UNIVERSITI PUTRA MALAYSIA

CLONING AND EXPRESSION OF ALKALINE PROTEASE GENE FROM BACILLUS STEAROTHERMOPHILUS STRAIN FI

SUHAILI BT AB. HAMID

FSAS 2000 25

CLONING AND EXPRESSION OF ALKALINE PROTEASE GENE FROM *BACILLUS STEAROTHERMOPHILUS* STRAIN F1

By

SUHAILI BT AB.HAMID

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

June 2000

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

CLONING AND EXPRESSION OF ALKALINE PROTEASE GENE FROM *BACILLUS STEAROTHERMOPHILUS* F1

By

SUHAILI BINTI AB. HAMID

June 2000

Chairman : Associate Professor Dr. Che Nyonya Ab. Razak

Faculty : Science and Environmental Studies

Cloning of alkaline protease gene from a thermophilic bacteria, *Bacillus stearothermophilus* F1 (BSF1) was done by PCR cloning method. Genomic DNA of BSF1 was extracted and the highest concentration obtained was 0.46 μg/μL and with a purity of 2.0. Three sets of primers were used including RM5 (5'-CA(CT) GG(ACGT) ACC AA(CT) GTG GC(CGT) GG-3) and RM6 (5'-(ACG)GG GGT (ACG)GC CAT GGA (CGT)CC-3', FOR900 (5-GCA TGC TAC GAT TAA ATA TC-3) and REV900 (5'-CGG CAA TAT CAC TTA GAG TAC C-3') and FOR900 (5'-GCA TGC TAC GAT TAA ATA TC-3') and REV1591 (5-TGC AGC AGA AAG AAG GAA-3') which resulted in amplification of 500-bp, 900-bp and 1500-bp products, respectively. Southern blotting analysis suggested that both 500-bp and 900-bp fragment were present within the 1500-bp fragment. *E. coli* TOP10F' harbouring recombinant plasmid of pCR2.1 TOPO vector and 1500-bp fragment showed proteolytic activity on skim milk agar (1%) by the formation of a clearing zone around the colonies after 5-6 hours incubation at 60°C, with the cultures that was initially grown at 37°C overnight. The complete DNA sequence of 1589-bp was determined, whereby, an open reading frame (ORF) of 1080-bp was found. This ORF was preceded by the putative Shine-Dalgarno (SD) sequence, AGGGGG, with a spacing of 9 bases and two putative *E. coli* -10 (ATTAAT) and -35 (TTTTCA) promoters and was followed by an inverted repeat sequence downstream to the stop codon. The ORF was translated into a peptide of 360 amino acid residues. Comparison of the amino acid sequence with the other proteases showed that it had 84% similarity with Ak1 protease from *Bacillus* sp. Ak1, 53% with PD498 protease from *Bacillus* sp. PD498 and 43% similarity with E79 protease from *Thermoactinomyces* sp. E79 and protease from *B. cereus*. The sequence of 25 amino acid residues corresponding to the signal peptide was found in this sequence. Analysis of the cloned protease showed that 90% of the total enzyme was found in the cell extract. Heating at 70°C for 1-3 hours resulted in an increment of the activity. The recombinant enzyme was partially purified and characterized. The optimum temperature and pH for hydrolysis of azocasein were 85°C and 8.0, respectively. The presence of 10 mM PMSF inhibited the activity of recombinant protease by 99%, suggesting that this enzyme is of serine protease. These result showed that this recombinant enzyme was comparable to the native.

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

PENGKLONAN DAN PENGEKPRESAN GEN YANG MENGKODKAN PROTEASE ALKALI DARI BAKTERIA *BACILLUS STEAROTHERMOPHILUS STRAIN* F1

Oleh

SUHAILI BT AB. HAMID

Jun 2000

Pengerusi : Profesor Madya Dr. Che Nyonya Ab. Razak

Fakulti : Sains dan Pengajian Alam Sekitar

Dalam kajian ini, gen yang mengkodkan protease alkali dari bakteria termofilik *Bacillus stearothermophilus* F1 telah diklonkan melalui kaedah pengklonan tindak balas berantai polimerase. Pengekstrakan genomik DNA berjaya mendapatkan sampel DNA dengan kepekatan dan ketulenan tertinggi iaitu 0.46 μg/μL dan 2.0. Tiga gabungan jujukan primer telah digunakan termasuk RM5 (5'-CA(CT) GG(ACGT) ACC AA(CT) GTG GC(CGT) GG-3) dan RM6 (5'-(ACG)GG GGT (ACG)GC CAT GGA (CGT)CC-3', FOR900 (5-GCA TGC TAC GAT TAA ATA TC-3) dan REV900 (5'-CGG CAA TAT CAC TTA GAG TAC C-3') dan FOR900 (5'-GCA TGC TAC GAT TAA ATA TC-3') dan REV1591 (5-TGC AGC AGA AAG AAG GAA-3') dan ini masing-masing menghasilkan 500 pasangan bes , 900 pasangan bes dan 1500 pasangan bes. Analisis pemblotan Southern mencadangkan

bahawa fragmen 500 pasangan bes dan 900 pasangan bes wujud dalam fragmen 1500 pasangan bes. Bakteria *E. coli* yang mengandungi plasmid rekombinan yang terdiri daripada vektor pCR2.1 TOPO dan fragmen 1500 pasangan bes menunjukkan aktiviti proteolisis di atas agar susu skim. Ini dapat dikesan melalui penghasilan kawasan zon cerah di sekeliling koloni selepas pengeraman 5-6 jam pada suhu 60°C, pada kultur yang telah dieramkan semalaman pada suhu 37°C. Jujukan DNA keseluruhan yang ditentukan, menghasilkan 1589 pasangan bes, di mana didapati satu rangka bacaan terbuka bersaiz 1080 pasangan bes. Rangka bacaan terbuka ini didahului dengan jujukan Shine-Dalgarno, AGGGGG pada jarak 9 pasangan bes dan juga 2 promoter *E. coli*, -10 (ATTAAT) dan -35 (TTTTCA) dan ini diikuti dengan jujukan ulangan berbalik aliran ke bawah daripada kodon penamat. Rangka bacaan terbuka telah ditranslasikan kepada peptida yang mengandungi 360 asid amino. Perbandingan dengan jujukan asid amino protease lain menunjukkan protease F1 mempunyai 84% homologi dengan protease dari bakteria *Bacillus* sp. Ak1, 53% dengan protease dari bakteria *Bacillus* sp. PD498 dan 43% homologi dengan protease dari bakteria *Thermoactinomyces* sp. E79 dan *B. cereus.* Dalam jujukan protease rekombinan ini juga, telah dijumpai jujukan 25 asid amino yang menyerupai jujukan peptida isyarat. Analisa penghasilan protease rekombinan dalam sel *E. coli* menunjukkan bahawa sebanyak 90% daripada jumlah enzim telah ditemui daripada ekstrak sel.

Pemanasan pada suhu 70°C selama 1-3 jam menunjukkan peningkatan aktiviti. Enzim ini telah separa ditulenkan dan dicirikan. Suhu dan pH optima untuk hidrolisis azokasein, masing-masing pada suhu 85°C dan pH 8.0. Kehadiran 10 mM PMSF, telah merencatkan aktiviti enzim rekombinan sebanyak 99 %. Ini mencadangkan bahawa enzim ini adalah dari kumpulan protease serine. Keputusan yang diperolihi menunjukkan enzim rekombinan ini adalah setara dengan enzim asal.

ACKNOWLEDGEMENTS

In the name of Allah, the most gracious, the most merciful. Alhamdulillah, I thank Allah for the strength and blessing in the completion of this study.

I would like to express my sincere appreciation and deep sense of gratitude to the chairman of the supervisory committee, Associate Prof Dr. Che Nyonya Ab. Razak for her advice, guidance, encouragement, patience and support and also for her motherly care throughout the whole course of study. My deep appreciation is also extended to my co-supervisors, Prof. Dr. Abu Bakar Salleh, Assoc. Prof. Dr. Mahiran Basri and Assoc. Prof. Dr. Khatijah Mohd Yusoff, for all their indispensable advice and suggestion. I would also like to thank Dr. Raja Nor Zaliha for her guidance and support. I would like to thank Assoc. Prof. Dr. Abdullah Sipat and Prof Dr. Noraini, for allowing me to use the equipment in their laboratories and also all the lecturers in the Department of Biochemistry and Microbiology for their supports.

My special appreciation is also extended to all my colleagues in Lab 139 and 140 especially Shidah, Leha, Thanges, Ee Lin and also to my ex-labmate, K.Yam (PORIM, Bangi) for their friendship, support and also their help in one way or another. My appreciation also extended to Chin, Goh, Lee, Shila, K. Hal, Laith, Bimo and Nazleen for their help and friendship. I also would like to thank my fellow graduate students in Lab 202 and the Virology Lab especially Filza, Ong Wai Kean, Husna, Majid, Pria, Subha and Wong Sing King for their help and suggestion especially in genetic work. My thanks also to our lab officer, K.Yati and also departmental staff for all the help and support.

Finally, my deepest appreciation and gratitude go to my husband, Ahmad Afandi Murdi and my son, Amirul Haziq and also my beloved parents and family members for their prayers and also for all their love that has given me the courage to accomplish this challenging task.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	adenine base nucleotide
bp	base pair
C	cytosine base nucleotide
dH_20	distilled water
DNA	deoxyribonucleic
EDTA	ethylenediamine tetra acetic acid
G	guanine base nucleotide
h	hour
kD	kilodalton
kb	kilo base pair
M	Molar
mg	milligram
ml	milliliter
mM	millimolar
min	minute
µg	microgram
µL	microliter
ng	nanogram
nm	nanometer
ORF	open reading frame

s	second
SDS	sodium dodecyl sulfate
rpm	revolutions perminute
T	thymine base nucleotide
TBE	tris-borate buffer
v/v	volume per volume
w/v	weight per volume

CHAPTER I

INTRODUCTION

Gene or molecular cloning can be generally defined as the method that facilitates the isolation and manipulation of a specific region in a particular genome by replicating them independently as a part of an autonomous vector (Turner *et al.*, 1997). The development of such technology has provided a direct approach for the production of a wide range of biochemical products from microorganisms (Kumar *et al.*, 1991).

Nowadays, industry such as food processing uses large amount of microbial enzymes; many of which are produced more efficiently and at a lower cost by using genetically engineered microorganisms. Atkinson (1989) reported that such recombinant proteins originating from a microorganism can be produced at expression levels of 4 to 50% of the soluble protein by the new hosts.

Proteases which is synonymous with the term peptide hydrolases (Mancheko, 1994) are proteolytic enzymes that attack the peptide bonds of protein molecules forming small peptides. They can be classified into 4 groups based on the nature of amino acids in the active site of the enzyme. These groups are serine proteases, metallo (neutral)-proteases, thiol (cysteine) proteases and acid (aspartic) proteases. Serine proteases possess the catalytic triad amino acids, His, Asp and Ser and their activities are inhibited by phenylmethylsulfonylfluoride (PMSF) and diisopropyl fluoro phosphate (DFP). Cysteine proteases have a catalytic dyad composed of Cys and His residues and these enzymes are sensitive towards some oxidizing and alkalizing agents. The activity of metallo-protease is inhibited by EDTA due to the presence of a divalent cation, such as Zn ion (Zn^{2+}) at the catalytic site. Meanwhile, an acid protease is composed of a catalytic dyad made of two Asp residues and is not sensitive towards EDTA or serine protease inhibitors (Doughter and Semler, 1993; Suhartono *et al.*, 1997).

Proteases can be derived from various sources. Traditionally, proteases are obtained from plants and animals. However, due to certain advantages such as ease of cultivation, microorganisms are becoming a major source of these industrially important enzymes (Gacesa and Hubble, 1989; Taylor and Leach, 1995). The genus *Bacillus* was reported as the profilic producer of extracellular proteases (Nishiya and Imanaka, 1990; MacIver *et al.*, 1994; Fujita *et al.*, 1995). The major extracelullar proteases from this genus are serine proteases and neutral proteases (Van der Laan *et al.*, 1991; Fujita *et al.*, 1995).

Proteases are important both physiologically and commercially (MacIver *et al.*, 1994). They play important roles in some biochemical reactions such as degradation of proteins into amino acids and peptides for nutrients, formation of spores and germination, coagulation cascade reactions, pathogen mechanisms, modulation of gene expression, enzyme modification and *secretion of various* proteins.

From the economic point of view, microbial proteases are one of the commercial enzymes, which have found wide application in various industrial processes such as detergent making, leather production, meat tenderizing, baking and brewing. Bacterial proteases are reported to produce 20% of the total industrial enzymes (Fujita *et al.*, 1995). Serine or alkaline proteases have been used in washing detergents on a large scale for some 15 years (Priest, 1984). Subtilisin, the second largest family of serine proteases (Barret and Rawling, 1994), was claimed to be a major component of enzyme complement in biological washing powder (Cowan, 1996).

Thermostability is one of the significant requirement for commercial enzymes since thermal denaturation is a common cause of enzyme inactivation. Efforts have been made to improve the thermostability of these enzymes. An alternative method of obtaining these enzymes with improved thermostability is to isolate these enzymes from naturally occurring thermophilic organisms. However, this approach to produce large quantities of enzymes from thermophiles is often impractical, as the yield may be low due to imprecise growth conditions and the high fermentation temperature may need specialized equipment which in turn may increase production cost. Therefore, the preferred method is to isolate, clone and express the thermophilic genes of interest from these thermophilic organisms in mesophilic host(s).

Using the conventional gene cloning methods, Enequist *et al.* (1991) have cloned and expressed thermostable neutral protease gene from *B. caldolyticus* into *B. subtilis*. In addition, MacIver *et al.* (1994) and Takami *et al.* (1992) have cloned thermostable alkaline proteases from *Bacillus* sp. Ak1 and *Bacillus* sp. no AH-101 into *E. coli*. However these conventional methods sometimes suffer from low efficiency and extensive manipulations.

An alternative cloning method involves the use of the polymerase chain reaction (PCR). In this method, a particular DNA segment can be specifically amplified *in vitro*. The efficiency of cloning can be increased by incorporating sequences for the creation of restriction sites to the primers as well as by direct insertion of a PCR fragment into a specific vector with a T-overhang. In this study, cloning has been performed by direct insertion of the PCR product into such vector.

B. stearothermophilus strain F1 is a thermophilic bacteria, which was isolated from decomposed oil palm branch. It was documented to produce the thermostable alkaline protease (Rahman *et al.*, 1994). Detergency studies on this protease showed that it was better than Savinase, which was used commercially in detergent industry (Gunasekaran, 1995). Thus, it has potential to be used as a detergent enzyme.

With respect to the potential in the industrial use of this enzyme, this research was carried out with the ultimate goal of enhancing protease production as well as maintaining their thermostability. Thus, this research was undertaken with the following objectives:

1. to clone and express the alkaline protease gene from *B. stearothermophilus* F1 in a mesophilic host;
2. to determine the nucleotide sequences of the protease gene;
3. to analyse and compare their nucleotide and predicted amino acid sequences to related proteases; and
4. to partially purify and characterise the recombinant enzyme.