Isolation of gene encoding xylanase from bagasse compost soil using sequence-based approach

By

Miss Panutda Mangkorn

A special project submitted to the Faculty of Biotechnology
Assumption University in part fulfillment of the requirements for
the degree of Bachelor of Science in Biotechnology

2011

Isolation of gene encoding xylanase from bagasse compost soil using sequence-based approach

By

Miss Panutda Mangkorn

A special project submitted to the Faculty of Biotechnology
Assumption University in part fulfillment of the requirements for
the degree of Bachelor of Science in Biotechnology
2011

Special Project

Isolation of gene encoding xylanase from bagasse compost soil using sequence-based approach

By

Miss Panutda Mangkorn

Year 2011

Title	**:**	**Isolation of gene encoding xylanase from bagasse compost soil using sequence-based approach**

Report done by	**:**	**Panutda Mangkorn**
Project advisor	**:**	**Dr. Tatsaporn Todhanakasem**
Project co-advisor	**:**	**Dr. Somchai Pongpattanakitshote**
Department	**:**	**Agro-Industry**
Academic year	**:**	**2011**
Advisor Committee	**:**	**Dr. Viyada Kunathigan**
		Dr. Chaniku Chutrakul
		A. Sireerat Laodheerasiri

Tatsaporn Todhanakasem
..... **Advisor's Signature**

(Dr. Tatsaporn Todhanakasem)

Instructor, School of Biotechnology

..... **Co - advisor's Signature**

(Dr. Somchai Pongpattanakitshote)

Researcher, BIOTEC

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to all those who gave me the possibility to complete this senior project. I want to thank Faculty of Biotechnology, Assumption University for giving me opportunity and support to do the senior project and also for all the laboratory instruments provided in order to complete the project.

I am extremely thankful to the scholarship from Young Scientist and Technologist Program (YSTP), National Science and Technology Development Agency (NSTDA) for the financial support during (2010-2011).

I am deeply indebted to my advisor Dr. Tatsaporn Todhanakesem, instructor from the Assumption University, and Dr. Somchai Pongpattanakitshote, Researcher of Enzyme Technology Laboratory from BIOTEC for all of help, suggestions and encouragement throughout the research and writing of this project. I sincerely thank, Enzyme Technology Laboratory stuff for encouraging me, supporting me and giving me a helpful advices to pass through all the obstruct I faced during the research.

My appreciation is extended to Dr. Viyada Kunathigan, A. Sireerat Laodheerasiri, committee members in Faculty of Biotechnology, for their time during the project presentation.

Finally, I would like to express my special deepest appreciation to my beloved parents, family members and my friends for their patient, love, and support which enabled me to complete this senior project.

Panutda Mangkorn

Isolation of gene encoding xylanase from bagasse compost soil using sequence-based approach

Abstract

Hemicelluloses are among the most abundant biomasses on earth and represent a considerably immense source of fixed carbon in nature. Xylanases are one of the major hemicellulose-degrading enzymes that randomly cleave backbone xylans containing within the plant cell walls. Consequently, these enzymes have significant biotechnological potential applications in various industries. In this study, culture-independent approach using sequence-based technique was employed to isolate the xylanase genes from bagasse compost soil. To obtain the genes responsible for xylanase enzyme production, degenerate primers designed from the conserved regions of bacterial xylanase (glycosyl hydrolase family 10) genes were used. The degenerate primers were used in the PCR reaction to amplify the DNA previously extracted from bagasse compost soil. The obtained PCR products of approximately 160 bp. Sequence analysis of the partial xylanase gene exhibit 64% amino acid identity to *Geobacillus thermoglucosidasius C56-YS93.* The full-length gene determination was performed using the genome walking approach. The 3' and 5'ends were obtained. The results shown that a full-length xylanase gene contained 2,649 bp open reading frame encoding 883 amino acid residues which exhibited 43% amino acid sequences identity to Endo-1,4-beta-xylanase from *Thermotoga sp. RQ.* Additionally, a putative carbohydrate binding module family 9 was also present at the C-terminus of the gene sequence.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER I
INTRODUCTION

Lignocellulosic wastes are estimated at several billion tons annually on earth [25]. The abundant amounts of lignocellulosic wastes are generated through forestry and agricultural practices, paper-pulp industries, timber industries and many agro - industries in which they pose environmental problems. Currently, lignocellulosic materials turn to be very important as they can be converted into various different value added products including biofuels, chemicals, and cheap energy sources for fermentation, animal feeds and human nutrients. Lignocellulosic materials consist of lignin, hemicellulose and cellulose. Hemicelluloses is the second most common polysaccharides in nature, represent about 20–35% of lignocellulosic biomass. Various agricultural residues, such as corn fiber, corn stover, wheat straw, rice straw, and sugarcane bagasse contain about 20 – 40%. However, the conversion of hemicellulose to fuels and chemicals is problematic. The effective enzymes use to degrade hemicellulose still be necessary. Therefore, there is a need of the lignocellulosic materials - degrading enzymes to convert into the fuel ethanol and other value-added fermentation products including bio-pulping of wood, coffee processing, fruit and vegetable maceration, and preparation of high fiber baked goods. Xylan-degrading enzymes hold great promise in saccharifying various pretreated agricultural and forestry residues to fermentable sugars. The utilization of hemicellulosic sugars is essential for efficient conversion of lignocellulosic materials to fuel ethanol and other value-added fermentation products [17].

Xylans are the most abundant hemicelluloses. In recent years, bioconversion of hemicelluloses has received much attention because of its practical applications in various agro-industrial processes, such as efficient conversion of hemicellulosic biomass to fuels and chemicals, delignification of paper pulp, digestibility enhancement of animal feedstock, clarification of juices, and improvement in the consistency of beer. Enzymes that degrade, or help to degrade, hemicellulose are of great interest to the paper and pulp industry due to their bleach-boosting properties (biobleaching of pulp), which reduces environmentally unfriendly chlorine consumption [8].

Enzyme industry is the result of rapid development of biotechnology i.e. recombinant gene. Currently, there is a huge potential market for fiber-degrading enzymes for lignocellulosic materials degradation technology and protein engineering. With better knowledge and purification of enzymes, a number of applications have increased

exponentially. Nowadays, the use for industrial enzymes has now extended to almost all industries handling organic compounds.

The glycosyl hydrolases consisted of a large family of enzymes (xylanases, cellulases, etc.) that are of great significance in industrial processes and the degradation of lignocellulosic materials [10].

1.1 Xylanases

1.1.1 Structure of xylan

Plant cell walls have three major polymeric constituents: cellulose (insoluble fibers of β-1,4-glucan), hemicallulose (non-callulolytic polysaccharides including glucans, mannans, and xylans), and lignin (a complex polyphenolic structure). Xylan is a complex polysaccharide comprising a backbone of xylose residues linked by β- 1,4-glycosidic bonds (Figure 1), The main chain of xylan is composed of β- Xylopyranose residues. Most xylans occur as heteropolysaccharides, containing different substituent groups which are acetyl, arabinosyl and glucuronosyl residues.

1.1.2 Xylanolytic enzymes

The utilization of these wastes for the production of strategic chemicals and fuel requires hydrolysis of all the components. Because xylan is a major plant structural polymer, xylanases and the microorganisms that elaborate them could be used in food processing and paper and the pulp, sugar, ethanol, feed, and agrofiber industries (Gomes et al., 1993). For most bioconversion processes, xylan must first be converted to xylose or xylooligosaccharides. They may be achieved by acid hydrolysis or through the use of xylanolytic enzymes [29].

Xylanolytic enzymes are glycosyl hydrolases which degrade xylan. Xylanases are widespread in nature. They have been reported to be present in marin and terrestrial bacteria, rumen and ruminant bacteria, fungi, marine and terrestrial Xylanases are usually composed of repertoire of hydrolytic enzymes: β-1,4- endoxylanase, β-xylosidase α-L-arabinofuranosidase, a-glucosidase, acetyl xylan esterase, and phenolic acid (ferulic and p-coumaric acid) esterase (Figure 1). Table 1 lists the enzymes involved in the degradation of xylan and their modes of action [9].

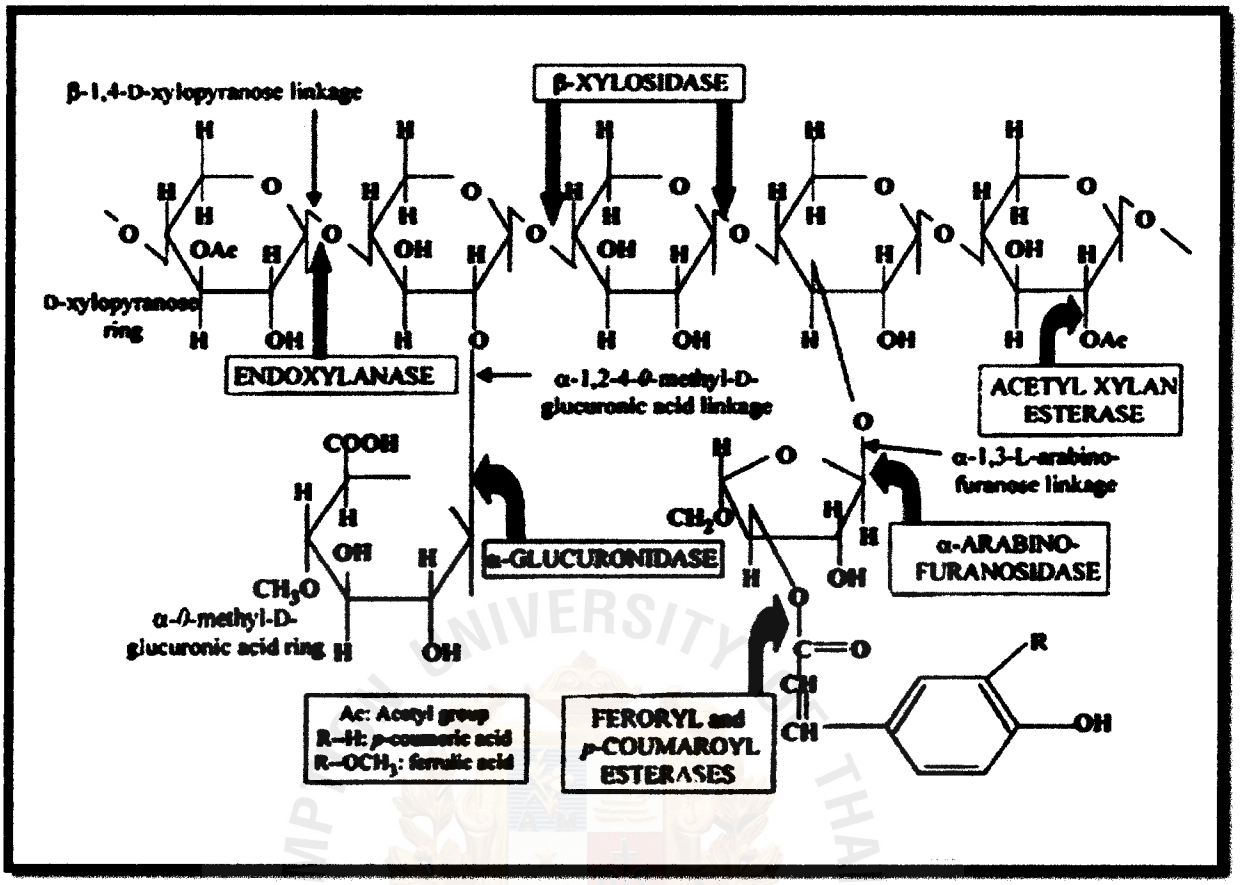

Figure 1: Xylanolytic enzymes system

Table 1: Enzymes involved in the hydrolysis of complex heteroarabinoxylans

Enzyme	Mode of action
Endo-xylanase	Hydrolyzes mainly interior β-1,4-xylose linkages of the xylan backbone
Exo-xylanase	Hydrolyzes β-1,4-xylose linkages releasing xylobiose β-Xylosidase Releases xylose from xylobiose and short chain xylooligosaccharides
α-Arabinofuranosidase	Hydrolyzes terminal nonreducing α-arabinofuranose from arabinoxylans
α-Glucuronidase	Releases glucuronic acid from glucuronoxylans
Acetylxylan esterase	Hydrolyzes acetylester bonds in acetyl xylans
Ferulic acid esterase	Hydrolyzes feruloylester bonds in xylans

1. β-1,4-Endoxylanases (1,4-β-D-xylan xylohydrolase; EC 3.2.1.8) cleave the intemal glycosidic linkages of the heteroxylan backbone, resulting in the production of xylooligosaccharides. As hydrolysis proceeds,these oligosaccharides will be further hydrolyzed to xylotriose, xylobiose, and xylose. There enzymes have been isolated from several fungi and bacteria such as *Aspergillus niger*, *Trichoderma koningii and Dictyoglomus thermophilum Rt46B.1*
2. β-D-Xylosidases (β-D-xylosids xylohydolase; EC 3.2.1.37) are exogly - cosidases that hydrolyze short xylooligosaccharides and xylobiose from the non - reducing end to liberate xylose. These enzymes have been reported in bacteria and fungi.
3. α-L-Arabinofuranosidase. There are two types of arabinases, the exo-acting α-L-arabinofuranosidase (EC 3.2.1.55), which is active against ρ-nitrophenyl - α-L-arabinoofuranoside and on branchred arabinans, and endo-1,5-a-L-arabinase (EC 3.2.1.99), which is active only toward linear arabinans. Endoarbinases have been reposted in Bacillus subtilis, Clostridium felsineum and fungi.
4. α-D-Glucosidases (3.2.1.20) hydrolyze the a-1,2 linkages between glucuronic acid and xylose residues in glucuronoxylan. These enzymes have been isolated from *Aspergillus niger* and *Streptomyces flavogriseus.*
5. Acetylxylan esterases (EC 3.1.1.6) remove the O-acetyl substituents at the C2 and C3 positions of xylose residues in acetylxylan. The production of acetylxylan esterase was found in fungi and bacteria.
6. Ferulic acid esterases (EC 3.1.1.73) cleave the ester linkages between arabinose side chains and ferulic acids in xylan. Similarly, ρ-coumaric acid esterases (EC 3.1.1.-) cleave the ester linkage between arabinose and ρ-coumaric acid. Few Ferulic and ρ-coumaric acid esterases have been purified and characterized.

1.1.3 Application of xylanases

Xylanases isolated from microorganisms have attracted a great deal of attention. In the last decade, particularly because of their biotechnological potential in various industrial processes, xylanases have shown an immense potential for increasing the production of several useful products in the most economical way. The main possibilities are the production of single call proteins (SCPs), enzymes, liquid or gaseous fuel, solvents and sugar syrups, which can be used as such or as feed stocks for other microbiological processes. Currently, the most promising application of xylanase is the per-bleaching of Kraft pulps. Enzyme application improves pulp fibrillation and water retention, reduction of beating times in virgin pulps, restoration of bonding and increased freeness in recycled fibers, and selective removal of xylans from dissolving pulps. Xylanases are also useful in yielding cellulose from dissolving pulps from rayon production and biobleaching of wood pulps [19].

In the bread and bakery industries, the efficiency of xylanases in improving the quality of bread has been shown with an increase in specific bread volume. This is further enhanced when amylase is used in combination with xylanase.

In the waste industry, xylan is present in large amount in waster from agricultural and food industries. Hence, xylanases are used for conversion of xylan to xylose in wastewater. The development of an efficient process of enzymatic hydrolysis offers new prospects for treating hemicellulosic wastes.

In the juice industry, xylanases are used concurrently with cellulase and pectinase for clarifying must and juices, and for liquefying fruits and vegetables.

In the animal - industry, depression in weight gain and feed conversion efficiency in rye-fed broiler chicks has been associated with intestinal viscosity. Incorporation of xylanase into a rye-based diet of broiler chicken results in reduced intestinal viscosity, thus improving the weight gain of chicks and their feed conversion efficiency.

In the fuel industry, xylanase acts in synergism with several other enzymes such as mannanase, ligninase, xylosidase, glucanase, glucosidase, etc. and can be used for the generation of biological fuel such as lignocellulosic biomass [28].

Moreover, agro-industrial and food-processing wastes are available in staggering quantities all over the world, which largely become a source of health hazard. The majority of these wastes contain cellulose, hemicelluloses, which consist of xylan about 20-40%, and lignin. The utilization of these wastes for the production of strategic chemicals and fuel are very important due to the shortage and high price of oils. With this problem, the production

of fuel requires hydrolysis of all the components in order to convert the xylan into xylose or xylooligosaccharides. They may be achieved by acid hydrolysis or through the use of xylanolytic enzymes [2].

1.2 The role of thermostable enzymes

As mention earlier, enzymes are useful in various industrial applications. One characteristic that is required in industrial processes is thermostability. Several substrates can only be degraded or converted by enzymes under extreme conditions. The increase in temperature has a significant influence on the solubility of organic compounds. Thermostable enzymes are required to degrade lignocellulosic materials because the substrate must be pre-treated at 160–260 °C before it is vulnerable to enzymatic digestion the biofuel production such as ethanol. The goal of pretreatment is to make the lignocellulosic materials accessible to hydrolysis for conversion to fuels. The pretreatment process will change the physical and chemical structure of the lignocellulosic biomass and improve hydrolysis rates [26]. Therefore, thermostable enzymes have gained wide industrial and biotechnological interests because they are suitable for harsh industrial processes. In addition, high temperature helps eliminate the risk of bacterial or viral contamination. For these reasons, the number of studies on extremophiles from extreme environments has grown rapidly in the past few years [7].

1.3 Direct DNA isolation from environment

Soil is a major source of organic carbons on earth and an important habitant for microorganisms. Moreover, soil microorganisms are the main sources for natural products e.g. antibiotics, anticancer drugs, antifungal compounds, and enzymes. A great diversity is found in soil microorganisms as revealed by experiments using 16S RNA sequence analysis, in situ hybridization or dot blot hybridization. Therefore, soli microbial diversity is rich and remains widely unexplored resource for novel industrial enzymes and bioactive compounds. For many years, the diversity of soli microorganisms has been established based on culture-dependent approach. However, this approach has a limitation in that approximately only 1%of soli microorganisms can be cultured using the standard methods. Thus, the remaining 99% of soli microorganisms have yet to be identified. Therefore, culture-independent approaches have been developed to overcome such limitation [24].

Thailand, being a tropical country, has been known for great biodiversity of microorganisms. One of the potential bioresources for obtaining novel enzymes in Thailand

is Phu Khieo Bio-Energy at Chaiyaphum province. Recently, the biodiversity of one of agricultural industry has been established. Main target is lignocellulolytic degrading enzymes. Therefore, the microorganisms that can grow by using plant biomass as a substrate are focused in which it should provide a concentrated pool of lignocellulolytic degrading microorganisms [5].

1.4 Culture - independent approaches (activity-based screening and sequence - based screening approaches)

In order to obtain the novel enzymes directly from soil without the cultivation of microbes, activity-based screening and sequence-based screenings are commonly used to aid this purpose.

Activity-based screening has the potential to detect entirely novel genes encoding new types and classes of enzymes or to identify new bioactive compounds. To isolate the novel enzymes from environmental DNA, total soil microbial genomes (metagenome) are extracted to construct metagenomic library and enzymatic reactions with specific substrates can then be screened. Many enzymes of industrial importance have been discovered using this strategy. The advantage of this method is the potential of accessing totally unknown sequence and selection for full-length genes with functional gene products. The limitations, however, are that this method requires the expression of the functional protein of interests in the appropriate host cell (e.g. *Escherichia coli*). Therefore, more than one type of hosts might be needed in order to obtain functional protein. The large size of the library is also required. Furthermore, efficient and economical screening methods for the desired traits must be established to facilitate the screening of vast libraries.

Sequence-based screening approach required oligonucleotide primers to identify target genes directly by polymerase chain reaction (PCR). Therefore, the conserved amino acid sequences must be known to design degenerate primers to match unknown target genes followed by genome walking PCR to retrieve the full-length gene. One advantage of the sequence-based screening approach is that this method is not dependent on the heterologous expression of cloned genes. Using this approach, it is possible to obtain target enzymes with low level of expression which could be undetected when activity-based screening was employed. Sequence-based screening approach has been used successfully in discovering various enzymes such as xylanases, amylases and polyketide synthases.

CHAPTER II
OBJECTIVES

The aim of this project is to isolate the full-length gene encoding xylanases discovered by sequence-based approach of bagasse compost soil.

CHAPTER III
MATRERIALS

3.1 Bacterial strains

Escherichia coli, DH5α was used as a host for plasmid propagation

3.2 Plasmid vectors

pTZ 57 R/T vector (InsTAclone™ PCR Cloning Kit) was used for all cloning steps. A physical map of pTZ 57 R/T vector is shown in Figure 2.

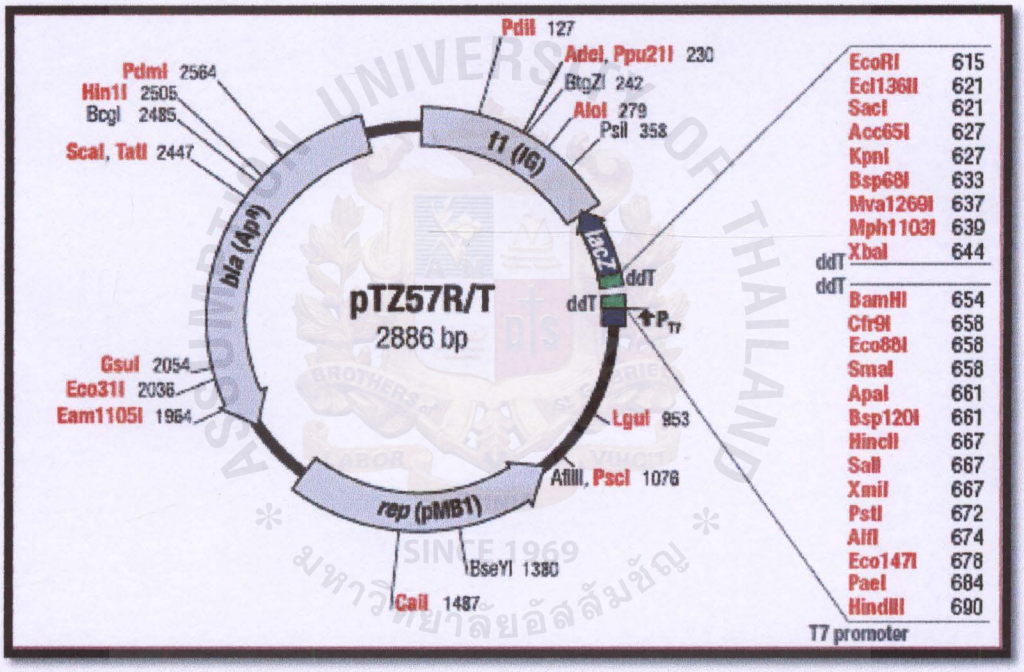

Figure 2: The physical map of pTZ57R/T (Fermentas)

The vector has been pre-cleaved with *Eco*32I (an isoschizomer of *Eco*RV) and treated with terminal deoxynucleotidyl transferase to create 3'-ddT overhangs at both ends. The PCR fragment with 3'-dA overhangs is ligated into the vector with 3'-ddT using DNA ligase.

3.3 Sampling site

Soil sample was obtained from Phu Khieo Bio-Energy at Chaiyaphum province in Thailand. The Samples at 1.2-1.5 m deep were collected and use for a source of lignocellulolytic degrading microorganisms.

3.4 Synthetic oligonucleotides

All synthetic oligonucleotides (primers) used in this study (Table 2) were synthesized by Bioscience Unit, BIOTEC (Thailand).

Table 2: Primers used in each experiment. For the degenerate oligonucleotides, the following addreviations are used (I=Inosine; M=A, C; N=A, T, C, G, R= A, G; S= C. G; W=A, T; Y= C, T)

Experiments and Primer names	**Sequence (5'-3')**	**Size (bp)**	**Tm (˚C)**
Adaptor Primer (API)	GTAATACGACTCACTATAGGGC	22	59
Nested Adaptor Primer (AP2)	ACTATAGGGCACGCGTGGT	19	71
Family 10 xylanase gene			
Partial xylanase gene			
XylF23	MGNGGICAYACNYTIGTTTGGCA	23	55
XYNFR	T(AC)GTT(GT)AC(AC)AC(AG)TCCCA	17	44
3'-end DNA amplification			
GSPF1	CTTACCCGTGTCCAGGGATGTTTTGATCC	29	68.1
GSPF2	TATATATCCACGGTTGTGGGAAGGTACAAAGG	32	66.9
GSPF3	CTCAAAACCGGCAAACTGGTAGGTT TCGATA	31	70.9
GSPF4	GAATCATGGAACCGAGTCCACCAG	24	69.5
5'-end DNA amplification			
GSPR1	AAAACATCCCTGGACACGGGTAAGGCAGG	29	68.1
GSPR2	GGCAGGGTTGGCCGGATCTTTAAAAATCC	29	69.5
Full-length DNA amplification			
Xyn F	TCATTTCAGTCGTAACACACCAAATACAGC	30	64
Xyn R	ATGAAACGGATTCAATGGATATCAATGCTCG	31	64.2

3.5 Culture media

3.5.1 Bacterial culture medium (for transformation containing pTZ 57 R/T (vector)

E. coli strain DH5α was grown in LB [1% (w/v) peptone (Difco), 0.5% (w/v) NaCl, and 0.5% (w/v) yeast extract (Difco)]. The *E. coli* transformants were grown in LB containing 100 μg/ml ampicillin (Sigma, USA). For agar plates, 2% (w/v) of Bacteriology agar was added. For X-gal/IPTG agar plates, 40 μl of a stock solution 5- Bromo-4-chloro-3-indodyl-β-D-galaactoside (X-gal; 20 mg/ml of X-gal in dimethylformanmide) and 4 μl of a solution of isopropylthio-β-D-galactoside (IPTG; 200 mg/ml of IPTG in water) were added in agar media.

3.6 Kits

InsTAclone™ PCR Cloning Kit (pTZ57R/T)	Fermentas, USA
GeneJET™ Gel Extraction Kit	Fermentas, USA
GeneJET™ Plasmid miniprep kit	Fermentas, USA
BD GenomeWalker™ Universal Kit	Clontech, USA

3.7 Enzymes

All enzymes were purchased for New England Biolabs, USA ; Boehringer Mannheim, Germany; Promega, USA; Gibco BRL, USA; Stratagene, USA; Sigma, USA; Fermentas, USA or Invitrogen, USA.

CHAPTER IV
METHODS

4.1 Genomic DNA extraction with SDS-based DNA extraction method (Zhou et al., 1996)

The total 5g of soil sample (Bagasse compost soil) from Phu Khieo Bio-Energy at Chaiyaphum province in Thailand was mixed with 13.5 ml of DNA extraction buffer (100 mM Tris-HCl [pH 8.0], 100 mM sodium EDTA [pH 8.0], 100 mM sodium phosphate [pH 8.0], 1.5 M NaCl, 1% (w/v) cetyltrimethylammonium bromide (CTAB)) and 100 µl of proteinase K (10 mg/ml) by horizontal shaking at 225 rpm for 30 min at 37℃. Next, 1.5 ml of 20% SDS was added, and the sample was incubated in a 65℃ water bath for 2 h with gentle end-over-end inversions every 15 to 20 min. The supernatant was collected after centrifugation at 6,000 x g for 10 min at room temperature and transferred to 50-ml centrifuge tubes. The soil pellet was extracted two more times by adding 4.5 ml of the extraction buffer and 0.5 ml of 20% SDS, vortexing for 10 sec, incubating at 65 ° C for 10 min, and centrifuged as before. Supermatants from the three cycles of extraction were combined and mixed with an equal volume of chloroform-isoamyl alcohol (24:1, v/v). The aqueous phase was recovered by centrifugation and precipitated with 0.6 volume of isopropanol at room temperature for 1 h. The pellet of crude nucleic acids was obtained by centrifugation at 16,000 x g for 20 min at room temperature, washed with cold 70% ethanol, and resuspended in sterile distilled water. The DNA was analyzed on 0.8% agarose gel electrophoresis.

4.2 Genomic DNA purification using troughing method

In order to get rid of humic acid contamination within bagasse compost sediment, the obtained genomic DNA will be subjected to further purification using the modified troughing method. This method relies on the difference in mobility efficiency of the DNA and humic acid compounds during electrophoresis in the agarose gel. Briefly, the genomic DNA is subjected to electrophoresis in a TAE containing ethidium bromide. The extent of a migration of the genomic DNA is visualized under the UV light and a rectangular well (trough) is made just below the DNA band directly in front of the path of migration. The well is then filled with troughing buffer which consists of 30% PEG8000 (Polyethyleneglycol, molecular weight 8000) in TAE buffer. Electrophoresis is continued until the DNA band migrated to the middle of the well. The DNA, which is free from humic acid is then collected [5].

4.3 DNA concentration and purity determination

The quantity and quality of DNA were determined by using a Nanodrop 1000 spectrophotometer (Thermo scientific, USA) at the absorbance of 260 nm and 280 nm. The DNA purity was determined by the absorbance ratio of A260/ A280. A ratio should be between 1.6 - 1.8 which represents a high purity of DNA.

4.4 DNA electrophoresis

The agarose gel was prepared by melting 0.8% (w/v) of agarose gel in lX TAE Tris-Acetate-EDTA (Tris acetate buffer 89 mM, acetic acid, 89 mM Tris-HCl, 2 mM EDTA, pH 8.0). The Gel wad poured into an electrophoretic tray and allowed to set at room temperature for 20-30 min. TAE buffer 1X was used as an eleteophoretic buffer. The digested DNA or PCR product was mixed with 30% (v/v) of loading dye (25% glycerol, 60 mM EDTA and 0.25% bromphenol blue), loaded into the prepared gel and run at a constant voltage at 110 volts of 1h. The gel was then stained in 2.5 g/ml of ethidium bromide solution for 5 min and destained in water for 15-20 min. The DNA was visualized under UV light [Gel Doc model 100 (Bio-Rad, USA) or BioDoc-It™ System (UVP, USA)] and photographed.

4.5 DNA purification by extraction kit (GeneJET™ Gel Extraction Kit)

A band containing crude extracted DNA was cut from the gel under a UV light. GeneJET™ Gel Extraction Kit was used to purify the PCR product. Specifically, binding buffer in the ratio of 1:1 was added to the excised gel containing the PCR product. The mixture was incubated at 50 - 60℃ for 10 min or until the gel was completely dissolved. The tube was mixed by inversion every few minutes to facilitate the melting process. Ensure that the gel is completely dissolved. Check the color of the solution. A yellow color indicates an optimal pH for DNA binding. If the color of the solution is orange or violet, add 10 µl of 3 M sodium acetate, pH 5.2 solution and mix. The color of the mix will become yellow. The mixture was then applied to the Gene Jet purification column and centrifuged at 12,000 x g for 1 min. After the flow through was discarded, 700 µl of Wash buffer was added to GeneJET™ purification column to remove all traces of agarose. The column was then centrifuged as described previously. The empty GeneJET™ purification column was centrifuged for an additional 1 min to completely remove residual wash buffer. The column was placed in a fresh 1.5 ml tube. A volume of 50 µl of elution buffer was added to the GeneJET™ purification column and the column was left at room temperature for 1 min to dissolve DNA. The tube was then centrifuged as above. The step of DNA elution with the

elution buffer was repeated to obtain the maximal amount of DNA. Then, 1-2 µl of concentrated DNA was analyzed by gel electrophoresis in order to determine its concentration.

4.6 DNA amplification of xylanase

4.6.1 Degenerate primers design

The degenerated oligonucleotide primers were designed based on the conserved regions of family 10 xylanase genes for xylanase.

4.6.2 PCR amplification

4.6.2.1 PCR amplification of partial xylanase gene

To amplify the partial sequence of xylanase gene from Phu Khieo Bio-Energy at Chaiyaphum province (Bagasse compost soil) sediment, PCR was employed using XylF20 and XYNFR primers (Table 3). PCR amplification was performed in 50-µl reaction mixtures containing 2 µl of purified DNA, 1 µM of each primer, 1X Tap DNA polymerase buffer [50 mM KCl, 10 mM Tris-HCl (pH 8) and 0.1% Triton® X-100], 1.5 mM $MgCl_2$ 0.5 mM of dNTPs and 1 units of Tap DNA polymerase. The reaction was performed with a Perkin-Elmer model 2400 GeneAmp apparatus for 35 cycles of denaturation at 94 °C for 30 sec, annealing at 46 °C for 1 min and extension at 72 °C for 1 min. PCR product was analyzed on 1.5% agarose gel electrophoresis, Then visualized under UV light and photographed.

4.7 Cloning of xylanase genes

4.7.1 Competent cell preparation using DMSO

A single colony of *E. coli* DH5α was inoculated into 250 ml SOB [2% peptone (Difco), 0.5% yeast extract (Dirfco), 0.05% Nacl] broth and aerated at 250 rpm 18 °C until OD_{600} reached 0.6. The cell was transferred into five sterile 50 ml polyprorylene centrifuge tubes and chilled on ice for 10 min. After centrifugation at 3,000 xg for 7 min at 4 °C, the call pellet was resupended in 20 ml of ice-cold TB buffer [10 mM PIPES, 55 mM $MnCl_2$,15 mM $CaCl_2$, and 250 mM KCl], and incubated on ice for 10 min. The cell pellets were then resuspended in 20 ml of ice-colded TB buffer and dimethy sulfoxide (DMSO) was added with gently swirling to give the final concentration of 7% (v/v). The cell suspension was finally dispensed into 100 µl aliquots in microcentrifuge tubes and stored at -80 °C.

4.7.2 DNA ligation

The purified DNA was ligated into the pTZ57R/T Easy vector in a molar ratio 1:3 (vector:insert). The appropriate amount of DNA used in ligation was calculated from the following formula:

$$ng\ of\ insert = \frac{ng\ of\ vector \times Kb\ size\ of\ insert \times (insert{:}vector\ molar\ ratio}{Kb\ size\ of\ vector}$$

In this study, 25 ng of pTZ57R/T Easy vectoe was used in a total volume of 10 µl. Ligation mixture contained 1X rapid ligation buffer of T4 DNA ligase [30 mM Tris-HCl (pH7.8), 10 mM $MgCl_2$, 10 mM EDTA,1 mM ATP and 5% polyethyleneglycol] and 3 units of T4 DNA ligase. The ligation mixture was mixed and incubated for overnight at 16 °C.

4.7.3 Transformation of competent *E. coli* DH5α cells

The total of 10 ng of ligated product was mixed with 100 µl of competent cells. The mixture was placed on ice for 30 min, immediately heat shocked at 42°C for 90 sec and was placed back on ice for 5 min. A volume of 900 µl of LB [2% peptone (Difco), 0.5% yeast extract (Difco), 0.05% NaCl] medium was added to the mixture which was then incubated at 37°C for 1 h with constant shaking. The cells were spreaded on LB agar plate containing 100 g/ml of ampicillin, 4 µl of 0.8 M IPTG (isopropylthio-β-D-galactoside) (IPTG and X-gal were spreaded over the surface of an LB plate). The agar plate was incubated at 37°C for 12-16 h.

4.7.4 Master plate preparation and recombinant clone screening

After incubation for 12-16 h, blue white colonies were observed. Approximately 20 white colonies were picked and spotted on a LB agar plate containing ampicillin, so called master plate. The master plate was incubated at 37 °C for 12-16 h. To screen for the presence of DNA inserts from recombinant clones, simplified rapid size screening was performed. One colony per clone was picked by toothpick from the master plate and lysed in 30 µl of pre-warm lysis buffer [5 mM EDTA, 10% (w/v) SDS. 100 Mm NaOH, 60 Mm KCI and 0.05% (w/v) bromphenol blue]. The lysed cells were incubated at 37°C for 5 min, placed on ice for 5 min and centrifuged at 13,000 xg for 5 min. Then, 20 µl of supernatant was analyzed on 0.8% agarose gel electrophoresis. Clones that contained the plasmid DNA with the larger size then that of the vector alone were selected for plasmid extraction.

4.8 Plasmid extraction

4.8.1 Extraction of plasmid DNA using GeneJET™ Plasmid Miniprep Kit (Fermantas)

The overnight cultured cells were collected by centrifugation at 12,000 x g for 1 min, and the supernatant was discarded. The cell suspension was transferred to a microcentrifuge tube. The pellet was resuspended in 250 µl of resuspenstion solution. The bacteria should be resuspended completely by vortexing or pipetting up and down until no cell clumps remain. A 250 µl of lysis buffer was added and the mixture was gently inverted for 4-6 times until the solution becomes viscous and slightly clear. At this step, do not vortex to avoid shearing of chromosomal DNA and do not incubate for more than 5 min to avoid denaturation of supercoiled plasmid DNA. A volume of 350 µl of the the Neutralization Solution was then added and the mixture was further inverted for 4-6 times immediately and thoroughly. It is important to mix thoroughly and gently after the addition of the Neutralization Solution to avoid localized precipitation of bacterial cell debris. The neutralized bacterial lysate is cloudy and viscous. The solution was then centrifuged for 5 min at 12,000 x g to pellet cell debris and chromosomal DNA. After centrifugation, the supernatant was transferred to the GeneJET™ spin column which was inserted in a collection tube by decanting of pipetting. This step has to avoid disturbing or transferring the white precipitate. The colume was then centrifuged at 12,000 x g for 1 min and the flow-through was discarded. The column was then washed with 500 µl of wash solution and centrifuged at 12,000 x g for 1 min. After the flow-through was discarded, the GeneJET™ spin column was washed again with 500 µl of wash solution and centrifuged at 12,000 x g for 1 min. After the flow-through was discarded the column was centrifuged for an additional 1 min to remove residual wash solution. The GeneJET™ spin column was then placed in a clean 1.5 ml microcentrifuge tube and the DNA was eluted by adding 50 µl of Elution buffer (10 mM Tris-Cl, pH 8.5) to the center of GeneJET™ spin column membrane to elute the plasmid DNA. The DNA was left standing for 2 min at room temperature and collected by centrifugation at 12,000 x g for 2 min. The plasmid DNA was stored at -80°C.

4.9 Insert size screening by restriction enzyme digestion

After recombinant plasmid was purified, the presence of DNA insert was checked by *Eco*RI and *Bam*HI digestion. Plasmid (1µl) was adder into the reaction mixture containing 1X buffer [90 mM Tris-HC1 (pH 7.5), 50 mM NaC1 and 10mM $Mgc1_2$] and 2-4 units of *Eco*RI and *Bam*HI restriction endonuclease enzyme (Promega) in a final

volume of 20 µl. The reaction mixture was incubated at 37 °C for 3h. The size of insert was analyzed using appropriate percentage of agarose gel.

4.10 DNA sequencing

After screening with restriction enzyme analysis, the recombinant clones harboring inserted DNA were sent for DNA sequencing (BioDesign Co., Ltd.)

4.11 Sequencing analysis

DNA sequencing obtained was translated into deduced amino acid sequences using Bioedit sequence Alignment Editor program. The DNA sequences from different clones were aligned by Clustal X program. The homology of nucleotide and the deuced amino acid sequences were searched in Genbank database (http://www.ncbi.nim.nih.gov) and compared to other related sequences by Clustal X Program.

4.12 Amplification of 5' and 3' ends of xylanase

4.12.1 Genome walking PCR (Clontech. USA) (Figure 3)

The construction of DNA libraries was performed by digestion with different restriction enzymes (*Eco*RV, *Dra*I, *Stu*I, and *Pvu*II) that recognize a 6-base site and generate blunt-ended fragments. Digested genomic DNA was then ligated to the Genome Walker adaptor. The ligation mixture contained 1 µl of Genome Walker adaptor, 1X ligation buffer and 3 units of T4 DNA ligase (Fermentas) in the total of 8 µl. After the ligation mixture was incubated for overnight at 16°C, two rounds of PCR amplification were performed using different primer pairs for 5'-end amplification. The first PCR amplification was performed using the outer adaptor primer (AP1) and outer gene specific primers (GSP1): GSPR1 for xylanase. The first PCR mixture was then used as a template for nested PCR amplification using the nested adaptor primer (AP2) and nested gene-specific primers (GSP2): GSPR2 for xylanase. Both PCR reactions contained 1 µl of digested genomic DNA template, 1X *Tth* PCR reaction buffer, 3 mM $Mg(OAc)_2$, 0.2 mM dNTPs, 0.2 µM of each primer and 1 units advantage *Tth* polymerase mix (Clontech) in a final volume of 25 µl. The primary PCR was performed using two-step cycles and the condition for PCR was 7 cycles at 94 °C for 2 sec and 72 °C for 3 min, followed by 32 cycles at 91 °C for 2 sec and 67 °C for 3 min and an final extension at 67 °C for 4 min. Then the secondary PCR amplification was performed for 5 cycles at 94 °C for 2 sec and 72 °C for 3 min followed by 25 cycles at 94 °C for sec and 67 °C for 3 min. The PCR

product was analyzed on 1.5% agarose gel electrophoresis. Then they were purified, subcloned and sequenced as described in protocol 4.5, 4.7, 4.8, 4.9, 4.10 and 4.11.For 3'- end amplification, PCR reaction was performed as described above in except that the following primers were used: GSPF1, GSPF2, GSPF3, and GSPF4 (for xylanase).

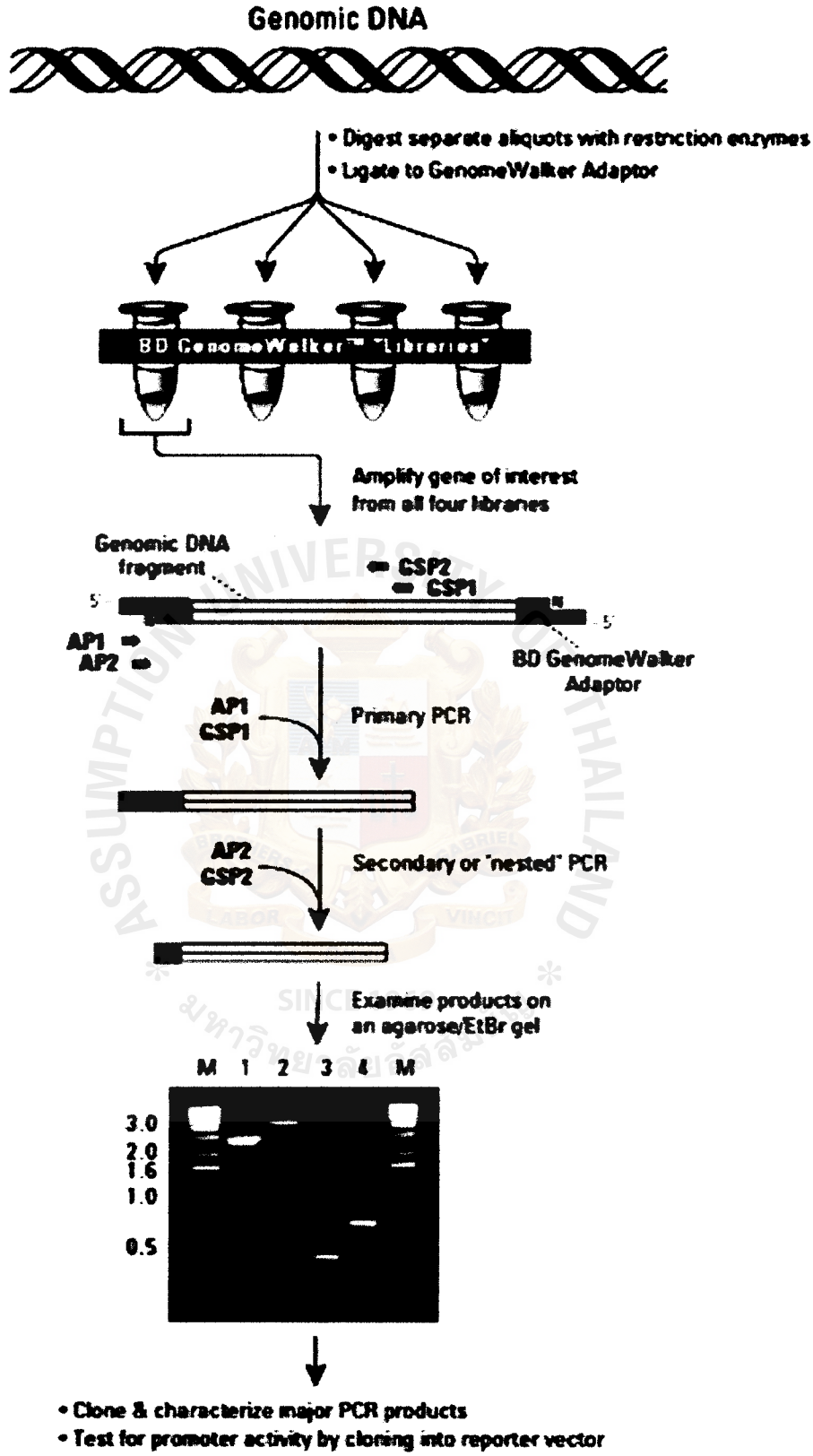

Figure 3: Schematic of genome walking PCR (Clontech)

4.13 Cloning and amplification of full - length

To amplify the full-length gene Xylanase gene, two gene specific primers and corresponding to the 5' and 3' ends (Table 3), respectively were designed according to the 5' and 3' sequences obtained from genome walking. The reaction consisted of 1 µl of genomic DNA template, 1X Prime STAR GC buffer (Mg^{2+} Plus), 2.5 mM dNTPs, 1 µM of each primer, 1 unit of Prime STAR HS DNA polymerase (Takara, Japan), and sterile milliQ water to a final volume of 50 µl. PCR was performed for 30 cycles of denaturation at 98 °C for 10 sec, annealing at 58 ℃ for 15 sec and extension at 72 °C for 3 min, followed by a final extension at 72 °C for 10 min. The PCR product was then analyzed on 0.8% agarose gel electrophoresis, purified, subcloned and sequenced as described in protocol 4.5, 4.7, 4.8, 4.9, 4.10 and 4.11.

CHAPTER V
RESULTS AND DISCUSSION

5.1 Sample collection

The main target of this research is to screen the novel xylanase from the baggase. Based on the hypothesis, the compost soil under sugarcane bagasse pile was naturally enriched by lignocellulosic biomass. This will provide a concentrated pool of lignocellulolytic degrading microorganisms.

In this research, this compost soil was selected as a source of for a source of lignocellulolytic degrading microorganisms. The samples were collected from power plant, Phu Khiao Bio-Energy at Chaiyaphum province (Figure 4). To collect the soil sample, bagasse was removed in order to collect compost soil under bagasse. The samples at 1.2-1.5 meters deep were collected and use for a source of lignocellulolytic degrading microorganisms.

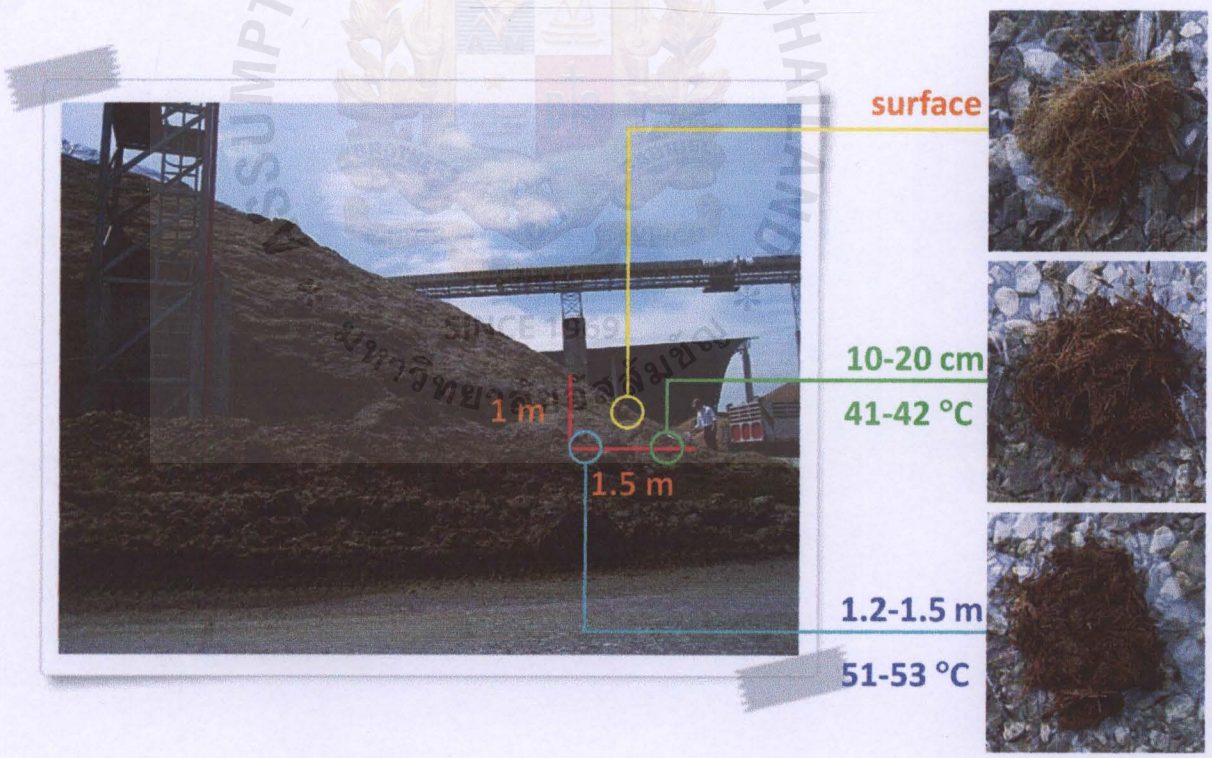

Figure 4: Compost soils under sugarcane bagasse piles at 1.2-1.5 m deep

5.2 Direct extraction and purification of genomic DNA from the sediment of Phu Khieo Bio - Energy at Chaiyaphum province (Bagasse compost soil)

Genomic DNA from the sediment of Phu Khieo Bio-Energy at Chaiyaphum province (Bagasse compost soil) was extracted using SDS based DNA extraction method. The yield of crude extracted DNA was 0.044 μg of DNA per gram of soil. According to Zhou et al, 2.5 to 26.9 μg of DNA per gram of soil could be obtained from this method, depending on the sources of the soil. The size of genomic DNA was larger than 1 Kb (Figure 5), which was in agreement with those from previous studies. To avoid the contamination from humic substances that might interfere in some downstream reactions such as PCR and restriction enzyme digestion, DNA purification using modified troughing method was performed in order to get rid of the humic acid contamination (Figure6). This was to investigate the quality of DNA is sufficiently free of humic acid for downstream experiments. This suggested that DNA was of high quality for further experiments.

Figure 5: Genomic DNA extraction using SDS based DNA extraction method. Genomic DNA from the sediment of Phu Khieo Bio-Energy at Chaiyaphum province was analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. Lane 1 shown 1 Kb DNA marker. Lane 2 and 3 shown the crude extracted DNA.

Figure 6: The purified genomic DNA using troughing method. Crude extracted DNA was purified by using modified troughing method. It was analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. Lane 1 represented 1 Kb DNA marker. Lane 2 represented the purified DNA using troughing method.

5.3 PCR amplification using xylanase degenerated primers

To directly obtain partial xylanase gene from unculturable microorganisms, degenerate primers (XylF23 and XYNFR primers) were designed based on the conserved amino acid sequences of glycosyl hydrolase family 10 xylanase. PCR product of approximately 160 bp was obtained (Figure 7) and purified using GeneJET™ Gel Extraction Kit (Fermentas). The purified fragment was then ligated to pTZ 57 R/T vector and transformed into *E. coli* DH5α. The transformants were selected on LB medium supplemented with ampicillin and recombinant plasmids were extracted. Restriction enzyme analysis with *Eco*RI and *Bam*HI digestion was performed and the result showed that each clone harbored various sizes of DNA inserts (Figure 8). The randomly selected recombinants were submitted for DNA sequencing. From DNA sequencing analysis, 7 different partial sequences encoding xylanases were found. These recombinant plasmids contained different amount of nucleotides which exhibited 58-98% identities. They were 58%, 64%, 98%

identity to Endo-1,4-beta-xylanase from *Butyrivibrio proteoclasticus B316*, *Geobacillus thermoglucosidasius C56-YS93*, and *Clostridium thermocellum* respectively.

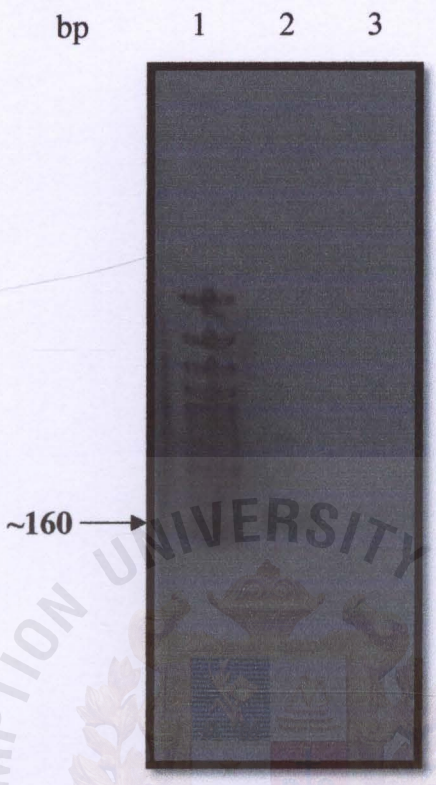

Figure 7: PCR amplification of partial xylanase gene. PCR amplification of partial xylanase gene using XylF23 and XYNFR primers. The size of expected PCR product was approximately 160 bp. PCR product were analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. Lane 1 shown 100 Kb DNA marker. Lane 2 shown the amplified PCR product of XylF23 and XYNFR primers. Lane 3 shown Negative control.

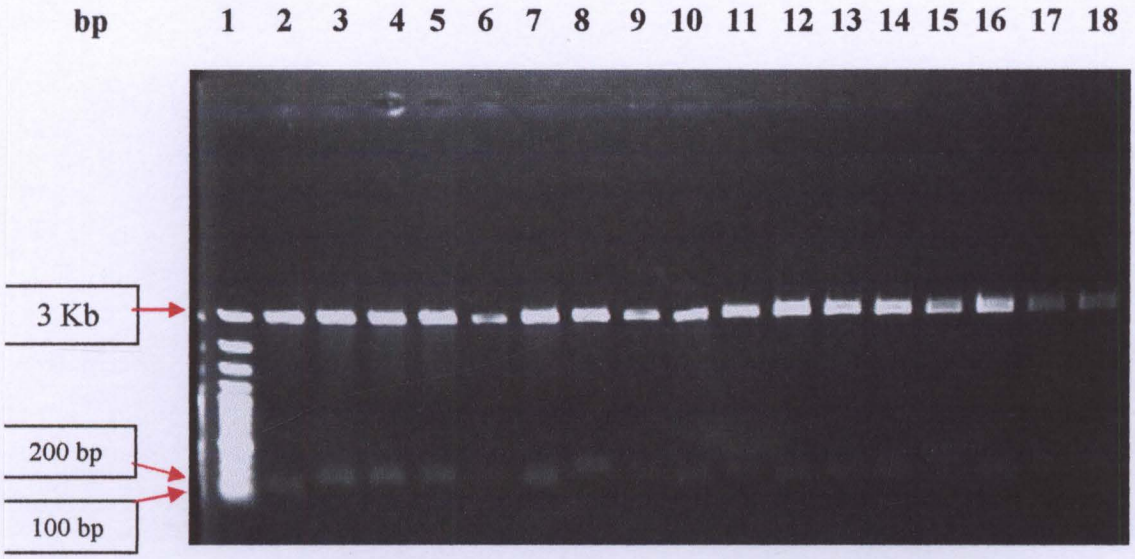

Figure 8: Restriction enzyme analysis with *Eco*RI and *Bam*HI digestion. Recombinant clones were digested with *Eco*RI and *Bam*HI restriction enzyme. The digested products were analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. Lane 1 represented 100 Kb DNA marker. Lane 2 to 18 represented the amplified PCR product of XylF23 and XYNFR primers.

Accordingly, the recombinant plasmids contained 133 nucleotides which 64% identities to Endo-1,4-beta-xylanase from *Geobacillus thermoglucosidasius C56-YS93* (figure9) was used for genome walking approach (Clontech) to obtain the full- length sequence of this xylanase gene.

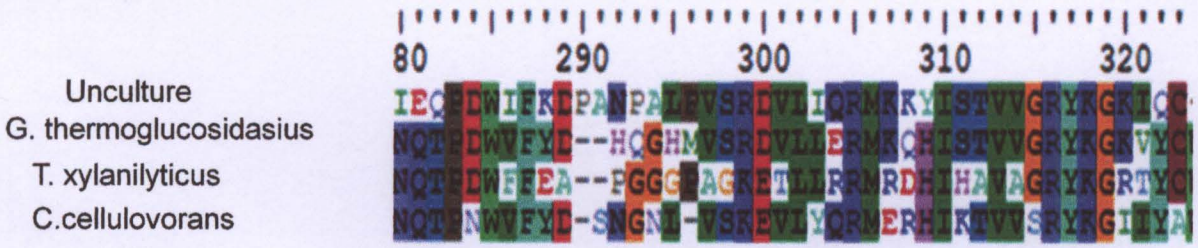

Figure 9: Alignment of the deduced amino acid sequences of partial xylanase containing 133 bp. The deduced amino acid sequences of partial xylanase was aligned by Clustal X with other bacterial xylanase genes. This sequence was homolog to xylanase from *Geobacillus thermoglucosidasius*m, *Thermoanaerobacterium xylanolyticum*, and *Clostridium cellulovorans*.

The reason why this clone was selected for obtaining the full- length sequence of this xylanase gene because there was the recently discovered extremophile – the bacterium *Geobacillus thermoglucosidasius* – which thrives in the high temperatures and pressures of petroleum reservoirs. It has ability to ferment the major C5 and C6 sugars (e.g., xylose and glucose) in lignocellulosic biomass and can tolerate high concentrations of ethanol [19]. These capabilities make *Geobacillus thermoglucosidasius* an ideal microbe for improved production of bio-ethanol and other more advanced biofuels. Based on the recently research, the solar energy stored in plant biomass is embedded within the complex sugars of lignocellulose, a matrix of sugars and lignin (the substance that gives strength and structure to plant cell walls). As lignocellulose is the most abundant organic material on earth, this stored energy represents a potential bonanza for biofuel production provided effective fermentation microbes can be found. Yeast, the most commonly used fermentation microbe, is unable to naturally ferment many of sugars in lignocellulose. Preliminary studies suggested *Geobacillus thermoglucosidasius*m as a candidate but measuring its full potential for biofuels production required detailed information on its metabolism [19]. Therefore, this recombinant plasmid was used as the starting point for genome walking approach in order to find the open reading frame of this xylanase.

5.4 Amplification of the 3' end of partial xylanase gene using genome walking

In order to obtain the full length xylanase gene, the set of specific primers were designed (Table 3) from the partial xylanase sequence which was obtained in the beginning (5.3). A set of specific primers will be designed based on forward or 3'- end and perform the genome walking until obtaining the STOP codon. On the other hand, the reverse side or 5' - end will also perform the genome walking until obtaining the START codon. This can be indicated that the open reading frame of the novel xylanase enzyme will be discovered.

To obtain the full-length xylanase gene, a set of gene specific-primers was designed (Table 3) to retrieve the 3' and 5'-ends of xylanase gene directly from the environmental DNA using genome walking PCR (Clontech). To retrieve the 3' – end of xylanase gene, PCR amplification was performed with two rounds of PCR amplification to increase specificity (nested PCR). The construction of four DNA libraries with different restriction enzymes (*Eco*RV, *Dra*I, *Stu*I, and *pvu*II) which were ligated to the Genome Walker adaptor were amplified using a set of specific primers which were provided in Table3.

For the PCR amplification, the first or primary PCR uses the outer adaptor primer (AP1) which provided in the kit and an outer gene specific primer (GSPF1) which designed based on the known sequence of partial xylanase. The primary PCR mixture was then diluted and used as a template for a secondary or "nested" PCR with the nested adaptor primer (AP2) and a nested gene-specific primer (GSPF2). The nested adaptor primer (AP2) was provided in the kit and a nested gene-specific primer (GSPF2) was obtained based on another known sequence of partial xylanase. The direction of a set of gene specific primers designed from the known sequence partial xylanase was shown in the figure 10.

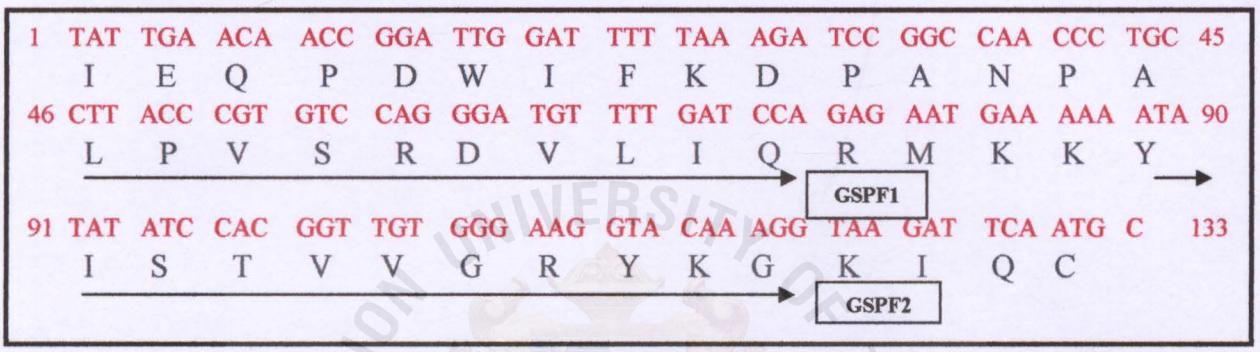

Figure 10: Nucleotides and deduced amino acid sequences of partial xylanase from the sediment of Phu Khieo Bio-Energy at Chaiyaphum province (bagasse compost soil). DNA sequencing result revealed 133 nucleotides of the partial xylanse gene. A set of gene specific primers 3' end (GSPF1, and GSPF2 primers) were designed and indicated by arrows.

In the first and second PCR, the amplified PCR products were obtained (Figure 11). The PCR products from every library both primary and secondary were analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. From the result, there were PCR product from the primary PCR which came from *Dra*I, *Pvu* II, and *Eco*RV. On the other hand, the PCR products obtained from secondary PCR came from all four libraries which were *Dra*I, *Stu*I, *Pvu* II, and *Eco*RV.

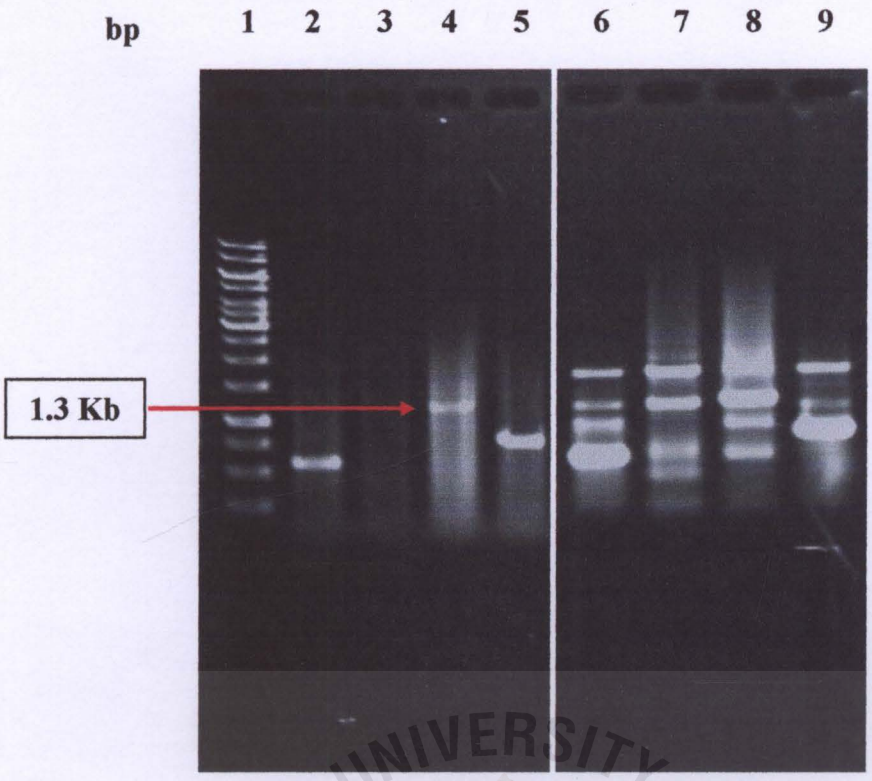

Figure 11: PCR amplification of xylanase for downstream walking at 3'end (Forward region). PCR amplification of downstream region (3'end) using a set of gene specific primers (GSPF1, and GSPF2 primers). The PCR products were analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. Lane 1 shown 1 Kb DNA marker. Lane 2 to 5 shown PCR product from primary PCR of *Dra* I, *Stu* I, *Pvu* II, *EcoR* V respectively. Lane 6 to 9 shown PCR product from nested PCR of *Dra* I, *Stu* I, *Pvu* II, *EcoR*V respectively.

Each of the DNA fragments with different sizes which began in a known sequence at the 3' -end of GSF1 and GSPF2 and extend into the unknown adjacent genomic DNA were then be purified and cloned into pTZ 57 R/T vector. Restriction enzyme digestion with *Eco*RI and *Bam*HI was performed to verify the DNA insert. The recombinant clones were subjected to DNA sequencing. DNA sequence analysis and the contig alignment found that obtianed PCR product of approximately 600 bp, 750 bp, and 1.8 kb from *Dra*I, *Eco*RV, and *Stu*I respectively shown no sequence similarity to family 10 xylanase gene. This can be determined that the PCR amplification produced the non specific products.

However, the obtained PCR product of approximately 1.3 kb from the *Pvu*II library revealed the length of 3'end sequence of 1,034 nucleotides, encoding 344 amino acids. This sequence identities to endo- 1,4 beta xylanase. Nucleotide sequence analysis showed that this sequence exhibited 43% identity to endo - 1,4 beta xylanase of *Dictyoglomus turgidum DSM 6724* (Figure 12).

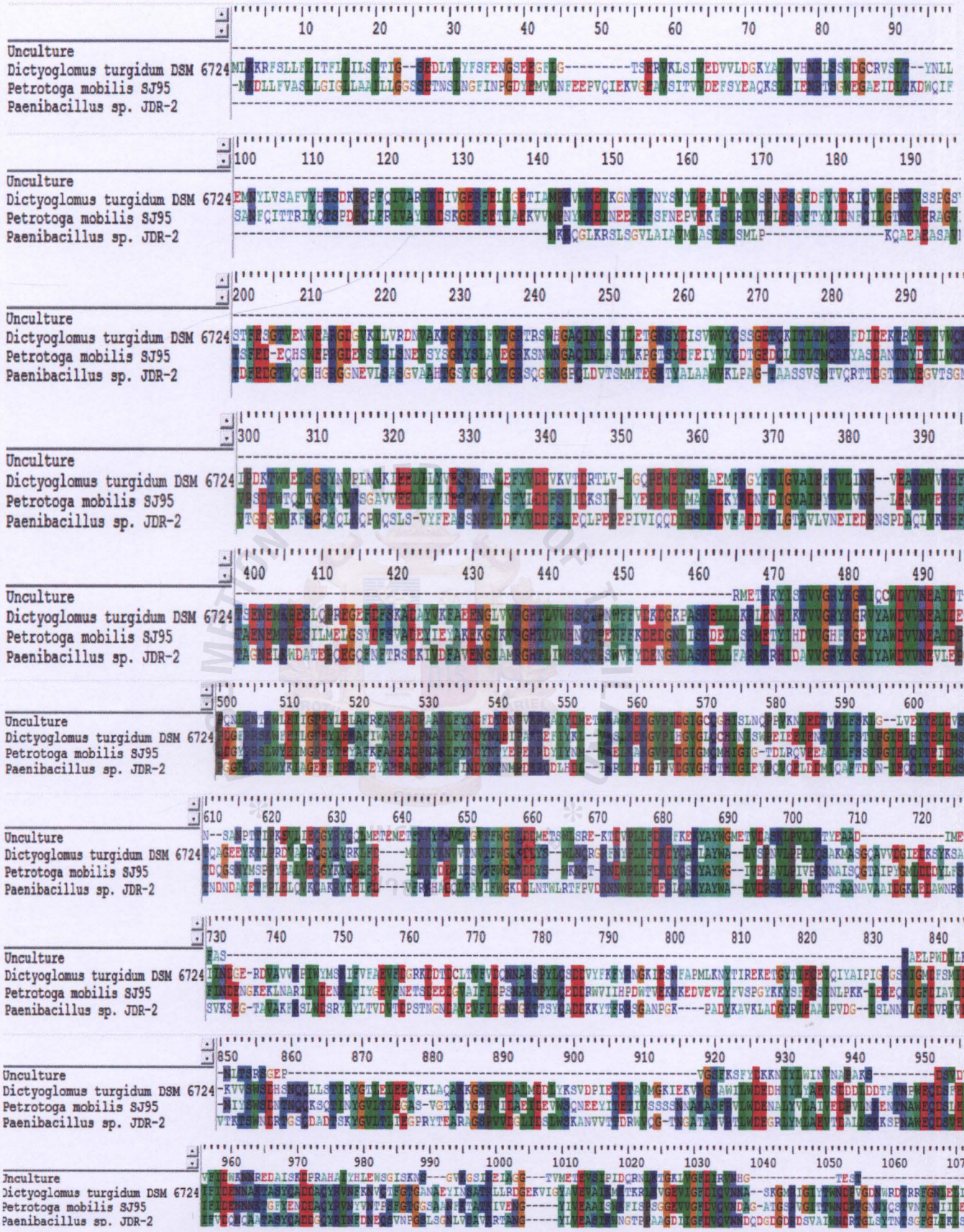

Figure 12: Amino acid alignment of *bagasse* xylanase gene against other bacterial xylanase genes from the first genome walking at 3'- end. The deduced amino acid sequence of

xylanase gene was aligned by Clustal X with other bacterial enzymes in xylanase family: *Dictyoglomus turgidum DSM 6724*, *Petrotoga mobilis SJ9*, and *Paenibacillus sp. JDR-2*. Shading of the alignment represented the degree of conservation. The color shading showed the most conserved region among all sequences compared.

The contig alignment and sequence analysis shown that the STOP codon was not found, therefore; the second genome walking was performed in order to obtain the STOP codon. Two set of primers were designed to obtain the full- length of this xylanase gene which were GSPF3 for primary PCR and GSPF4 for secondary PCR (Table3). The positions of a set of gene specific primers designed from the known sequence obtaining from first genome walking were shown in the figure 13.

753 TAT GGA TCA ATG TGA ATG CCC CTG CAA AAA GCG ACT CCG TGG ACG 797

798 TAT TTA TTG ATT GGA AAA ATA ACA GGG AGG ACG CTA TTT CAA AAG 842

343 ATC CGC GTG CGC ACG CCA TTT ATC ATT TGG AAT GGT CAG GAA TCA 887

888 GCA AGA ACA GTG GCG TCA AAG GTT CCA TCC GGG AAA TTG CCG GAG 932

933 GCA CCG TCA TGG AGG TAT CAA TTC CCA TTG ATC AGC GCA ATC TCA 977

GSPF3

978 AAA CCG GCA AAC TGG TAG GTT TCG ATA TCC GGG TGA ATC ATG GAA 1022

GSPF 4

1023 CCG AGT CCA CCA G 1034

Figure 13: Nucleotides and deduced amino acid sequences of partial xylanase from the sediment of Phu Khieo Bio-Energy at Chaiyaphum province (bagasse compost soil). The 3'-end of xylanase gene was obtained from unculturable microorganisms using genome walking method (Clonech). DNA sequencing revealed 1,095 nucleotides encoding 363 amino acids for the first genome walking. The positions of primers for second genome walking were indicated by arrows.

To obtain the remaining unknown sequence at 3'-end, a set of gene specific-primers was designed (Table 3). The PCR amplification was performed with two rounds of PCR amplification to increase specificity. The four DNA libraries were constructed by digesting with different restriction enzymes (*Eco*RV, *Dra*I, *Stu*I, and *pvu*II) and then ligated to the Genome Walker adaptor. All DNA libraries were amplified using a new set of specific primers.

For the second genome walking, the first or primary PCR uses the outer adaptor primer (AP1) which provided in the kit and an outer gene specific primer (GSPF3) which designed based on the known sequence of xylanase obtaining from the first genome walking. The primary PCR mixture was used as a template for a secondary or "nested" PCR with the nested adaptor primer (AP2) and a nested gene-specific primer (GSPF4). The nested adaptor primer (AP2) was provided in the kit and a nested gene-specific primer (GSPF4) was designed from the known sequence of xylanase obtaining from the first genome walking.

For the PCR amplification, the amplified PCR products were obtained in both two rounds (figure 14). The PCR products were analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. The results shown that there was only PCR product from *Dra*I library in the primary PCR. This single band was cloned for further DNA sequencing in pTZ 57 R/T vector. The insert size screening was done using restriction enzyme digestion with *Eco*RI and *Bam*HI to verify the DNA insert. The recombinant clones were subjected to DNA sequencing.

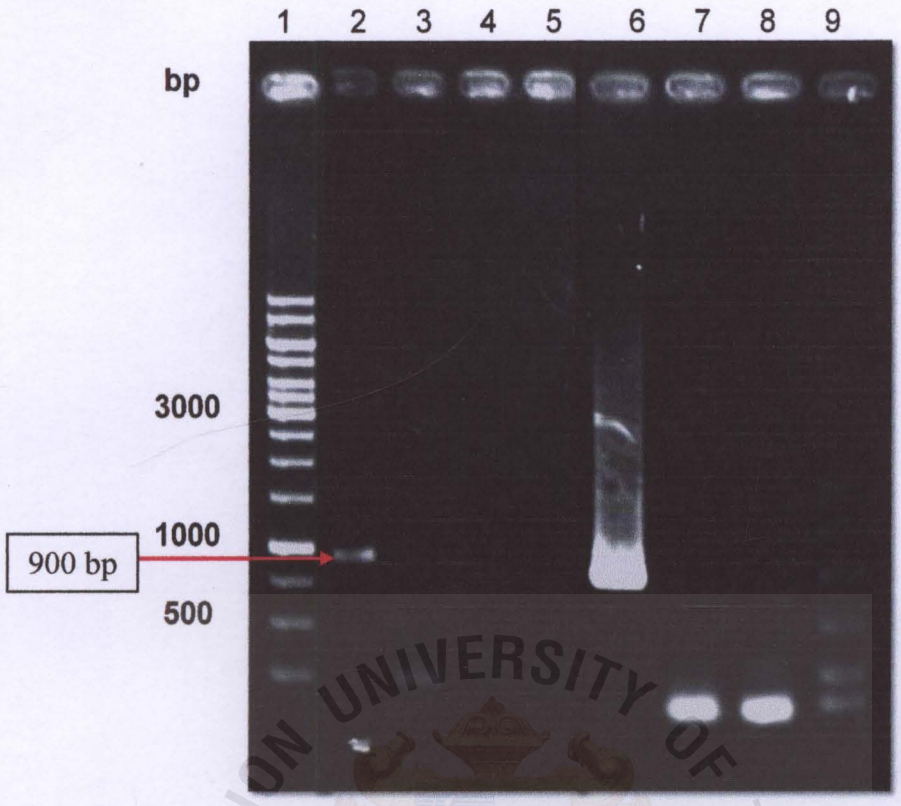

Figure 14: PCR amplification of xylanase for the remaining downstream walking at 3'end (Forward region). PCR amplification of downstream region (3'end) using a set of gene specific primers (GSPF3, and GSPF4 primers). The PCR products were analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. Lane 1 represented 1 Kb DNA marker. Lane 2 to 5 represented PCR product from primary PCR of *Dra* I, *Stu* I, *Pvu* II, *EcoR* V respectively. Lane 6 to 9 represented PCR product from nested PCR of *Dra* I, *Stu* I, *Pvu* II, *EcoR* V respectively.

The DNA sequence analysis and the contig alignment were analyzed that the obtained PCR product of approximately 900 bp from the *Dra*I library revealed the length of 3'end sequence of 634 nucleotides, encoding 211 amino acids. Nucleotide sequence analysis showed that this sequence exhibited 49 % identity to endo - 1,4 beta xylanase of *Thermotoga naphthophila RKU-10* (Figure 15).

The contig alignment and sequence analysis shown that the deduced amino acid sequence included a putative STOP codon in the second genome walking.

Therefore, the downstream region was successfully obtained since the STOP codon was found. The total of the 3'end sequence was 1,668 nucleotides encoding 556 amino acids which identity to endo - 1,4 beta xylanase (Figure 16 and 17).

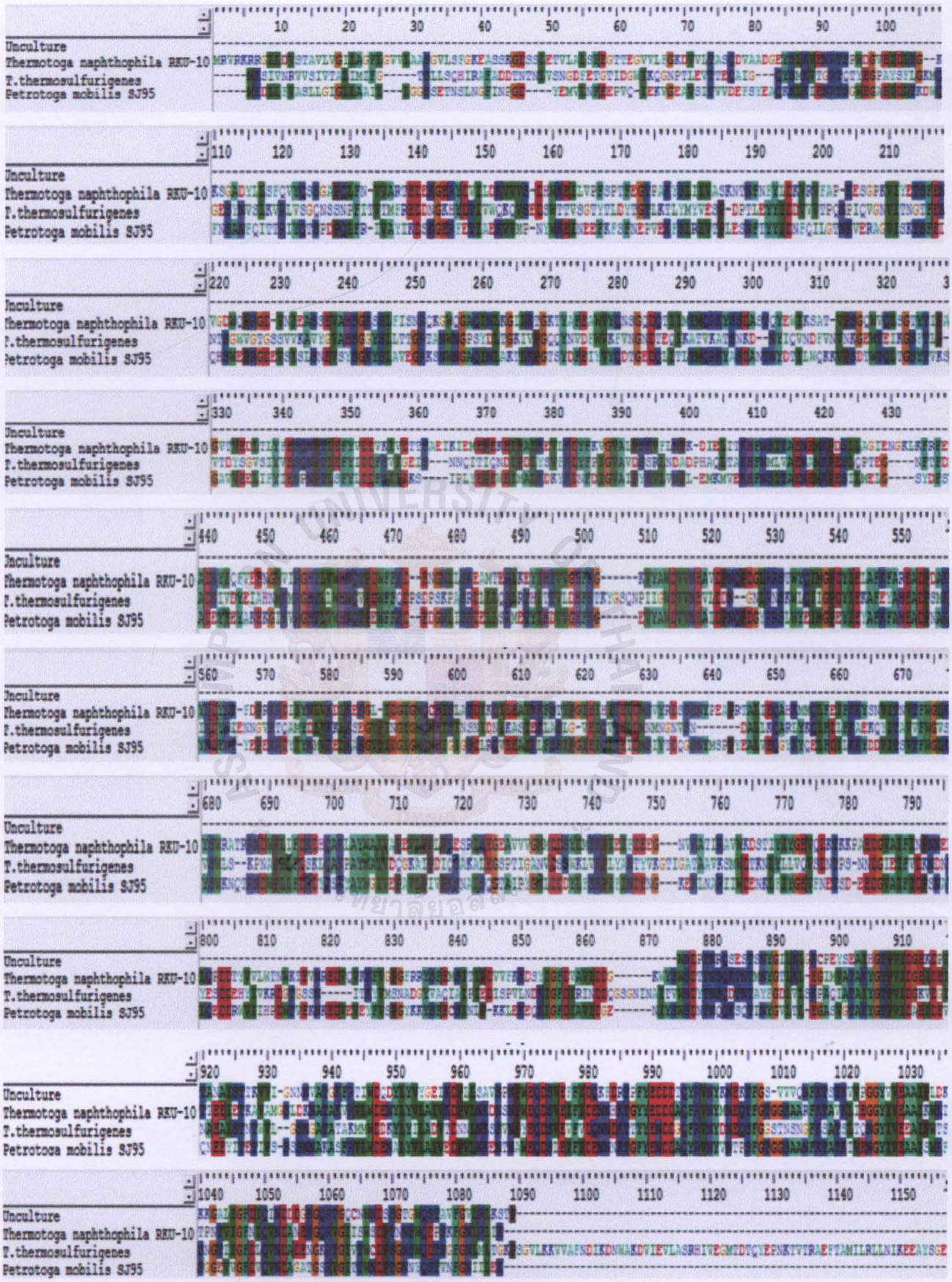

Figure 15: Amino acid alignment of *bagasse* xylanase gene against other bacterial xylanase genes from the first genome walking at 3'-end. The deduced amino acid sequence of xylanase

gene was aligned by Clustal X with other bacterial enzymes in xylanase family: *Thermotoga naphthophila RKU-10*, *Thermoanaerobacterium thermosulfurigenes*, and *Petrotoga mobilis SJ95*. Shading of the alignment represented the degree of conservation. The color shading showed the most conserved region among all sequences compared.

TGGGATGTTGTTAATGAGGCCATTGATACAACCCAACCTCAAAATTTACGAAATACAAAATGGCTTG
AGATCATCGGACCCGAATATTTGGAACTTGCTTTCAGGTTTGCCCATGAAGCCGACCCGGCGGCAAA
GCTTTTTTATAACGATTTCGATACCGAAAATCCGGTGAAACGGCAGGCCATATATGACATGGTTAAA
GCCCTTAAAGAAAAAGGCGTTCCCATTGACGGTATCGGTTGTCAGGGACATATATCGCTAAACCAAC
CTCCGGTTAAAAATATTGAAGATACTGTGAAACTCTTTTCCAAACTGGGTCTGGTTGAAATTACCGA
ACTCGATGTGTCCGTATTTAATAGCGCAAATCCGACAACGATACCCAAAGAAGTTTTAATTGAGCAG
GGTTACCGGTATCAGCAACTCATGGAAATGTTTAAAAAATATAAAAACGTTCTCACCGGTGTAACAT
TTTGGGGACTCAAGGATGATATGTCCTGGCTCTCAAGGGAAAAGACCGATGTGCCACTTTTGTTTGA
CAAGAGGTTCAAGGAAAAATACGCCTATTGGGGTATGGTTGATGCTTCGAAGCTCCCGGTTCTTATC
AAAACTTATGAAGCAGCGGATATCATGCCGGTTTTTGCCTCAAAAGCAGAACTTCCCTGGGATATCC
TTAAAACTGAAAACCTAACCAGCCGGTCCGGTGAACCGGTGGGAAGTTTTAAATCCTTCTATGACAA
AAAAAATATTTATTTATGGATCAATGTGAATGCCCCTGCAAAAAGCGACTCCGTGGACGTATTTATT
GATTGGAAAAATAACAGGGAGGACGCTATTTCAAAAGATCCGCGTGCGCACGCCATTTATCATTTG
GAATGGTCAGGAATCAGCAAGAACAGTGGCGTCAAAGGTTCCATCCGGGAAATTGCCGGAGGCAC
CGTCATGGAGGTATCAATTCCCATTGATCAGCGCAATCTCAAAACCGGCAAACTGGTAGGTTTCGAT
ATCCGGGTGAATCATGGAACCGAGTCCACCAGTTGGAATGATCCCACCAACCGCCAGTCGGAGAGC
ACTTCAAATTATGGAATCTTAAAACTTGGTAAATGTCCGGAATATTCGGAAGCTATCCATGGAACCC
CTGTTATCGATGGAGAAAAGGACCCAATCTGGGAAACGGCAAATGCTATTTCTACCACCAAGGTAG
TTATAGGTAATAATAACGTTGCCACTGGAAAATTCAGAACCATCTGGGATCAAGATTACCTGTATGT
TTATGGAGAAATAAAGGATCCGTTGCTTAGCGCAGTCAGCCCAAATCCCTGGGAACAGGATTCAGT
GGAATTCTTCATTGACCAAAAATTTGATAGAACTCCGTTTTACGAGGATGATGATACCCAATACCGG
GTGAATTACAAGAATGAAAAAACTTTTGGGAGTGTGGTTGTGCAAAATTTCAAAACCAGCACCAAG
GTAGTCCCCGGGGGATATGTTGTGGAAGCTGCGATTCCATTGGATAAAATACCAGTTAAGAAAGGC
GCCCTAATTGGATTTGACATTCAGATTAATGATGATGATGGAAAGGGTCAAAGGACAGGACAATGT
AATTGGAATGACTCATCTGGAACCGGTTGGCAGTCAACGGCTGTATTTGGTGTGTTACGACTGAAA
TGA

Figure 16: Nucleotide sequences of the 3' end of xylanase gene

WDVVNEAIDTTQPQNLRNTKWLEIIGPEYLELAFRFAHEADPAAKLFYNDFDTENPVKRQ
AIYDMVKALKEKGVPIDGIGCQGHISLNQPPVKNIEDTVKLFSKLGLVEITELDVSVFNSAN
PTTIPKEVLIEQGYRYQQLMEMFKKYKNVLTGVTFWGLKDDMSWLSREKTDVPLLFDKR
FKEKYAYWGMVDASKLPVLIKTYEAADIMPVFASKAELPWDILKTENLTSRSGEPVGSFKS
FYDKKNIYLWINVNAPAKSDSVDVFIDWKNNREDAISKDPRAHAIYHLEWSGISKNSGVK
GSIREIAGGTVMEVSIPIDQRNLKTGKLVGFDIRVNHGTESTSWNDPTNRQSESTSNYGIL
KLGKCPEYSEAIHGTPVIDGEKDPIWETANAISTTKVVIGNNNVATGKFRTIWDQDYLYVY
GEIKDPLLSAVSPNPWEQDSVEFFIDQKFDRTPFYEDDDTQYRVNYKNEKTFGSVVVQNF
KTSTKVVPGGYVVEAAIPLDKIPVKKGALIGFDIQINDDDGKGQRTGQCNWNDSSGTGW
QSTAVFGVLRLK-

Figure 17: Amino acid sequence of the 3' end of xylanase gene

3' end of xylanase gene was obtained from environmental DNA using genome walking. DNA sequencing revealed 1,668 nucleotides encoding 556 amino acids including a putative STOP codon. BLAST analysis showed similarity of this sequence with those belonging to family 10 glycosyl hydrolase.

Once the STOP codon was found, the 5' end of this xylanase gene will be retrieved in order to obtain the full length of the xylanase gene. A set of specific primers were designed (Table 3) from the partial xylanase sequence which was obtained in the beginning (5.3). Two specific primers were designed based on forward or 5'- end region to perform the genome walking.

To retrieve the 5' – end of xylanase gene, PCR amplification based on genome walking approach (Clontech) was performed with two rounds to increase specificity. The construction of four DNA libraries was done using different restriction enzymes (*Eco*RV, *Dra*I, *Stu*I, and *pvu*II) and was then ligated to the Genome Walker adaptor. These DNA libraries were amplified using a set of specific primers which were provided in Table3.

For the genome walking approach, the first or primary PCR uses the outer adaptor primer (AP1) which provided in the kit and an outer gene specific primer (GSPR1) which designed based on the known sequence of partial xylanase. The primary PCR mixture was then diluted and used as a template for a secondary or "nested" PCR with the nested adaptor primer (AP2) and a nested gene-specific primer (GSPR2). The nested adaptor primer (AP2) was provided in the kit and a nested gene-specific primer (GSPR2) was obtained based on another known sequence of partial xylanase. The direction of a set of gene specific primers designed from the known sequence partial xylanase was shown in the figure 18.

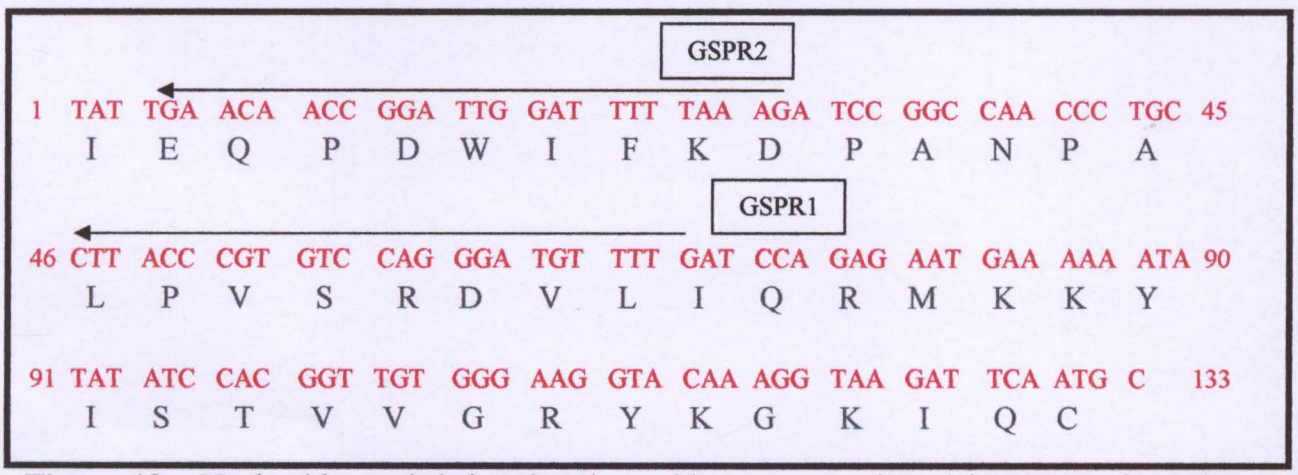

Figure 18: Nucleotides and deduced amino acid sequences of partial xylanase from the sediment of Phu Khieo Bio-Energy at Chaiyaphum province (Bagasse compost soil)

DNA sequencing result revealed 133 nucleotides of the partial xylanse gene. A set of gene specific primers for 5' (GSPR1 and GSPR2 primers) end were designed and indicated by arrows.

For the PCR amplification, the amplified PCR products were obtained (figure 19). The PCR products were analyzed on 0.8% agarose gel electrophoresis. The gel was strained with ethidium bromide and visualized under UV. From the figure, there was no PCR product from the primary PCR. The PCR products from the secondary PCR were obtained.

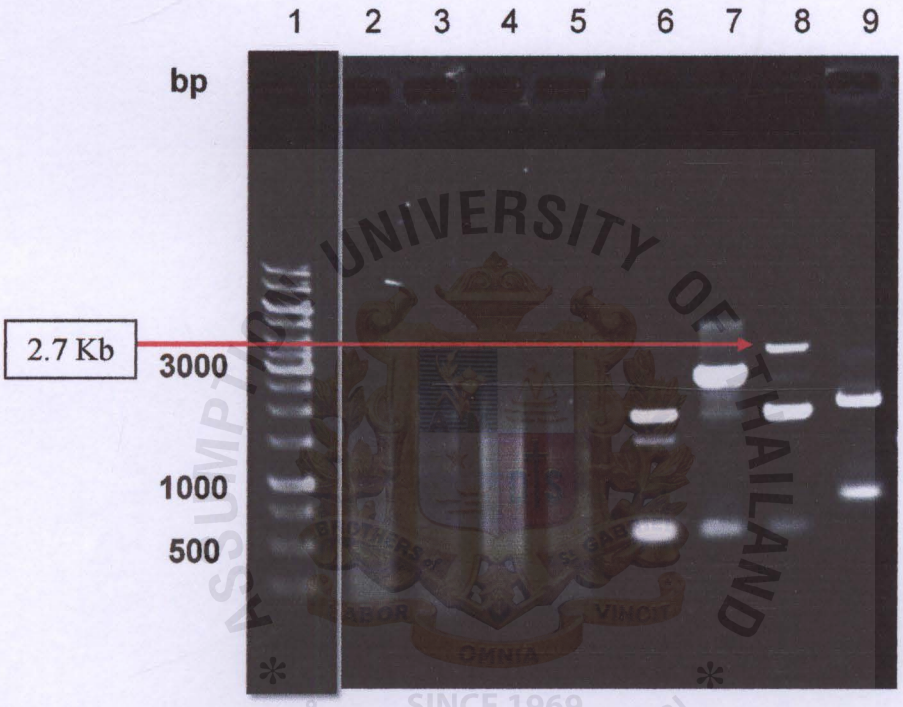

Figure 19: PCR amplification of xylanase for downstream walking at 5'end (Reverse region) PCR amplification of downstream region (3'end) using a set of gene specific primers (GSPF1, and GSPF2 primers). The PCR products were analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. Lane 1 represented 1 Kb DNA marker. Lane 2 to 5 represented PCR product from primary PCR of *Dra* I, *Stu* I, *Pvu* II, *Eco*RV respectively. Lane 6 to 9 represented PCR product from nested PCR of *Dra* I, *Stu* I, *Pvu* II, *Eco*RV respectively.

Each of the DNA fragments with different sizes which began in a known sequence at the 5' -end of GSR1 and GSPR2 and extended into the unknown adjacent genomic DNA were then be purified and cloned into pTZ 57 R/T vector. Restriction enzyme digestion with *Eco*RI and *Bam*HI was performed to verify the DNA insert. The recombinant clones were subjected for the DNA sequencing. DNA sequence analysis and the contig alignment found that obtianed PCR product of approximately 1.5 Kb, 2 Kb, and 2.5 Kb *Dra*I, *Stu*I, and *Eco*RV respectively shown no sequence similarity to family 10 xylanase gene. This can be determined that the PCR amplification produced the non specific products which were not related to family 10 xylanase gene [36].

However, the obtained PCR product of approximately 2.7 kb from the *Pvu*II library revealed the length of 5'end sequence of 848 nucleotides, encoding 282 amino acids. This nucleotide sequence analysis showed that this sequence exhibited 42% identity to endo- 1,4 beta xylanase of *Thermotoga sp. RQ2* (Figure 20).

The contig alignment and sequence analysis shown that the deduced amino acid sequence included a putative START codon in the first genome walking at 5'end region. Therefore, the upstream region was successfully obtained since the START codon was obtained. The upstream sequence was 848 nucleotides encoding 282 amino acids which identity to endo - 1,4 beta xylanase (Figure 21 and 22).

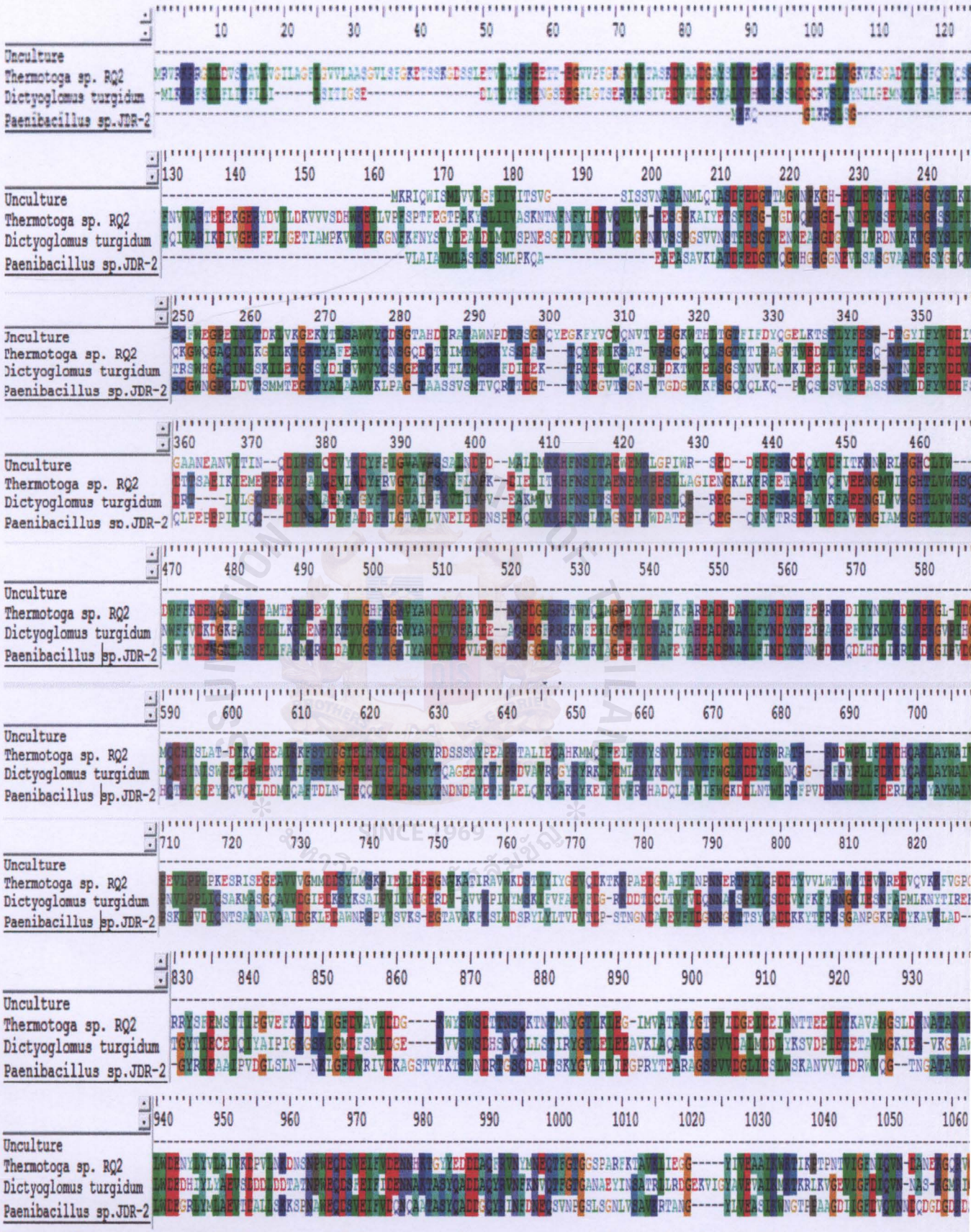

Figure 20: Amino acid alignment of *Bagasse* xylanase gene against other bacterial xylanase genes from the first genome walking at 5'-end. The deduced amino acid sequence of xylanase gene was aligned by Clustal X with other bacterial enzymes in xylanase family: *Thermotoga*

sp. RQ2, *Dictyoglomus turgidum DSM 6724*, and *Paenibacillus sp. JDR-2*. Shading of the alignment represented the degree of conservation. The color shading showed the most conserved region among all sequences compared.

ATGAAACGGATTCAATGGATATCAATGCTCGTTGTTTTAGGATTCATAATTGTGATTACTTC
CGTGGGAAGTATTTCTTCGGTAAATGCTTCAGCCAATATGTTACAGATCGCATCGGATTTT
GAGGATGGCACGACCATGGGCTGGAATCCGAAAGGACATGAAAAACTTGAAGTAAGTAC
CGAGGTAGCCCATTCGGGTAAATACAGTTTGAAGATTGCCGGACGGAGCCAGTTTTGGGA
AGGACCTGAAATCAATCTCACCGACAAACTTGTCAAAGGAGAGAAATATACTTTATCGGCA
TGGGTATACCAAGATAGCGGGACCGCTCATGATATACGGGCGACGGCTTGGAATCCGGAT
ACTTCTAGCGGCAATCAGTATGAAGGTAAATTCTACGTATGTGTTCAAAACGTCACGGTGG
AAAGTGGAAAATGGACCCACCTGACCGGAACATTTATCTTTGATTATCAGGGTGAATTGAA
AACATCTACTTTATATTTTGAGTCACCGGATACTGGTTATATTTTTTATGTCGATGATATAAC
CATCACGGGCGCAGCCAATGAAGCAAATGTAATAACTATCAATCAAGATATTCCTTCACTTT
GCGAGGTTTATAAAGACTACTTCCCAATCGGTGTCGCGGTTCCCAGTTCGGCTTTGAATGA
CCCCGACATGGCGCTGCTCATGAAAAAACATTTCAATAGCATAACGGCGGAATGGGAAAT
GAAACTTGGTCCCATCTGGAGAAGTGAAGATGACTTCGATTTTTCAAAATGTGATCAGTAT
GTGGATTTTATTACGAAAAATAATATGAGATTACGGGGTCACTGTCTGATTTGGCA

Figure 21: Nucleotide sequences of the 5' end of xylanase gene

MKRIQWISMLVVLGFIIVITSVGSISSVNASANMLQIASDFEDGTTMGWNPKG
HEKLEVSTEVAHSGKYSLKIAGRSQFWEGPEINLTDKLVKGEKYTLSAWVYQDSG
TAHDIRATAWNPDTSSGNQYEGKFYVCVQNVTVESGKWTHLTGTFIFDYQGEL
KTSTLYFESPDTGYIFYVDDITITGAANEANVITINQDIPSLCEVYKDYFPIGVAVPS
SALNDPDMALLMKKHFNSITAEWEMKLGPIWRSEDDFDFSKCDQYVDFITKNN
MRLRGHCLIW

Figure 22: Amino acid sequence of the 5' end of xylanase gene

5' end of xylanase gene was obtained from environmental DNA using genome walking. DNA sequencing revealed 848 nucleotides encoding 282 amino acids including a putative start codon. BLAST analysis showed similarity of this sequence with those belonging to family 10 glycosyl hydrolase.

5.5 Amplification of the full – length bagasse xylanase gene

After the 5' and 3' ends of bagasse xylanase sequences were indentified. XynF and XynR (Table3) were used for PCR amplification to obtain the full-length xylanase gene from baggase compose soil. The PCR product of approximately 2.6 Kb was obtained as shown in the figure 23.

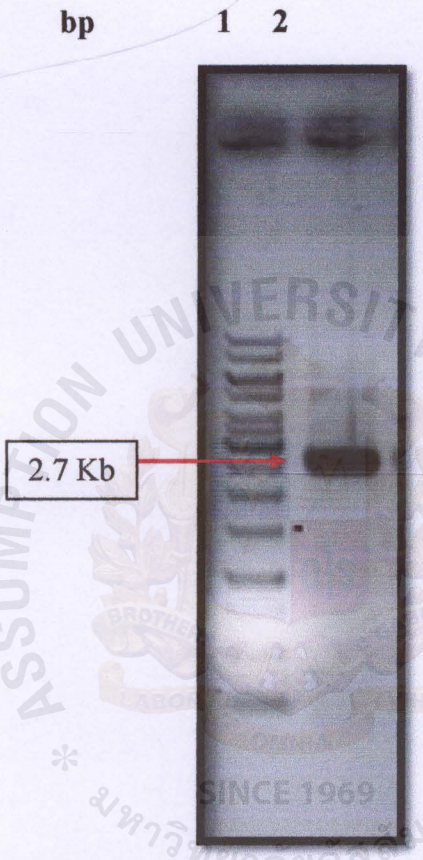

Figure 23: PCR amplification of full length xylanase. The PCR product of xylanase using XynF and XynR primers to obtain the full length gene and analyzed on 0.8% agarose gel electrophoresis and the gel was strained with ethidium bromide. Lane 1 shown 1 Kb DNA marker. Lane 2 shown the amplified PCR product of full length xylanase gene.

The fragment of PCR product was purified and cloned into pTZ 57 R/T vector. DNA sequencing was performed and the result revealed that the full-length of xylanase gene contained 2,649 nucleotides encoding 883 amino acid sequences with a calculated mass of 92 k Da (Figure 24).

The analysis of sequence revealed that this full length xylanase exhibited 43% amino acid sequences identity to Endo-1,4-beta-xylanase from *Thermotoga sp. RQ*, 41% identity to

Endo-1,4-beta-xylanase from *Paenibacillus sp. JDR-2* and 40% identity to Endo-1,4-beta-xylanase from *Dictyoglomus turgidum DSM 672* (Figure 25).

The full-length xylanase gene contained the signal peptide which was analyzed using SignalP 3.0. The full-length xylanase gene presented the location of signal peptide cleavage sites in amino acid sequences from Gram positive bacteria. This xylanase has the most likely cleavage site between position 30 and 31 of the sequences. The signal peptide probability was 1.000. A signal peptide was a short (3-60 amino acids long) peptide chain that directed the transport of a protein. It can be indicated that this protein can be secreted out of the cell when the bacteria synthesized in the cell. It will be very useful in term of the study and application of using enzyme.

The four highly conserved regions related to catalytic and substrate binding site among the xylanase family could be indentified in the full-length xylanase sequence. The 3D structure was preliminary predicted from the deduced amino acid sequences using SWISS-MODEL program in figure 26. Three carbohydrate binding module and one catalytic site were found in the obtained full-length xylanase gene from bagasse compost sediment. Two carbohydrate-binding modules (CBM9) and one carbohydrate-binding CenC like domain (CBM_4_9) were contained on the sequence of full-length xylanase gene. Moreover, the glycosly hydrolase family 10 was found on the full-length xylanase sequence.

O-Glycosyl hydrolases EC:3.2.1. were a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety. A classification system for glycosyl hydrolases, based on sequence similarity, has led to the definition of 85 different families. This classification was available on the CAZy (CArbohydrate-Active EnZymes) web site.

Glycoside hydrolase family 10 GH10 comprised enzymes with a number of known activities; xylanase (EC:3.2.1.8); endo-1,3-beta-xylanase (EC:3.2.1.32); cellobiohydrolase (EC:3.2.1.91). These enzymes were formerly known as cellulase family F.

The microbial degradation of cellulose and xylans required several types of enzymes such as endoglucanases (EC:3.2.1.4), cellobiohydrolases (EC:3.2.1.91) (exoglucanases), or xylanases (EC:3.2.1.8). Fungi and bacteria produced a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family F or as the glycosyl hydrolases family 10 [33].

For CBM_4_9, it functioned as the hydrolase activity, acting on glycosyl bonds. This family included diverse carbohydrate binding domains. The 1,4-beta-glucanase CenC from *Cellulomonas fimi* contains two cellulose-binding domains, CBD(N1) and CBD(N2), arranged in tandem at its N terminus. These homologous CBDs were distinct in their selectivity for binding amorphous and not crystalline cellulose. The tertiary structure of CBDN1 is strikingly similar to that of the bacterial 1,3-1,4-beta-glucanases, as well as other sugar-binding proteins with jelly-roll folds [23].

For CBM9, it contained the modules of approximately 170 residues found so far only in xylanases. The cellulose-binding function has been demonstrated in one case. The 3D structure was β-sandwich. CBM9 or DOMON-like type 9 carbohydrate binding module of xylanases; Family 9 carbohydrate-binding modules (CBM9) played a role in the microbial degradation of cellulose and hemicellulose (materials found in plants). The domain has previously been called cellulose-binding domain. The polysaccharide binding sites of CBMs with available 3D structure have been found to be either flat surfaces with interactions formed by predominantly aromatic residues (tryptophan and tyrosine), or extended shallow grooves. The CBM9 domain frequently occurs in tandem repeats; members found in this subfamily typically co-occur with glycosyl hydrolase family 10 domains and are annotated as endo-1,4-beta-xylanases. CBM9 from *Thermotoga maritima xylanase 10A* was reported to have specificity for polysaccharide reducing ends [24].

Accordingly, this obtained full-length xylanase can be used to study the expression and activity in the further experiment. Due to the predicted structure from the sequence, this enzyme might work well and can be used in the industrial application.

```
atgaaacggattcaatggatatcaatgctcgttgttttaggattcataattgtgattact
 M  K  R  I  Q  W  I  S  M  L  V  V  L  G  F  I  I  V  I  T
tccgtgggaagtatttcttcggtaaatgcttcagccaatatgttacagatcgcatcggat
 S  V  G  S  I  S  S  V  N  A  S  A  N  M  L  Q  I  A  S  D
tttgaggatggcacgaccatgggctggaatccgaaaggacatgaaaaacttgaagtaagt
 F  E  D  G  T  T  M  G  W  N  P  K  G  H  E  K  L  E  V  S
accgaggtagcccattcgggtaaatacagtttgaagattgccggacggagccagttttgg
 T  E  V  A  H  S  G  K  Y  S  L  K  I  A  G  R  S  Q  F  W
gaaggacctgaaatcaatctcaccgacaaacttgtcaaaggagagaaatatactttatcg
 E  G  P  E  I  N  L  T  D  K  L  V  K  G  E  K  Y  T  L  S
gcatgggtataccaagatagcgggaccgctcatgatatacgggcgacggcttggaatccg
 A  W  V  Y  Q  D  S  G  T  A  H  D  I  R  A  T  A  W  N  P
gatacttctagcggcaatcagtatgaaggtaaattctacgtatgtgttcaaaacgtcacg
 D  T  S  S  G  N  Q  Y  E  G  K  F  Y  V  C  V  Q  N  V  T
gtggaaagtggaaaatggacccacctgaccggaacatttatctttgattatcagggtgaa
 V  E  S  G  K  W  T  H  L  T  G  T  F  I  F  D  Y  Q  G  E
ttgaaaacatctactttatattttgagtcaccggatactggttatatttttttatgtcgat
 L  K  T  S  T  L  Y  F  E  S  P  D  T  G  Y  I  F  Y  V  D
gatataaccatcacgggcgcagccaatgaagcaaatgtaataactatcaatcaagatatt
 D  I  T  I  T  G  A  A  N  E  A  N  V  I  T  I  N  Q  D  I
ccttcactttgcgaggtttataaagactacttcccaatcggtgtcgcggttcccagttcg
 P  S  L  C  E  V  Y  K  D  Y  F  P  I  G  V  A  V  P  S  S
gctttgaatgaccccgacatggcgctgctcatgaaaaaacatttcaatagcataacggcg
 A  L  N  D  P  D  M  A  L  L  M  K  K  H  F  N  S  I  T  A
gaatgggaaatgaaacttggtcccatctggagaagtgaagatgacttcgatttttcaaaa
 E  W  E  M  K  L  G  P  I  W  R  S  E  D  D  F  D  F  S  K
tgtgatcagtatgtggattttattacgaaaaataatatgagattacggggtcactgtctg
 C  D  Q  Y  V  D  F  I  T  K  N  N  M  R  L  R  G  H  C  L
atttggcatattgaacaaccggattggatttttaaagatccggccaaccctgccttaccc
 I  W  H  I  E  Q  P  D  W  I  F  K  D  P  A  N  P  A  L  P
gtgtccagggatgttttgatccagagaatgaaaaaatatatatccacggttgtgggaagg
 V  S  R  D  V  L  I  Q  R  M  K  K  Y  I  S  T  V  V  G  R
tacaaaggtaagattcaatgctgggatgttgttaatgaggccattgatacaacccaacct
 Y  K  G  K  I  Q  C  W  D  V  V  N  E  A  I  D  T  T  Q  P
caaaatttacgaaatacaaaatggcttgagatcatcggacccgaatatttggaacttgct
 Q  N  L  R  N  T  K  W  L  E  I  I  G  P  E  Y  L  E  L  A
ttcaggtttgcccatgaagccgacccggcggcaaagctttttataacgatttcgatacc
 F  R  F  A  H  E  A  D  P  A  A  K  L  F  Y  N  D  F  D  T
gaaaatccggtgaaacggcaggccatatatgacatggttaaagcccttaaagaaaaaggc
 E  N  P  V  K  R  Q  A  I  Y  D  M  V  K  A  L  K  E  K  G
gttcccattgacggtatcggttgtcagggacatatatcgctaaaccaacctccggttaaa
 V  P  I  D  G  I  G  C  Q  G  H  I  S  L  N  Q  P  P  V  K
aatattgaagatactgtgaaactcttttccaaactgggtctggttgaaattaccgaactc
 N  I  E  D  T  V  K  L  F  S  K  L  G  L  V  E  I  T  E  L
gatgtgtccgtatttaatagcgcaaatccgacaacgatacccaaagaagttttaattgag
 D  V  S  V  F  N  S  A  N  P  T  T  I  P  K  E  V  L  I  E
cagggttaccggtatcagcaactcatggaaatgtttaaaaaatataaaaacgttctcacc
 Q  G  Y  R  Y  Q  Q  L  M  E  M  F  K  K  Y  K  N  V  L  T
ggtgtaacattttggggactcaaggatgatatgtcctggctctcaagggaaaagaccgat
 G  V  T  F  W  G  L  K  D  D  M  S  W  L  S  R  E  K  T  D
gtgccacttttgtttgacaagaggttcaaggaaaaatacgcctattggggtatggttgat
 V  P  L  L  F  D  K  R  F  K  E  K  Y  A  Y  W  G  M  V  D
gcttcgaagctcccggttcttatcaaaacttatgaagcagcggatatcatgccggttttt
 A  S  K  L  P  V  L  I  K  T  Y  E  A  A  D  I  M  P  V  F
gcctcaaaagcagaacttccctgggatatccttaaaactgaaaacctaaccagccggtcc
 A  S  K  A  E  L  P  W  D  I  L  K  T  E  N  L  T  S  R  S
```

```
ggtgaaccggtgggaagttttaaatccttctatgacaaaaaaaatatttatttatggatc
 G  E  P  V  G  S  F  K  S  F  Y  D  K  K  N  I  Y  L  W  I
aatgtgaatgcccctgcaaaaagcgactccgtggacgtatttattgattggaaaaataac
 N  V  N  A  P  A  K  S  D  S  V  D  V  F  I  D  W  K  N  N
agggaggacgctatttcaaaagatccgcgtgcgcacgccatttatcatttggaatggtca
 R  E  D  A  I  S  K  D  P  R  A  H  A  I  Y  H  L  E  W  S
ggaatcagcaagaacagtggcgtcaaaggttccatccgggaaattgccggaggcaccgtc
 G  I  S  K  N  S  G  V  K  G  S  I  R  E  I  A  G  G  T  V
atggaggtatcaattcccattgatcagcgcaatctcaaaaccggcaaactggtaggtttc
 M  E  V  S  I  P  I  D  Q  R  N  L  K  T  G  K  L  V  G  F
gatatccgggtgaatcatggaaccgagtccaccagttggaatgatcccaccaaccgccag
 D  I  R  V  N  H  G  T  E  S  T  S  W  N  D  P  T  N  R  Q
tcggagagcacttcaaattatggaatcttaaaacttggtaaatgtccggaatattcggaa
 S  E  S  T  S  N  Y  G  I  L  K  L  G  K  C  P  E  Y  S  E
gctatccatggaacccctgttatcgatggagaaaaggacccaatctgggaaacggcaaat
 A  I  H  G  T  P  V  I  D  G  E  K  D  P  I  W  E  T  A  N
gctatttctaccaccaaggtagttataggtaataataacgttgccactggaaaattcaga
 A  I  S  T  T  K  V  V  I  G  N  N  N  V  A  T  G  K  F  R
accatctgggatcaagattacctgtatgtttatggagaaataaaggatccgttgcttagc
 T  I  W  D  Q  D  Y  L  Y  V  Y  G  E  I  K  D  P  L  L  S
gcagtcagcccaaatccctgggaacaggattcagtggaattcttcattgaccaaaaattt
 A  V  S  P  N  P  W  E  Q  D  S  V  E  F  F  I  D  Q  K  F
gatagaactccgttttacgaggatgatgatacccaataccgggtgaattacaagaatgaa
 D  R  T  P  F  Y  E  D  D  D  T  Q  Y  R  V  N  Y  K  N  E
aaaacttttgggagtgtggttgtgcaaaatttcaaaaccagcaccaaggtagtccccggg
 K  T  F  G  S  V  V  V  Q  N  F  K  T  S  T  K  V  V  P  G
ggatatgttgtggaagctgcgattccattggataaaataccagttaagaaaggcgcccta
 G  Y  V  V  E  A  A  I  P  L  D  K  I  P  V  K  K  G  A  L
attggatttgacattcagattaatgatgatgatggaaagggtcaaaggacaggacaatgt
 I  G  F  D  I  Q  I  N  D  D  D  G  K  G  Q  R  T  G  Q  C
aattggaatgactcatctggaaccggttggcagtcaacggctgtatttggtgtgttacga
 N  W  N  D  S  S  G  T  G  W  Q  S  T  A  V  F  G  V  L  R
ctgaaatga
 L  K  -
```

Figure 24: Nucleotides and deduced amino acid sequences of full length xylanase gene from bagasse compost soil.

The nucleotides sequences of full - length xylanase contained 2,649 nucleotides encoding 883 amino acid sequences. The deduced amino acid sequences were shown in a single letter code under the respective codon. BLAST analysis showed similarity of this sequence with those belonging to family 10 glycosyl hydrolase.

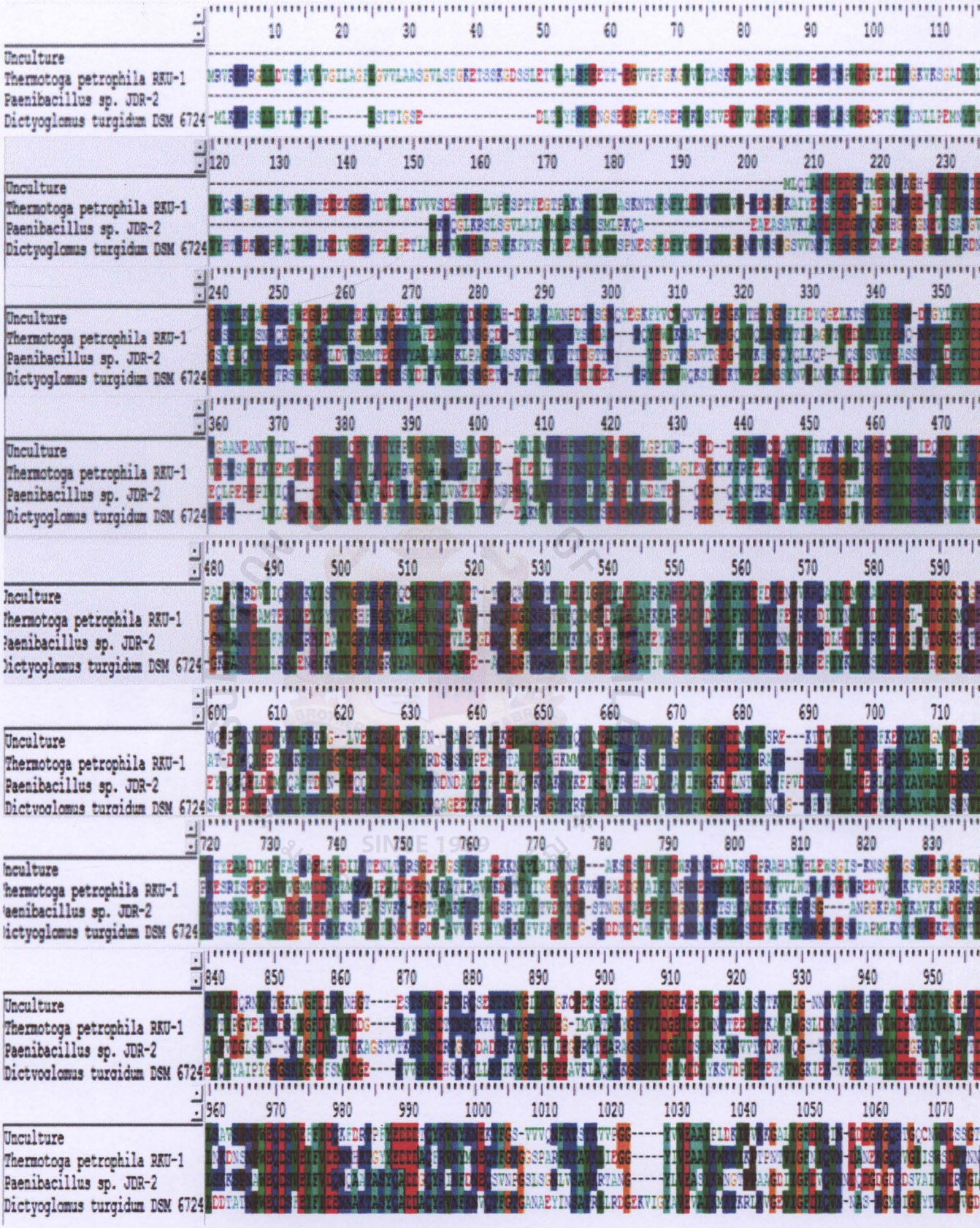

Figure 25: Amino acid sequences alignment of full length *Bagasse* xylanase gene against other bacterial xylanase from bagasse compost soil. The deduced amino acid sequences of full length xylanase were aligned by clustal W with other bacterial enzymes in xylanase

family: *Thermotoga sp. RQ, Paenibacillus sp. JD* and *Dictyoglomus turgidum DSM 672.* Shading of the alignment represented the degree of conservation. The color shading showed the most conserved region among all sequences compared.

Figure 26: 3D structure prediction

Predicted tertiary structure of full – length bagasse xylanase was performed using An Automated Comparative Protein Modelling Server, (SWISS-MODEL) in www.expasy.org/swissmod/SWISS-MODEL.html

CHAPTER VI
CONCLUSIONS

In this study, a full-length xylanase gene was successfully obtained from the environmental DNA at power plant, Phu Khiao Bio-Energy, Chaiyaphum province. A genome walking approach was used to obtain the full-length gene from the known partial xylanase sequence.

The result shown that a full-length xylanase gene contained 2,649 bp open reading frame encoding 883 amino acid residues with a calculated of 92 kDa. BLAST analysis showed similarity of this sequence with those belonging to family 10 glycosyl hydrolase. This enzyme classified as Endo-1,4-beta-xylanase with three carbohydrate binding module. The predicted 3D structure of Xylanase depicted $(\beta / \alpha)_8$ and contained signal peptide in the sequence.

This xylanase enzyme might have some benefits in the enzyme application due to the predicted properties from the sequence. However, the cloning for expression and test for the activity of this enzyme will be further analyzed in the future research.

REFERENCES

1. Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201.
2. Beg QK, Kapoor M, Mahajan L, Hoondal GS. Microbial xylanase and their industrial application: review. Appl Microbiol Biotechnol. 2001; 56: 326-338.
3. Bieley P. Vrsanska M. Tenkanen M. Kluepfel D. Endo-β-1,4 xylanase families: differences in catalytic properties. J Biotechnol. 1997; 57: 151-166.
4. Daniel R. The soil metagenome - a rich resource for the discovery of novel natural products. Curr Opin. Biotechnol. 2004; 15: 199-204.
5. Dhillon A., Gupta J.K., Jauhari B.M. and Khanna S. 2000. DNA isolation from soil samples for cloning, *Bioresource Technology 73*, 273–277.
6. Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis 18.
7. Haki GD, Rakshit SK. Development in industrially important thermostable enzyme: a review. Biores Technol. 2003; 89: 17-34.
8. Henrissat B. A classification of glycosly hydrolase based on amino acid sequence similarities. Bioresource Technol. 1991; 89: 17-34.
9. Henrissat B. Davies G. Structural and sequence - based classification of glycosly hydrolase. Curr Opin Struct Biol. 1997; 7: 637-644.
10. Henrissat B. Bairoch A. New families in the classification of glycosly hydrolases based on amino acid sequence similarities. Biochem J. 1993; 293: 781-788.
11. Henrissat B. Bairoch A. Updating the sequence - based classification of glycosly hydrolases. Biochem J. 1996; 316; 695-696.
12. Hugenholtz, P., and N. R. Pace. 1996. Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotechnol.14:190–197.
13. Johnson KG, Harrison BA, Schneider H, Mackenzie CR, Fontana FD. Microbial degradation of hemicellulosic materials. Appl Biochem Biotechnol. 1989; 20/21: 245-258.
14. Jones DT. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195-202.

15. Kirk O, Borchert TV, Fuglsang CC. Industrial enzyme applications. Curr Opini Biotechnol. 2002; 13: 345-351
16. K. Smalla, N. Cresswell, L.C. Mendonca-Hagler, A. Wolters and J.D. van Elsas. Rapid DNA extraction protocol from soil for polymerase chain reaction-mediated amplification. J Appl Bactriol. 1993; 74: 78-85.
17. Kuhad, RC, Singh A. Lignocellulosic biotechnology: current and future prospects. Crit Rev Biotechnol. 1993; 13: 151-172.
18. L. Arlene Porteous and John L. Armstrong. Recovery of bulk DNA from soil by a rapid, small-scale extraction method. Curr Microbiol. 1991; 22: 345-348.
19. Lynn Yarris. Driving for Biofuels: New Technique Speeds Search for Biofuel Microbes. 2009
20. Wong, KKY, Tan LUL, Saddler JN. Multiplicity of β- 1,4 xylanase in microorganisms: function and applications. Microbiol Rev. 1988; 52: 305-317.
21. Malherbe S, Cloete TE (2003). Lignocellulose biodegradation: fundamentals and applications: A review. Environ. Sci. Biotechnol.1:105-114.
22. Marchler-Bauer A, Bryant SH (2004), *"CD-Search: protein domain annotations on the fly."*, Nucleic Acids Res.32(W)327-331
23. Marchler-Bauer A et al. (2009), *"CDD: specific functional annotation with the Conserved Domain Database."*, Nucleic Acids Res.37(D)205-10.
24. Marchler-Bauer A et al. (2011), *"CDD: a Conserved Domain Database for the functional annotation of proteins."*, Nucleic Acids Res.39(D)225-9
25. Michelle r. Rondon, et al. Cloning the Soil Metagenome: a Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. Applied and environmental microbiology, 2000; 66: 2541–2547.
26. Okeke, B.C. and Obi, S.K.C. 1994. Lignocellulose and sugar compositions of some agro-waste materials. *Bioresource Technology 47*, 283–284.
27. Parveen Kumar, Diane M. Barrett, Michael J. Delwiche and Pieter Stroeve. Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. I&EC research. 2009; 48: 3715-3716
28. Prade RA. Xylanases: from biology to Biotechnology. Biotechnol Genet Eng Rev. 1995; 13:101-131.

29. Puls J, Schuseil J. Chemistry of hemicelluloses: relationship between hemicelloulose structure and enzymes required for analysis. In: Hemicellulose and hemicellulases. Portland Press, London. 1993: 1-28.
30. Sun Y, Cheng J (2002). Hydrolysis of lignocellulosic material from ethanol production: A review. Biores. Technol. 83: 1-11.
31. Sunna A, Antanikian G. Xylanolytic enzymes from fungi and bacteria. Crit Rev Biotechnol. 1997; 17: 39-67.
32. Sunna A, Bergquist P. A gene encoding a novel extremely thermostable β-1,4 xylanase isolated directly from environmental sample. Extremeophiles. 2003; 7: 63-70.
33. Schwede T, Kopp J, Guex N, and Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research 31: 3381-3385.
34. Thomas WJ. Biochemistry and genetics of microbial xylanases. Curr Opin Biol. 199; 7: 337-342.
35. Tsai YL, Olson BH. Rapid Method for Separation of Bacterial DNA from Humic Substances in Sediments for Polymerase Chain Reaction. Appl Environ Microbiol. 1992; 2292-2295.
36. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF and Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life Journal of Molecular Biology, 337, 635-645.
37. Zdobnov E.M. and Apweiler R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro Bioinformatics 17: 847-848.
38. Viriya Nitteranon. Cloning and expression of enzymes from glycosyl hydrolase family (xylanase and pullulanase) from the sediment of Bor Khlueng Hot Spring. Mahidol University. 2006.