UNIVERSITI PUTRA MALAYSIA

DETERMINATION OF THE NUCLEOTIDE SEQUENCE OF A THERMOSTABLE XYLANASE GENE FROM *Bacillus coagulans* ST-6

AINU HUSNA BINTI M S SUHAIMI

FSAS 1999 30

DETERMINATION OF THE NUCLEOTIDE SEQUENCE OF A THERMOSTABLE XYLANASE GENE FROM *Bacillus coagulans* ST-6

By

AINU HUSNA BINTI M S SUHAIMI

Thesis submitted in Fulfilment of the Requirements
for the Degree of Master of Science in the
Faculty of Science and Environmental Studies,
Universiti Putra Malaysia.

July 1999

ACKNOWLEDGMENTS

In the name of Allah , the most gracious , the most merciful. Alhamdullillah, I thank Allah for allowing me to complete this project and thesis.

I would like to thank my chairperson, Assc. Prof Dr Abdullah Sipat , Department of Biochemistry and Microbiology, Universiti Putra Malaysia, for his guidance, encouragement, patience and constructive suggestions throughout the whole course of study. I would also like to thank my co-supervisors , Assc. Prof Dr Khatijah Yusoff and Dr. Harikrishna a/l Kulaveerasingam for all their support and discussions. Not forgetting Dr. Raha Abdul Rahim, Dr. Tan Wen Siang and Dr. Che Nyonya Abdul Razak, for their indispensable advices and suggestions.

All my friends in Lab 202 and 143 especially Pin ,Geok Yong, Wai Kean, Yousif, Omeima, Subha, Chiew Ling, Putery, Chui Fong, Pria, Law , Laili, VJ, Alan, thank you for all the help and support and for making my life in UPM a memorable and enjoyable one.

My beloved parents, parents in-laws and all family members, thank you so much for all the love, support and guidance. To my love and inspiration my husband , Azran Adzmi and my son, Ahmad Amiruddin Azran, thank you for making my life so complete, to both of you..this thesis is for you..

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF ABBREVIATIONS

A	adenine based nucleotide
bp	basepair
C	cytosine based nucleotide
cDNA	complementary DNA
DEAE	diethylaminoethyl
DNA	deoxyribonucleic acid
EDTA	ethylenediamine tetra acetic acid
G	guanine based nucleotide
kA	kilodalton
kb	kilo basepair
M	Molar
mA	milliamphere
mg	milligram
ml	milliliter
mM	millimolar
MW	molecular weight
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
pI	Isoelectric point
RBB	Remazol Brilliant Blue
rpm	revolutions per minute
SDS	sodium dodecyl sulfate

T	thymine based nucleotide
TBE	tris-borate buffer
TEMED	N.N.N',N'-tetramethyl-ethylenediamine
µg	microgram
µl	microlitre
µmol	micromol
YT	yeast tryptone

Abstract of the thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Master of Science.

THE DETERMINATION OF THE NUCLEOTIDE SEQUENCE OF A THERMOSTABLE XYLANASE GENE FROM *Bacillus coagulans* ST-6

By

AINU HUSNA M S SUHAIMI

June 1999

Chairman : Associate Professor Dr. Abdullah Sipat, Ph.D.

Faculty : Science and Environmental Studies.

PBNX1 is a recombinant plasmid containing vector plasmid pUC19 and a 2.6 kb fragment of *Bacillus coagulans* ST-6 genomic DNA which contains a xylanase gene. PBNX2 is another recombinant plasmid with the same insert DNA but in a pUC18 vector. Both plasmids expressed xylanase activity when grown on RBB-Xylan agar plates. Preliminary to nucleotide sequencing, the recombinant plasmids were modified using restriction enzyme deletion to remove a segment of the insert DNA. The original insert DNA which was 2.6 kb was successfully reduced to a 0.8 kb and a 1.8 kb fragment in pBNX1A and pBNX2A respectively. The deletion was done using restriction endonuclease *Sal*1 and the resulting deletion mutants together with the original clones were used to determine the nucleotide sequence of the xylanase gene.

By primer walking, 1420 bp of the forward sequence was obtained where an open reading frame (ORF) was found at 544 bp of the insert DNA. This 630 bp frame

was preceeded by the putative *E.coli* -10 and -35 promoters. No sequence corresponding to the signal peptide was found in this sequence.

The open reading frame (ORF) was translated into a peptide of 210 amino acid residues. This protein belonged to Family G11 of the Glycosyl Hydrolase family and had 59% homology with *Bacillus stearothermophilus* xylanase and 54% homology with xylanase from *Aeromonas caviae*. Eleven out of 20 completely conserved amino acids in this family were also conserved in this sequence and two conserved glutamate residues, E104 and E186 were directly involved in the enzyme's acid-catalytic mechanism. Secondary structure prediction showed that this enzyme consisted of two α-helices and 10 β-strands. Phylogenic studies showed that the primary structure of the enzyme was most closely related to *Bacillus pumilus* xylanase's primary structure.

The analysis of the deduced amino acid sequence showed that there were five cysteine residues in this sequence compared to none in four other mesophilic xylanases. These cysteine residues can form internal disulfide bonds among themselves which can increase the stability of the protein. Analyzing the predicted secondary structure, an extra α-helical structure which is a more stable secondary structure was observed in comparison to other mesophilic xylanases. These two factors namely the presence of cysteine residues and the extra α-helical structure may have an important role in determining the thermostability of this enzyme.

Abstrak tesis yang dikemukan kepada Senat Universiti Putra Malaysia untuk memenuhi keperluan Ijazah Master Sains.

PENENTUAN JUJUKAN NUKLEOTIDA GEN XILANASE YANG TERMOSTABIL DARI *Bacillus coagulans* ST-6

Oleh

AINU HUSNA M S SUHAIMI

Jun, 1999

Pengerusi : Profesor Madya Dr. Abdullah Sipat, PhD.

Fakulti : Sains dan Pengajian Alam Sekitar.

PBNX1 ialah satu plasmid rekombinan yang mengandungi plasmid vektor . pUC19 berserta satu DNA selitan yang merupakan sebahagian dari DNA genomik *Bacillus coagulans ST-6* bersaiz 2.6 kb yang mengandungi gen xilanase. PBNX2 ialah satu lagi plasmid rekombinan yang mengandungi DNA selitan yang sama tetapi didalam plasmid vektor pUC18. Kedua-dua plasmid ini mengekspreskan aktiviti xilanase apabila dipilih menggunakan agar media bercampur RBB-xilan. Sebelum penjujukan DNA dijalankan, klon-klon ini telah diubah dengan menggunakan enzim pembatas untuk menyingkirkan sebahagian daripada DNA selitan ini. Pensubklonan telah dilakukan dengan menggunakan enzim pembatas *Sal*I. DNA selitan yang pada asalnya bersaiz 2.6 kb telah dikurangkan kepada 0.8 kb di dalam PBNX1A dan 1.8 kb di dalam PBNX2A. Mutan-mutan yang terhasil bersama-sama klon-klon asal telah digunakan untuk melakukan penjujukan DNA.

Dengan menggunakan 'primer walking' 1420 daripada jujukan kehadapan telah diketahui di mana rangka bacaan terbuka telah dijumpai pada bes yang ke 544 pada DNA seluran itu. Rangkaan bacaan terbuka yang bersaiz 630 bp ini didahului oleh promoter *E Coli* –35 dan –10.

Rangkaan bacaan terbuka ini ditranslasikan kepada peptid yang mengandungi 210 bes amino asid. Protein ini tergolong di dalam kumpulan G11 dari keluarga Glikosil Hidrolase dan mempunyai 59% homologi dengan *B. stearothermophilus* dan 54 % homologi dengan xilanase dari *Aeromonas caviae*. Sebelas dari 20 amino asid yang terpulihara di dalam keluarga ini juga terpulihara di dalam jujukan ini. Dua daripada residu ini, glutamat E104 dan E186 terlibat secara langsung dengan mekanismanya yang bersifat hidrolisis berasid.

Penganalisaan ke atas jujukan amino asid ini, menunjukkan kehadiran 5 residu sistina berbanding dengan tiada residu sistina di dalam empat xilanase lain yang mesofilik. Residu-residu sistina ini boleh membentuk ikatan disulfida dalaman diantara mereka. Penganalisaan ke atas struktur sekunder jujukan ini juga menunjukkan kehadiran struktur α-heliks yang lebih berbanding dengan xilanase mesofilik yang lain. Struktur α-heliks ini adalah yang lebih stabil berbanding dengan struktur β-strand. Kedua-dua faktor ini mungkin memainkan peranan utama didalam menentukan termostabiliti enzim ini.

CHAPTER I

INTRODUCTION

According to Aspinall (1973), much of plant matter consists of celluloses, hemicelluloses, pectic substances, other polysaccharides and glycoproteins. In nature, cellulose is always associated with hemicellulose. While cellulose is a homopolymer consisting of just β-D-

one unit for its main chain as xylan or as a heteropolymer with two or more units such as glucomannans (Fengel, 1983). Xylan is the major component of plant hemicellulose. It is composed of a β-(1,4) linked D-xylopyranosyl

substituted with various side chains (Joseleau et al., 1992) .

Many end products of xylan degradation have potential to be used industrially, such as in the food industry (Biely, 1985) and feed industry (Bedford and Classen , 1992). In order to obtain these end products, xylan needs to be hydrolyzed. For xylan hydrolysis, enzyme hydrolysis using microbial xylanase is the most efficient. However, the application of microbial xylanase in the industrial world has not been proven to be economically feasible. The cost is too high and it is difficult to get sufficiently large amount of xylanase enzyme. A lot of studies have been done on

xylanase purified from a variety of microorganisms. A lot more however, needs to be known about xylanase function and structure in order to increase the efficiency of microbial xylanase for industrial applications. This can be done by studying the molecular aspects of the xylanase structure and function.

One common way to study the molecular aspects of xylanase is by cloning the gene involved and obtaining the nucleotide sequence. To date, more than 60 xylanase genes have already been cloned and sequenced, and from the data obtained, xylanases can be divided into two families, the F10 and the G11 of the Glycosyl Hydrolase according to the nature of their catalytic sites (Henrissat, 1995). As the compilation of data on xylanase increases, understanding of the enzyme also increases. Thus, manipulation of the enzyme to better serve humankind is viable.

This thesis is a report on the sequencing of the xylanase gene in a recombinant plasmid pBNX1. The plasmid consists of the vector plasmid pUC19 with an insert DNA of size 2.6 kb fragment from the thermophilic, xylanolytic bacteria *Bacillus coagulans* ST-6. The results obtained from the sequencing will add to the data already gathered on xylanase in the databank and the analysis of the deduced amino acid sequence will also provide us with a better understanding of xylanase structure and function.

CHAPTER II

LITERATURE REVIEW

Polyoses

Hemicellulose is defined as fractions isolated or extracted from plants with dilute alkali (Schulz,1891). Staudinger and Reineke (1939) however, referred to hemicellulose as polyoses. Polyoses differ from cellulose in three aspects (Fengel and Wegener, 1983), they have shorter molecular chains, composed of more than one sugar and have branching of chain molecules.

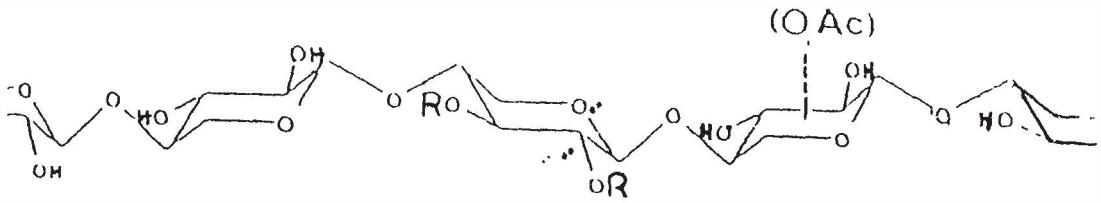
Among the polyoses, xylans are the most represented in nature (Joseleau, 1992). After cellulose, it is the most abundant renewable polysaccharide in nature. The main component of xylan is D-xylose, a five-carbon sugar. Xylans also have traces of L-arabinose. In nature, xylans exist in different forms, depending on its botanical origin and cytological localization. Xylans are the most abundant polyoses in the cell wall of land plants where they constitute at least 30 % of the dry weight and are constituents of secondary walls of tissues having structural functions (Joseleau et al., 1992). Xylans also function as reserves in primary walls of seeds and bulbs and primary walls of growing cell (Joseleau et al., 1992).

Xylan

Chemical Structure

Like most polysaccharides of plant origin, the structure of xylan varies. This is due to its presence in a variety of plant species and their distribution in more than one type of tissues and cells. Basically, all land xylans are characterized by a β-1,4 linked -D-xylopyranosyl main chain. This structure is shown in Figure 1. Examples of this unbranched xylan are found in esparto grass (Chanda et al., 1950) and tobacco stalks (Eda et al., 1976). Its complexity increases with the addition of variable substituents and side chains. Xylans can be divided into four main families based on the side chains, namely, linear homoxylan, which has no side chains at all; arabinoxylan, which has only one side chains of single terminal units of α-L-arabinofuranosyl substituents; glucuronoxylans, which has α-D-glucuronic acid or/and its 4-O-methyl ether derivative and glucuronoarabinoxylan, where α-L-arabinose and α-D- glucuronic (and 4 - O -methyl - α - D glucuronic) are present at the same time (Joseleau et al.,1992).

There are differences between xylans obtained from hardwood (angiosperms) and softwood (gymnosperms). Xylans from angiosperms are acetylated, usually at C2 and C3 where the OH group is substituted with an acetyl group. This acetylation contributes to the solubility of hardwood xylan. Gymnosperms xylans on the other



R : α-D-GlcpA(1
4-Ome-α-D-GlcpA(1→2)Xyl.
α-L-Araf(1→3)Xyl...
α-L-Araf(1→2)Xyl...
β-D-Galp(1→5)α-L-Araf(1→3)Xyl...
β-D-Xylp(1→2)α-L-Araf(1→3)Xyl...
α-L-Araf(1→2,1→3 and 1→2,3 Araf)$_n$ (1→ 3)Xyl...
Feruloyl
p-coumaroyl
Lignin

Figure 1 : General Structure of Xylan (adapted from Joseleau et al., 1992)

hand has no acetyl group, and instead has the presence of arabinofuranose and methylglucuronic acid (Fengel and Wegener, 1983; Woodward, 1984).

In all ultrastructural localization in plants, xylans do not occur on its own, it will always interact with other polysaccharides, lignin or other phenolic acids where the bonding involved are either covalent or non-covalent bonds (Joseleau et al.,1992). Non-covalent bonding essentially hydrogen bonding, interacts xylans with other polysaccharides and covalent bonds interconnect xylans with lignin and other phenolic acids.

Hydrolysis and Utilization of Its End Products

There are three ways to hydrolyze xylan, namely using alkali, acid or enzyme treatment. The method of choice however, is enzyme hydrolysis, using xylanase. It is a more efficient process as it is more specific and will be able to produce xylose free from other monosaccharides. Other advantages of xylanase hydrolysis include it is milder and there is less production of toxic by-products when compared to acid hydrolysis. This method however, is relatively new and is still at the research stage.

Upon hydrolysis, the end-product sugar moieties obtained can be used for a variety of applications. One very important application is the use of xylans as a source of fuel. As our fossil fuels are depleting fast, a renewable substitute is greatly needed. Products from processes such as processing woody biomass, hardwood and softwood,

agricultural residues such as sugar cane pulp and corn stalks and waste products from paper industry can provide us with valuable fuel (Woodward, 1988). The sugar produced can also be fermented to produce ethanol and butanol, both much in demand. (Woodward, 1988).

Xylanase

Physiological Role

In nature, xylanase plays a major role in the biodegradation of plant litter. The function of xylans here is to provide a source of metabolizable energy. It is shown that the rate of biodegradation of the complex polysaccharides in soil plant litter is dependant on the microbial population in the soil (Woodward, 1984). If the area is heavy with a metal such as fertilizer derived from copper, the litter decomposition is retarded as the microbial population is inhibited, resulting in a decrease in xylanases and cellulases. Xylanases and other carbohydrases are also involved in bark and wood degradation. *Aspergillus fumigatus*, for example, produces an extracellular xylanase when grown on beech, oak and sawdust (Flannigan and Sagoo, 1977).

The dietary intake of ruminant animals can comprise of up to 40% hemicellulose. It is an important source of energy and is degraded by xylanases from their rumen microbes. William and Withers (1981) discovered that most hemicellulolytic strains belong to *Eubacterium*, *Ruminococcus* and *Bacteroides*,

Xylanases are also beneficial to plant pathogens. *Colletorichum lindemuthanianum* and *Helminthosporium maydis* are two pathogens that produce polysaccharide-degrading enzymes including xylanases to disrupt the cell walls of bean and corn plants respectively for their carbon source (Anderson, 1978).

Other roles of xylanase include acting as a hydrolase on stored reserves present in the endosperm of cereal grains and as one of the enzymes associated with over ripening (Woodward, 1984).

Occurrence

Sources of xylanases are ubiquitous. They can be obtained from both prokaryotes and eukaryotes (Dekker and Richard, 1976) as well as from marine and terrestrial environments. Xylanase can be found in bacteria occurring naturally in the environment, also in rumen caecal bacteria and in fungi (saprophytes, phytopathogens and mycorrhiza). Xylanases can also be found in protozoa, in the intestines of insects, snails and in germinating plant seeds (Bastawde,1992). The presence of xylanases has not been demonstrated in vertebrate animal tissues, although they have been found in the intestines, rumen and caecum of ruminants. However, this xylanase is produced by the microbial flora and fauna in the vertebrates and not by the vertebrates tissues themselves (Woodward, 1984). Among all sources, xylanase from bacteria and fungi are the most studied. Some of these microorganisms include the genus *Bacillus*, *B.coagulans 26* (Esteban et al 1983) and *B.circulans* (Yang , 1989) . Other

microorganisms that produce xylanase include *Streptomyces lividans* (Chen et al., 1998), *Cryptococcus albidus* (Boucher et al., 1988) , *Butyrivibrio fibrisolvens* H17C (Lin and Thomson, 1991) and *Aureobasidium pullulans* Y-2311-1 (Li et al., 1994). There are also microorganisms that produce thermophilic xylanase and these include *Thermoanaerobacterium saccharolyticum* B6AR1 (Lee, 1993) and *Bacillus stearothermophilus* 21 (Baba et al., 1994).

Classification

Xylanases can be classified as follows:

1. β-Xylosidases - break down short xylooligosaccharides to xylose.
2. Exo-xylanases - produce xylose at high rates from xylan but produce short oligosaccharides at low rates.
3. Endo- xylanases are of four types:
   a) those that cannot cleave L-arabinosyl initiated branch points, and produce mainly xylobiose and xylose as final products. Generally, they are as active on xylooligosaccharides as xylotriose but at much lower rates on larger substrate.
   b) those that cannot cleave branch points and produce mainly oligosaccharide fragments larger than xylobiose. They are generally inactive on xylotetraose and smaller substrates.
   c) those that can cleave branch points and produce mainly xylobiose and xylose
   d) those that can cleave branch points and produce mainly xylooligosaccharides of intermediate size.

Purification

Purification of any enzyme is important for its characterization. Many methods are involved in the purification of xylanases. These methods are based on the different properties of the enzyme. Some of the methods used are ammonium sulfate precipitation, isoelectric focusing, ion-exchange chromatography, ultrafiltration, gel filtration, affinity chromatography, adsorption chromatography and hydroxyapatite chromatography.

Xylanase can be purified through a single purification step or a combination of two or more steps. The initial step is usually to concentrate the crude enzyme. Ammonium sulfate precipitation was used to purify xylanase from *Bacillus* sp. No. C-125 (Honda et al., 1985) and ion exchange chromatography was used to purify xylanase from *B. circulans* (Esteban et al., 1982).

On the other hand, there have been cases where the enzyme was purified using a combination of two purification methods. Two endoxylanases were purified from *Bacillus* sp. No C-125 by using Sephadex G-75 and Bio-Gel P-30 followed by DEAE-cellulose chromatography (Honda et al., 1985). Three endoxylanases were purified from *Clostridium stercorarium* (Berenger et al., 1985) using gel filtration and ion-exchange chromatography.