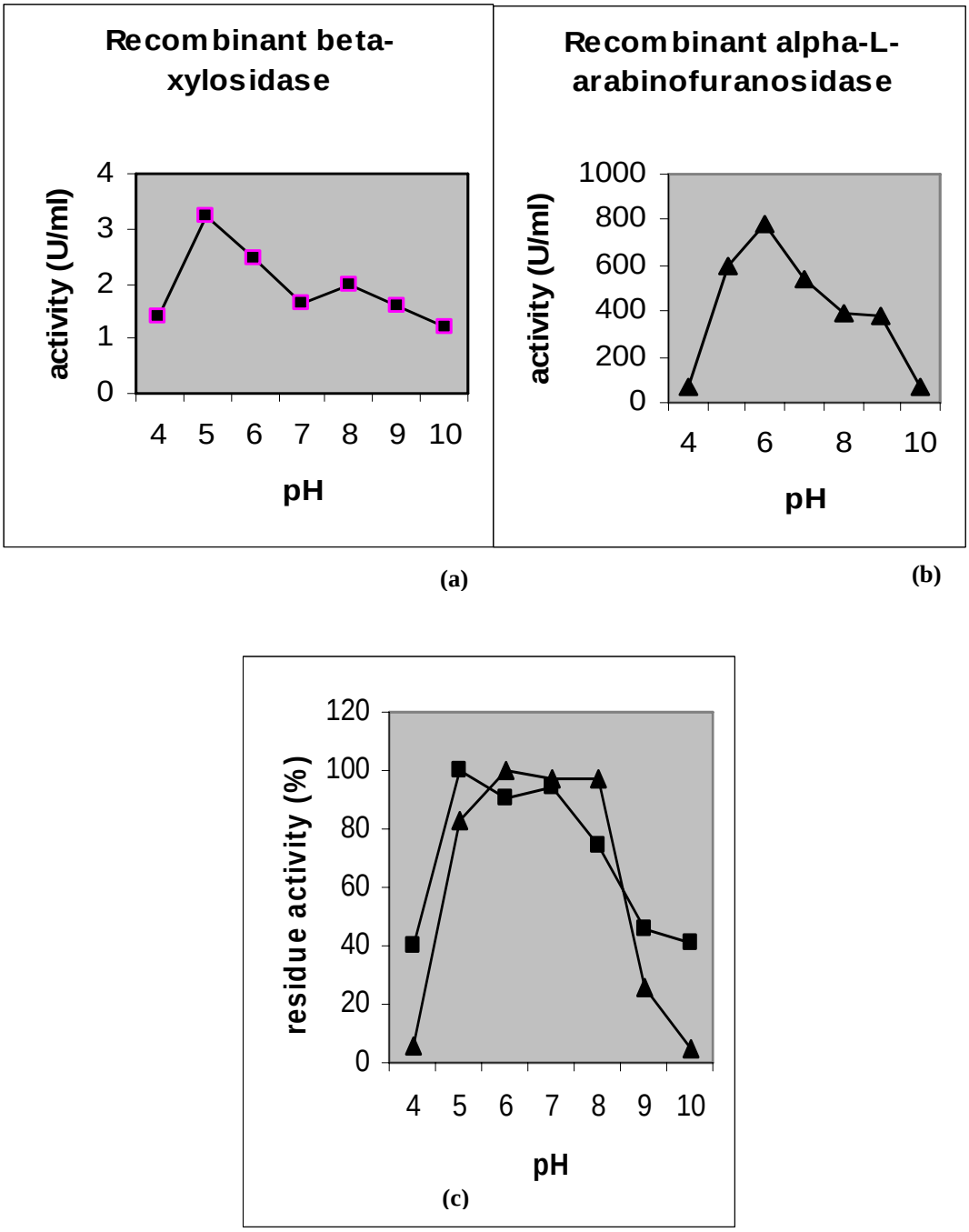

Figure 5. Optimum pH of (a) β-xylosidase, (b) α-L-arabinofuranosidase and (c) pH stability of β-xylosidase (-■-) and α-L-arabinofuranosidase (-▲-)

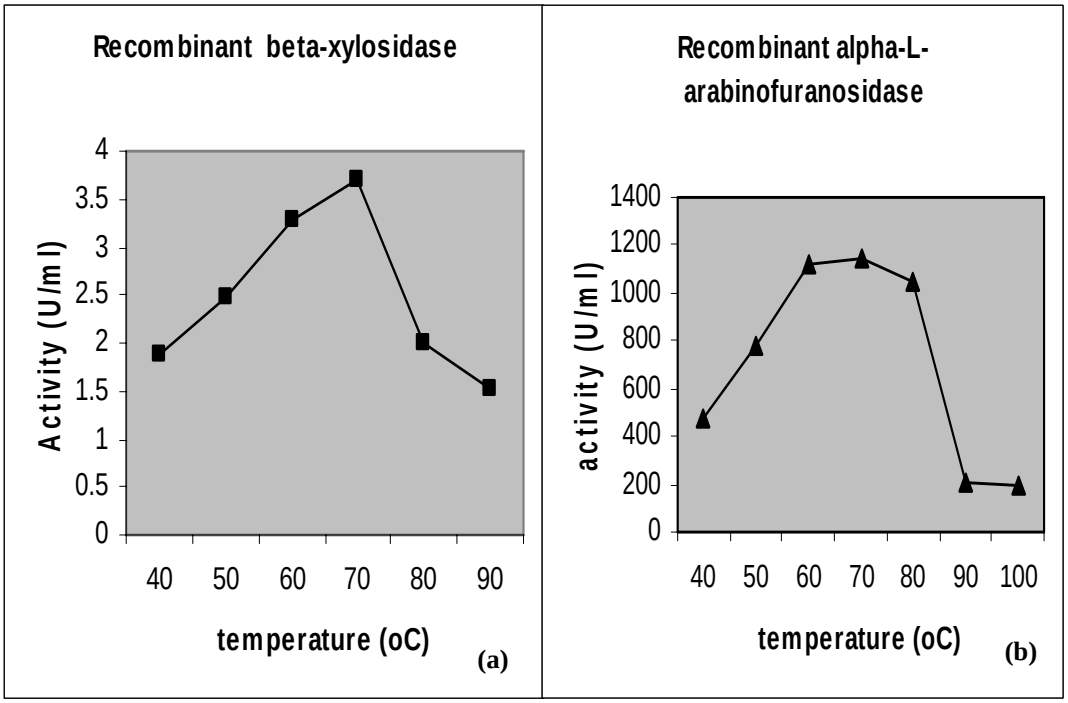

Figure 6. Optimum temperature of (a) β-xylosidase and (b) α-L-arabinofuranosidase

Although the recombinant Abfa produced by DH5α (pSA2) could hydrolyzed oat spelt xylan, birchwood xylan, and pNP-a, but could not hydrolyzed pNP-X. Those results assumed that synergistic mechanism between xylanolytic complexes enzymes were needed to hydrolyze various xylan substrates. Analysis by SDS-PAGE revealed that the deduced protein had a molecular weight of about 70 kDa (Exo-Xyl), and 60 kDa (Xyl and Abfa).The protein expressed from pBKS II (+) was used as control and showed no 60 and 70 kDa band. The same molecular weight of Abfa and Xyl will be interested for further characterization in the future research plan. The optimum pH of β-xylosidase was 5.0, and α-L-arabinofuranosidase was 6.0 (Figure 5a and 5b). pH stability of both enzymes were 5.0 – 8.0 (Figure 5c). Both enzymes showed the same optimum temperature 70^{o}C respectively (Figure 6).

Acknowledgements

This research was part of JSPS Scientific Exchange Program-Japan, and RUTXII, Ministry of Research and Technology, Indonesia.

REFERENCES

Beg QKM. 2001. Microbial xylanase and their industrial application: a review, *Appl Microbiol Biotechnol* **56**: 326-338.

Kulkarni NA. 1999. Molecular and hydrological aspects of xylanases. *FEMS Microbiol Rev* **23**: 411-456.

Laemmli UK. 1970. Cleavage of structural protein during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Miller LG. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar, *Anal Chem* **31**: 426-428.

Polizeli MTLM. 2005. Xylanases from fungi : properties and industrial applications, *Appl. Microbiol.Biotechnol.* **67**:557-591.

Puspaningsih, NNT. 2004. Isolation and characterization of β-xylosidase and α-L-arabinofuranosidase from *Bacillus thermoleovorans* IT-08. *Mie Bioforum* 2003, Ise-shima, Japan,2003. In : Ohmiya,K. et al., Biotechnology of Lignocellulose Degradation and Biomass Utilization. Tokyo, Japan, Uni Publishers Co,Ltd., **p**:442-450.

Saha BC. 2003. Hemicellulose bioconversion, *J Ind Microbiol Biotechnol* **30**: 279-291.

Sakka K. 1990. Cloning and expression in *E. coli* of *Clostridium stercorarium* strain F9 genes related to xylan hydrolysis, *Agric Biol Chem*. **54**: 337-342.

Sambrook J. 2001. *Molecular Cloning. A Laboratory Manual* 3rd. ed. New York: Cold Spring Harbor Laboratory Press.

Subramaniyan S & Prema P. (2002). Biotechnology of Microbial Xylanases : Enzymology, Molecular Biology, and Application, *Critical Rev Biotechnol*. **22**: 33-64.

Tan I. 2001. Characterization of xylanase from a xylanolytic thermophilic bacteria isolated from Gunung Pancar Hot Spring, *Biotropia* **17**:1-8.

Wolfenden, R. 1998. *J.Am.Chem.Soc*. **120**: 6814-6815.

Attachment. Genes encoding xylanases in pTP510. (a). Gene encoding *exo-xyl* was located at 2654 – 4468 (1815 bp); (b). *xyl* gene at 4824 – 6359 (1536 bp), and (c). *abfa* gene at 6626 – 8134 (1509 bp).

```
2654 atg act tta cag acg aat aaa aaa tca aat tat ata ctt tgt tat
      M   T   L   Q   T   N   K   K   S   N   Y   I   L   C   Y
2699 acc aga cta cca aaa gaa gat atc att tat tcg gcc aaa tta gct
      T   R   L   P   K   E   D   I   I   Y   S   A   K   L   A
2744 tat agt atg cat ctt gct tac agt aat gat ggc ata aat ttt gaa
      Y   S   M   H   L   A   Y   S   N   D   G   I   N   F   E
2789 ccg tta aat cac aat tct gga att tta ttt gca aag gct act gaa
      P   L   N   H   N   S   G   I   L   F   A   K   A   T   E
2834 aat gag aac ggt tcc ctt aac gca aag agt tta aaa aat cca tat
      N   E   N   G   S   L   N   A   K   S   L   K   N   P   Y
2879 ata ttc cat ttg aaa gac ggg aat ttc gga gtc ata gct gtt cgg
      I   F   H   L   K   D   G   N   F   G   V   I   A   V   R
2924 act aaa cca gat ggt aca aac gat gag gag agt aaa gga aaa gtc
      T   K   P   D   G   T   N   D   E   E   S   K   G   K   V
2969 ctt att ttt tca tcc ccg gat tta ttg caa tat aaa gaa att gga
      L   I   F   S   S   P   D   L   L   Q   Y   K   E   I   G
3014 ttg cta gac ttg aag gga aac act ttt gtt aat gac gta gtt tgt
      L   L   D   L   K   G   N   T   F   V   N   D   V   V   C
3059 cat tac gac gat gaa aaa aag gtt tat ctc atc aaa tgg agt gat
      H   Y   D   D   E   K   K   V   Y   L   I   K   W   S   D
3104 ggt ttg ggg aat tat tat atg aat ttg ata gag gac att atc act
      G   L   G   N   Y   Y   M   N   L   I   E   D   I   I   T
3149 ctg aat cat atc tct gat cct aaa ccg att gaa tct ttt cag tta
      L   N   H   I   S   D   P   K   P   I   E   S   F   Q   L
3194 gat tct gtc caa aca aaa ata gaa gga atc gta cct aga aat gtt
      D   S   V   Q   T   K   I   E   G   I   V   P   R   N   V
3239 ata caa gtt tcg gag gaa att gcc cgc aga tta att tgt aaa cta
      I   Q   V   S   E   E   I   A   R   R   L   I   C   K   L
3284 act gta ccc aca aat att aaa ata gaa gtt ccc gaa aag gta gca
      T   V   P   T   N   I   K   I   E   V   P   E   K   V   A
3329 gta aaa act gag aaa gat tta gaa aat gtg aaa gca atc gca ata
      V   K   T   E   K   D   L   E   N   V   K   A   I   A   I
3374 tac agt gat ggt aca acc gat aca aag cgt gtt aat tgg gat aag
      Y   S   D   G   T   T   D   T   K   R   V   N   W   D   K
3419 aat ggt att gat tgg agc aaa cca ggg act tat aga ata act ggt
      N   G   I   D   W   S   K   P   G   T   Y   R   I   T   G
3464 aca gta tac caa gat cat tac tca ttc ccg att gct att gat cga
      T   V   Y   Q   D   H   Y   S   F   P   I   A   I   D   R
3509 gcc gat ccg tgt atc act aaa tgg aat gga aag tac tat ttc ata
      A   D   P   C   I   T   K   W   N   G   K   Y   Y   F   I
3554 gcc acg aat gat gca gac gga aat cat tct tta tac atc aag gaa
      A   T   N   D   A   D   G   N   H   S   L   Y   I   K   E
3599 gcc gac acg att cct gga tta gtt gat gct gag gaa ata ttg ata
      A   D   T   I   P   G   L   V   D   A   E   E   I   L   I
3644 ctt gat tct gac aca tac gaa gat att aaa ggc tta tta tgg gcg
      L   D   S   D   T   Y   E   D   I   K   G   L   L   W   A
3689 cca gag ttt cat atc att gaa ggc gat ctt tac atc ttt cat ggt
      P   E   F   H   I   I   E   G   D   L   Y   I   F   H   G
3734 gcc aca tca aat ggg ttt tac tat gaa caa tca cac gtt atg aaa
      A   T   S   N   G   F   Y   Y   E   Q   S   H   V   M   K
3779 tta cga aaa ggc ggg aat cca gtt tgt gca aaa gat tgg tca agg
      L   R   K   G   G   N   P   V   C   A   K   D   W   S   R
3824 cca tat cgt gtt gta aag aaa gat ggc aca tat ttg tgt gaa gca
      P   Y   R   V   V   K   K   D   G   T   Y   L   C   E   A
3869 ggg aag act ata tct ttg gat atg act gta att aaa tgg aat gaa
      G   K   T   I   S   L   D   M   T   V   I   K   W   N   E
3914 gaa tat tat gtg gtt tgg tca cag agg gag ttt ata cct gag gat
      E   Y   Y   V   V   W   S   Q   R   E   F   I   P   E   D
```

```
3959 ctt gga gca tgg ctt tac att gca aag gtt gat cca aag gag cct
      L   G   A   W   L   Y   I   A   K   V   D   P   K   E   P
4004 tgg agg cta gtt tgc gat cca gtt gtt ctt act aaa ccg gaa tat
      W   R   L   V   C   D   P   V   V   L   T   K   P   E   Y

4049 gga tgg gaa aat aac aac gtt ttt gta gtt gaa gga cct ttc gcc
      G   W   E   N   N   N   V   F   V   V   E   G   P   F   A
4094 ttg att agg aat aac aaa tta ttc ttg aca tac tct ggt tcc ttg
      L   I   R   N   N   K   L   F   L   T   Y   S   G   S   L
4139 atc gat gaa act tac gtt att gga ttg tta acg gca gaa aaa ggc
      I   D   E   T   Y   V   I   G   L   L   T   A   E   K   G
4184 gca gat ttg tta aat cct gct tct tgg aca aag tgt aat tac ccg
      A   D   L   L   N   P   A   S   W   T   K   C   N   Y   P
4229 ttg ctt act tct agg agt gtc ccg gga gaa tat ggt ccc gga cat
      L   L   T   S   R   S   V   P   G   E   Y   G   P   G   H
4274 aat tcc tat gtg atg gac gat tac gga act ata tgg aac gtc tat
      N   S   Y   V   M   D   D   Y   G   T   I   W   N   V   Y
4319 cac gca agg cca ggt ata aaa ggt cct cgg tcc tcc ggg att cgg
      H   A   R   P   G   I   K   G   P   R   S   S   G   I   R
4364 cgt gta cat ttt gat ata gat ggc tat cca aga cta gat tta aca
      R   V   H   F   D   I   D   G   Y   P   R   L   D   L   T
4409 gaa gat aag gat ctc aat agg gat tta acg aaa gtc gct atg gat
      E   D   K   D   L   N   R   D   L   T   K   V   A   M   D
4454 gtc att ttg act taa 4468
      V   I   L   T   *
```

(a)

```
4824 atg gaa tat tct aac cca gta att aaa ggg ttt tat ccg gat ccc
      M   E   Y   S   N   P   V   I   K   G   F   Y   P   D   P
4869 agt att tgt cga gta ggt agt gat tat tac tta gtt aca agt tca
      S   I   C   R   V   G   S   D   Y   Y   L   V   T   S   S
4914 ttc cag tac ttc cct ggg gtt cca att ttt cat agt act aat tta
      F   Q   Y   F   P   G   V   P   I   F   H   S   T   N   L
4959 att aat tgg aat aag ata gga tat tgt tta att aga cca agt caa
      I   N   W   N   K   I   G   Y   C   L   I   R   P   S   Q
5004 ctt atg tta aat aat gca aca aat aga agt ggt ata ttt gca cct
      L   M   L   N   N   A   T   N   R   S   G   I   F   A   P
5049 acc ctt cgt tat cat gag gga att ttt tat tta ata aca aca aac
      T   L   R   Y   H   E   G   I   F   Y   L   I   T   T   N
5094 gta aca cta aag aag aac ttt att gtt atg tcc gaa gat cta caa
      V   T   L   K   K   N   F   I   V   M   S   E   D   L   Q
5139 ggg gaa tgg tct gaa ccg att tgg att gat gga tgg gga ggc att
      G   E   W   S   E   P   I   W   I   D   G   W   G   G   I
5184 gat cca tca cta ttt ttt gat aac gat ggg aag gtt tat att acc
      D   P   S   L   F   F   D   N   D   G   K   V   Y   I   T
5229 ggg aca aat gat aat gct agg ggt gaa gaa tta gga att tac caa
      G   T   N   D   N   A   R   G   E   E   L   G   I   Y   Q
5274 gca gaa ata gat tta aag aaa gga agt att ata ggt gaa aga aaa
      A   E   I   D   L   K   K   G   S   I   I   G   E   R   K
5319 ctc ata tgg aaa ggt aca ggc ggc tct tac ccg gaa gcc ccc cat
      L   I   W   K   G   T   G   G   S   Y   P   E   A   P   H
5364 tta tat aaa gtt aat ggc tgg tat tat tta tta atc gca gaa gga
      L   Y   K   V   N   G   W   Y   Y   L   L   I   A   E   G
5409 ggt aca gag tat ggt cat atg gtg acc gtt gca agg agt aaa tat
      G   T   E   Y   G   H   M   V   T   V   A   R   S   K   Y
5454 ccc ttc ggt cct ttc gaa agt tgt cct ttt aat cca ata tta act
      P   F   G   P   F   E   S   C   P   F   N   P   I   L   T
5499 cat aga agc aca aac cat cct ctt cag gca atc ggt cat gct gat
      H   R   S   T   N   H   P   L   Q   A   I   G   H   A   D
5544 att gtt cag tat cat gac gga agt tgg tgg gca gtt ttt cac ggt
      I   V   Q   Y   H   D   G   S   W   W   A   V   F   H   G
```

```
5589 act cgt ccc atc tct tat cca ccg aaa cac cat ttg ggc aga gag
      T   R   P   I   S   Y   P   P   K   H   H   L   G   R   E
5634 act tgt tta gct cct atc aag tgg aca gac gat ggt tgg cct att
      T   C   L   A   P   I   K   W   T   D   D   G   W   P   I
5679 att ggt tac aac gga aga att gat att aaa atg gat gct ggt tat
      I   G   Y   N   G   R   I   D   I   K   M   D   A   G   Y
5724 ctg cct gtg aaa gaa aaa aat att ggg gat gag atc att gaa gat
      L   P   V   K   E   K   N   I   G   D   E   I   I   E   D
5769 gat ttt aac agt gat att ttt tct aca gat tgg aat ttt att caa
      D   F   N   S   D   I   F   S   T   D   W   N   F   I   Q
5814 aac cct cgc ctt gaa cac tat tct ttg aag gga cgt cct agt tgg
      N   P   R   L   E   H   Y   S   L   K   G   R   P   S   W
5859 tta aaa atg cgg ggt aca gaa aaa aca ttg aat gat ata aat tcc
      L   K   M   R   G   T   E   K   T   L   N   D   I   N   S
5904 cca acg ttt att ggg cgg cgc caa gaa cat ttt gtt tgt aat gtg
      P   T   F   I   G   R   R   Q   E   H   F   V   C   N   V
5949 tcg aca tta tta gaa ttt aaa ccg aat cag gat aat gag gaa gct
      S   T   L   L   E   F   K   P   N   Q   D   N   E   E   A
5994 ggg cta acc gtt tat atg aat gaa aag cac cac tat gaa att gcc
      G   L   T   V   Y   M   N   E   K   H   H   Y   E   I   A
6039 cta aca aag aaa aat gga cga ata aat gta gtt ttg aag aaa act
      L   T   K   K   N   G   R   I   N   V   V   L   K   K   T
6084 gta ggg gat att cag gtt gtt gta aat tca tta gag tat ttc tct
      V   G   D   I   Q   V   V   V   N   S   L   E   Y   F   S
6129 aat acg att att ttt tct att caa gct aat ccg gaa gaa tac aag
      N   T   I   I   F   S   I   Q   A   N   P   E   E   Y   K
6174 ttt tca ttt gtt gat cct aat aca ggt cag act tat cta tta gga
      F   S   F   V   D   P   N   T   G   Q   T   Y   L   L   G
6219 aca gga ctt act aca ctt tta tct acg gag gtt gca gga ggg ttc
     T  G  L  T  T  L  L  S  T  E  V  A  G  G  F
6264 aca ggc gtt tac ttt ggg tta tat gcc act ggt aat gga aaa gtt
      T   G   V   Y   F   G   L   Y   A   T   G   N   G   K   V
6309 tgt acg gct ccc gcc ttt ttt gat tgg ttt aaa tat att cct gaa
     C  T  A  P  A  F  F  D  W  F  K  Y  I  P  E
6354 ata tag  6359
      I   *
```

(b)

```
6626 atg gct aca aaa aaa gca acc atg atc atc gaa aaa gac ttc aaa
      M   A   T   K   K   A   T   M   I   I   E   K   D   F   K
6671 att gct gaa atc gac aaa cgc atc tat ggc tcg ttt att gaa cac
      I   A   E   I   D   K   R   I   Y   G   S   F   I   E   H
6716 ctc ggc cgc gcg gta tac ggg ggg att tat gag ccg agc cat ccg
      L   G   R   A   V   Y   G   G   I   Y   E   P   S   H   P
6761 cag gcc gat gaa aac ggc ttc cgg cag gat gtc att gaa atg gtg
      Q   A   D   E   N   G   F   R   Q   D   V   I   E   M   V
6806 aaa gag tta caa gtg ccc att atc cgc tat ccg ggc ggg aat ttt
      K   E   L   Q   V   P   I   I   R   Y   P   G   G   N   F
6851 gtg tcc ggt tac aac tgg gag gac gga gtc ggg cca aaa gaa aag
      V   S   G   Y   N   W   E   D   G   V   G   P   K   E   K
6896 cgg ccg cgg cgg ctt gat ttg gca tgg aag tca gtg gaa acg aat
      R   P   R   R   L   D   L   A   W   K   S   V   E   T   N
6941 gaa att ggc ttg aat gaa ttt gtc gat tgg gcc aag atg gtc gga
      E   I   G   L   N   E   F   V   D   W   A   K   M   V   G
6986 gcc gaa gtg aat atg gcc gtc aac tta ggg acg cgc ggc att gat
      A   E   V   N   M   A   V   N   L   G   T   R   G   I   D
7031 gcg gca cgc aac ttg gtt gaa tat tgc aac cac ccg tcg ggc tcg
      A   A   R   N   L   V   E   Y   C   N   H   P   S   G   S
7076 tat tac agc gat ttg cgc att tcc cac ggc tat aaa gag ccg cat
      Y   Y   S   D   L   R   I   S   H   G   Y   K   E   P   H
7121 aaa att aaa aca tgg tgt cta ggc aat gag atg gac ggt ccg tgg
```

```
      K   I   K   T   W   C   L   G   N   E   M   D   G   P   W
7166 caa att ggc cac aag aca gcc gtt gag tac gga cga atc gct tgt
      Q   I   G   H   K   T   A   V   E   Y   G   R   I   A   C
7211 gaa gcg gcc aaa gtg atg aaa tgg gta gat ccg acc att gaa ctt
      E   A   A   K   V   M   K   W   V   D   P   T   I   E   L
7256 gtt gcg tgc gga agt tca ggc aga aat atg ccg acg ttt gcg gaa
      V   A   C   G   S   S   G   R   N   M   P   T   F   A   E
7301 tgg gaa gcg acg gtt ctt gat cac acg tat gag cat gtc gat tat
      W   E   A   T   V   L   D   H   T   Y   E   H   V   D   Y
7346 att tcc ctc cat caa tac ttt gga aat cga gat aat gac acg gcg
      I   S   L   H   Q   Y   F   G   N   R   D   N   D   T   A
7391 aat tat ttg gcg ctg tcg ctg gaa atg gat gat ttt atc cgt tcg
      N   Y   L   A   L   S   L   E   M   D   D   F   I   R   S
7436 gtt gtg gcc att gcc gat tac gtg aag gcg aaa aaa cga agc aag
      V   V   A   I   A   D   Y   V   K   A   K   K   R   S   K
7481 aag acg att cat ctg tcg ttt gac gaa tgg aac gta tgg tac cac
      K   T   I   H   L   S   F   D   E   W   N   V   W   Y   H
7526 tcg aat gag gcg gat aag caa att gaa ccg tgg acc gtc gcg ccg
      S   N   E   A   D   K   Q   I   E   P   W   T   V   A   P
7571 cct ttg ttg gag gat att tat aac ttt gaa gat gcg cta ctt gtc
      P   L   L   E   D   I   Y   N   F   E   D   A   L   L   V
7616 ggc tgc atg ctc att acg ctc atg aaa cat gcc gat cgg gtg aaa
      G   C   M   L   I   T   L   M   K   H   A   D   R   V   K
7661 att gcc tgc ttg gct cag tta gtg aat gtc att gca ccg atc atg
      I   A   C   L   A   Q   L   V   N   V   I   A   P   I   M
7706 acg gaa ccg aac ggg ccg gca tgg aag caa acc att tac tat ccg
      T   E   P   N   G   P   A   W   K   Q   T   I   Y   Y   P
7751 ttt atg cat gcc tcg gtt tac ggc aga ggg gtg gcg ttg cac cca
      F   M   H   A   S   V   Y   G   R   G   V   A   L   H   P
7796 gtt att tca agc ccg aaa tac gac agc aaa gac ttc aca gat gtt
      V   I   S   S   P   K   Y   D   S   K   D   F   T   D   V
7841 ccg tat tta gag tcg atc gct gtt tac aat gaa gaa aaa gaa gaa
      P   Y   L   E   S   I   A   V   Y   N   E   E   K   E   E
7886 gtg acg att ttt gcg gtc aac cgt gat atg gac gat tcg tta ttg
      V   T   I   F   A   V   N   R   D   M   D   D   S   L   L
7931 ctt gaa tgc gat gtc cgc cat ttt gac gat tat cgc gtt att gaa
      L   E   C   D   V   R   H   F   D   D   Y   R   V   I   E
7976 cat atc gta ttg gaa cat gaa aac gtg aaa caa acg aat tcc gcg
      H   I   V   L   E   H   E   N   V   K   Q   T   N   S   A
8021 caa tct tcc ccg gtc gtt ccg cac cgc aac ggc gat gct caa cta
      Q   S   S   P   V   V   P   H   R   N   G   D   A   Q   L
8066 tcc ggc ggg aaa gtg tcg gcg acg ttg tcg aag tta tcg tgg aat
      S   G   G   K   V   S   A   T   L   S   K   L   S   W   N
8111 gtg att cgt tta gga aaa cga taa 8134
      V   I   R   L   G   K   R   *
```

(c)