



UNIVERSITI PUTRA MALAYSIA

**CLONING, SEQUENCING AND EXTRACELLULAR
EXPRESSION OF THE ALKALIN E PROTEASE GENE FROM
BACILLUS STEAROTHERMOPHILUS F1**

FU ZHIBIAO

FSAS 2001 6

**CLONING, SEQUENCING AND EXTRACELLULAR
EXPRESSION OF THE ALKALINE PROTEASE GENE FROM *BACILLUS
STEAROTHERMOPHILUS* F1**

By

FU ZHIBIAO

**Thesis Submitted in Fulfilment of the Requirement for the Degree of
Master of Science in the Faculty of Science and Environmental Studies
Universiti Putra Malaysia**

July 2001



Abstract of thesis submitted to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

**CLONING, SEQUENCING AND EXTRACELLULAR EXPRESSION
OF THE ALKALINE PROTEASE GENE FROM *BACILLUS
STEAROTHERMOPHILUS* F1**

By

FU ZHIBIAO

July 2001

Chairman : Associate Professor Che Nyonya Abdul Razak, Ph.D.

Faculty : Science and Environmental Studies

The gene of a highly thermostable alkaline protease was amplified from *Bacillus stearothermophilus* F1 by polymerase chain reaction using consensus primers based on the sequences of serine protease genes from related species. Nucleotide sequence analysis of the gene revealed an open reading frame containing 1,206 bp which encodes for a polypeptide of 401 amino acid residues. The polypeptide composed of a signal peptide (25 amino acid residues), a prosequence (97 amino acid residues), and a mature protein of 279 amino acid residues. Amino acid sequence comparison revealed that it shared high homology with those of other serine proteases from a number of *Bacillus* spp. The recombinant F1 protease was efficiently excreted into culture medium using *E. coli* XL1-Blue harbouring two vectors: pTrcHis bearing the protease gene and pJL3 containing the bacteriocin-release-protein (BRP). Both vectors contain the *E. coli lac* promoter-operator system. In the presence of 40 μ M isopropyl- β -D-



thiogalactopyranoside (IPTG), the recombinant F1 protease and the BRP were expressed and the mature F1 protease was released into the culture medium. The enzyme was purified through a one-step heat treatment at 70°C for 3 h, and this method purified the protease to near homogeneity. The purified enzyme showed a pH optimum of 9.0, temperature optimum of 80°C, and was stable at 70°C for 24 h in pH ranges from 8.0 to 10.0. The enzyme exhibited a high degree of thermostability with half-lives of 3.5 h at 85°C, 25 min at 90°C, and was inhibited by the serine protease inhibitor, phenylmethanesulfonyl fluoride (PMSF).



Abstrak tesis yang dikemukakan kepada senat Universiti Putra Malaysia
sebagai memenuhi keperluan Ijazah Master Sains

**PENKLOANAN, PENJUJUKAN DAN PENGEKSPRESAN
EKSTRASELULAR GEN PROTEASE ALKALI DARIPADA
BACILLUS STEAROTHERMOPHILUS F1**

Oleh

FU ZHIBIAO

Julai 2001

Pengerusi : Prof. Madya Che Nyonya Abdul Razak, Ph.D.

Fakulti : Sains Dan Pengajian Alam Sekitar

Gen protease alkali yang termostabil daripada *Bacillus stearothermophilus* F1 telah digandakan dengan tindak balas berantai polimerase dengan kehadiran primer; primer yang berdasarkan jujukan gen protease serine daripada spesies berkaitan. Analisis jujukan menunjukkan rangka bacaan terbuka bersaiz 1206 bp yang mengkodkan polipeptida yang terdiri daripada 401 residu asid amino. Polipeptida tersebut terdiri daripada peptida isyarat (25 residu asid amino), satu pro-jujukan (97 residu asid amino) dan satu protein matang yang terdiri daripada 279 residu asid amino. Perbandingan asid amino menunjukkan homology yang tinggi dengan protease serina daripada beberapa spesies *Bacillus*. Protease daripada F1 rekombinan telah berjaya dirembeskan secara efisien ke dalam medium kultur menggunakan *E. coli* XL1-Blue yang membawa dua vector: pTrcHis yang mengandungi gen protease dan vektor pJL3 yang membawa gen protein-pembebasan-bakteriosin (BRP). Kedua-dua vektor mengandungi sistem operator promoter lac *E. coli*. Dengan kehadiran 40 μ M IPTG,



protease F1 rekombinan dan BRP diekspres dan protease matang F1 dirembeskan ke dalam medium kultur. Enzim dituliskan melalui kaedah pemanasan satu langkah pada 70°C selama 3 jam, kaedah ini berjaya menuliskan protease hampir mencapai homogeneiti. Enzim tulen menunjukkan pH optimum pada 9.0, suhu optimum pada 80°C dan stabil pada 70°C selama 24 jam di dalam pH antara 8.0 hingga 10.0. Enzim ini menunjukkan darjah kestabilan yang tinggi dengan hayat separuh 3.5 jam dan 25 minit pada suhu 85°C dan 90°C masing-masing. Enzim ini direncat oleh perencat protease serina, penilmetanasulfonil florida (PMSF).

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to the chairman of my supervisory committee, Associate Professor Dr. Che Nyonya Abdul Razak, for giving me the opportunity to work on this very interesting and practical project, and also for her invaluable guidance, encouragement, and constant support during the period of this study. My true appreciation is also extended to my co-supervisors, Prof. Dr. Abu Bakar Salleh, Associate Prof. Dr. Mahiran Basri and Dr. Raja Nor Zaliha Rahman for their indispensable supervision. My sincerest thanks go to other lecturers in the Department of Biochemistry and Microbiology for their help.

I am also grateful to all staff and graduate students in the Department of Biochemistry and Microbiology, especially Ain, Sim, Leow, Chin, Goh, Lee, Shidah, Leha, Ee Lin, Laith, Nazleen, Bimo, Maji, Swee Tin, Ko Lian, Tang, Wai Kean and Kiu Lin for their help.

Finally, my deepest gratitude goes to my wife, Xu Beida and my family for their love and support.



TABLE OF CONTENTS

		Page
ABSTRACT		i
ABSTRAK		iii
ACKNOWLEDGEMENTS		v
APPROVAL SHEETS		vi
DECLARATION FORM		viii
TABLE OF CONTENTS		ix
LIST OF TABLES		xii
LIST OF FIGURES		xiii
LIST OF ABBREVIATIONS		xv
CHAPTER		
I	INTRODUCTION	1
II	LITERATURE REVIEW	5
	Protease	5
	Introduction	5
	Classification of proteases	5
	Serine proteases	6
	Cystein proteases	7
	Aspartic proteases	8
	Metallo-proteases	8
	Microbial proteases	9
	Thermostable protease from <i>B. stearothermophilus</i> F1	12
	PCR Amplification of the Protease Gene from <i>B. stearothermophilus</i> F1	13
	Purification of Recombinant Protease	14
	Applications of proteases	15
	Production of recombinant proteins	16
	Protein expression in <i>E. coli</i>	18
	Protein expression in Baculovirus	21
	Protein expression in Yeast- <i>pichia pastis</i>	22
	Heterologous proteins targeting in <i>E. coli</i> system	23
	Cytoplasmic expression	23
	Periplasmic expression	24
	Extracellular secretion	25
	Bacteriocin-release-protein (BRP) induced release by <i>E. coli</i>	27
III	MATERIALS AND METHODS	30
	Materials	30
	Methods	32
	Microorganism and culture condition	32
	Genomic DNA extractin	32
	Quantitation of genomic DNA	33
	PCR amplification of protease gene	34



Purification of the PCR product	35
Cloning of PCR product into pCR2.1	35
Screening of the positive clones	36
Single colony PCR	36
Streaking on SMA plate	36
Plasmid extraction and DNA sequencing	36
Cloning and expression of the protease gene in expression vector	37
Assay of protease activity	38
Assay of extracellular and total protease activity	39
Co-transformation of <i>E. coli</i> strains with pJL3 and recombinant pTrcHis	39
Optimization of bacteriocin-release-protein (BRP) induced release	41
Effect of BRP on cell growth--Quasi-lysis test	41
Effect of host strain on protease release	42
Effect of IPTG on protease release	42
Effect of culture medium on protease release	42
Effect of temperature on protease release	42
SDS-PAGE analysis of bacteria proteins	43
Detection of proteolytic activity	44
Purification of recombinant protease	44
Crude enzyme preparation	44
Heat-treatment	45
Characterization of recombinant protease	45
Effect of inhibitors on enzyme activity	45
Effect of pH on enzyme activity	45
pH stability	46
Effect of temperature on enzyme activity	46
Thermostability	46
IV RESULTS AND DISSCUSSION	47
Growth of <i>Bacillus stearothermophilus</i> F1	47
Genomic DNA extraction	47
PCR amplification of the protease gene	50
Cloning of the PCR product	50
Nucleotide sequence analysis of the protease gene	55
Deduced amino acid sequence analysis	57
Expression of F1 protease in <i>E. coli</i>	61
Extracellular expression of F1 protease gene in <i>E. coli</i> with BRP System	62
Quasi-lysis test	62
Effect of host strain on protease release	64
Effect of IPTG on protease release.	65
Effect of different medium on protease release	65
Effect of temperature on protease release	67



SDS-PAGE analysis of bacterial proteins	68
Purification of the recombinant F1 protease through a single step	69
Native discontinuous PAGE and detection of proteolytic activity	71
Characterization of recombinant F1 protease	74
Effect of inhibitors	74
Effect of pH on protease activity	74
Effect of pH on protease stability	77
Effect of temperature on protease activity	79
Thermostability	79
V CONCLUSION	82
BIBLIOGRAPHY	86
VITA	95
APPENDICES	
Appendix A: <i>Escherichia coli</i> strains	
Appendix B: Liquid and solid media	



LIST OF TABLES

Table		Page
1	Comparison of recombinant protein expression systems	17
2	Post-translational processing of recombinant proteins in major expression systems	17
3	Preference for expression of different recombinant genes	18
4	Extracellular F1 protease activity with various <i>E. coli</i> strains at 5 and 24 h of induction	64
5	Total and extracellular F1 protease activity with various IPTG concentrations at 10, 24, and 30 h induction in <i>E. coli</i> XL1-Blue in LB medium	66
6	OD ₆₆₀ of the culture medium after 10 h and 24 h induction with various IPTG concentrations	66
7	Extracellular F1 protease activity in various broth at 5 and 24 h after induction with 40 μ M IPTG in <i>E. coli</i> XL1-Blue	67
8	Effect of cultivation temperature on the total and extracellular production of F1 protease in <i>E. coli</i> XL1-Blue in LB after induction with 40 μ M IPTG for 24 h	68



LIST OF FIGURES

Figures	Page
1. Bacteriocin-release-protein (BRP) expression plasmid pJL3 and recombinant F1 protease secretion vector pTrcHis	40
2. Production of protease by <i>B. stearrowthermophilus</i> F1	48
3. Genomic DNA of <i>B. stearrowthermophilus</i> F1	49
4. PCR product of 1.5 kb amplified using primers FOR 900 and REV 1591	51
5. Single colony PCR detecting the recombinant cells	53
6. Clearing haloes of <i>E. coli</i> TOP 10 harbouring protease gene on skim milk agar	54
7. Nucleotide sequence of the alkaline protease gene from <i>B. stearrowthermophilus</i> F1 and the deduced amino acid sequence	56
8. Homology lineup of alkaline serine protease from several bacteria species	60
9. Quasi-lysis of <i>E. coli</i> XL1-Blue cells harbouring two plasmids (pJL3 and recombinant pTrcHis)	63
10. SDS-PAGE analysis of F1 protease excretion by the <i>E. coli</i> XL1-Blue cells harbouring the two plasmids (pJL3 and recombinant pTrcHis)	70
11. SDS-12% polyacrylamide gel electrophoresis of the purified protease	72
12. Electrophoresis in 10% polyacrylamide gel under non-denaturing conditions	73
13. Effect of inhibitors on protease activity	75
14. Effect of pH on enzyme activity	76
15. Effect of pH on enzyme stability	78



16. Effect of temperature on protease activity	80
17. Effect of different temperatures on enzyme stability	81



LIST OF ABBREVIATIONS

A	adenine base nucleotide
APMSF	(4-aminophenyl)-methanesulfonyl fluoride
bp	base pair
C	cytosine base nucleotide
dH ₂ O	distilled water
DNA	deoxyribonucleic acid
EDTA	ethylenediamine tetra acid
G	guanine base nucleotide
h	hour
kDa	kilodalton
kb	kilo base pair
M	Molar
mg	milligram
ml	milliliter
mM	millimolar
min	minute
PMSF	Phenylmethanesulfonyl fluoride
PCMB	p-Chloromercuribenzoic Acid
TLCK	L-1-Chloro-3-(4-tosylamino)-7-amino-2 heptanone- hydrochloride
TCA	trichloroacetic acid
μg	microgram



μ l	microliter
ng	nanogram
nm	nanometer
ORF	open reading frame
sec	second
SDS	sodium dodecyl sulfate
T	thymine base nucleotide
v/v	volume per volume
w/v	weight per volume

CHAPTER I

INTRODUCTION

Proteases are important group of enzymes both academically and commercially. Microbial proteases dominate the commercial applications, with a large market share taken by subtilisin proteases from *Bacillus* spp. for laundry detergent applications (Ward, 1983). A major requirement for commercial applications of enzyme is thermal stability, because thermal denaturation is a common cause of enzyme inactivation. And there have been a number of recent efforts to improve the thermostability of the enzymes on the basis of the currently limited knowledge of protein engineering (Suzuki *et al.*, 1989).

An alternative method for obtaining enzymes with improved thermostability is to isolate enzyme from naturally occurring thermophilic organisms. However, a disadvantage of this approach is that it is impractical to produce large quantities of enzymes from such organisms, as yield may be low because of imprecise growth conditions. Furthermore, high-temperature fermentations may require special equipment (Sonnleitner and Fiechter, 1983). Therefore, the preferred method is to use gene technology to clone and express the thermophilic genes of interest into mesophilic organisms.



Microbial proteases have been divided into four groups based on their mechanisms of action. These groups are serine proteases, metallo (neutral)-proteases, thiol (cysteine) proteases and acid (aspartic) proteases. Serine proteases are characterized by the presence of a serine group in their active site, and recognized by their irreversible inhibition by PMSF. Serine proteases are generally active at neutral and alkaline pH, with optimum pH between 7 and 11. Metallo-proteases are characterized by the requirement for a divalent metal ion for their activity. They are inhibited by chelating agents such as EDTA but not by sulfhydryl agents or DFP. Thiol proteases depend on a catalytic triad consisting of cysteine and histidine for their activity. They are susceptible to sulfhydryl agents such as PCMB, but are unaffected by DFP and metal-chelating agents. While aspartic proteases depend on aspartic acid residues for their catalytic activity, and show maximal activity at low pH 3 to 4 (Rao *et al.*, 1998).

The extracellular proteases of the genus bacillus are mainly either serine (alkaline) or metal (neutral) enzymes (Priest, 1977). Neutral and alkaline proteases from *Bacillus* spp. have been cloned and expressed in the mesophilic host, either *Bacillus subtilis* or *E. coli*. Metallo-(neutral) proteases from *Bacillus* genus are widely distributed secretory enzymes, and the metallo-protease genes have been cloned and sequenced (Fuji *et al.*, 1983; Vasantha *et al.*, 1984; Takagi *et al.*, 1985; Kubo and Imanaka, 1988; Saul *et al.*, 1996). 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*. The serine protease genes from *B.*



amyloliquefaciens, *B. licheniformis*, *B. subtilis*, and the alkalophilic *Bacillus* spp. strain YaB have also been cloned and sequenced (Stahl and Ferrari, 1984; Jacobs *et al.*, 1985; Yoshimoto *et al.*, 1988; Kaneko *et al.*, 1989;).

However, there have been fewer reports on the release of recombinant protease into the culture medium of *E. coli*. There are several advantages to a system that releases proteins in a regulated manner into the culture medium. Among them are, purification of the protein of interest is simplified, the culture medium provides a larger space for accumulation of the protein, and release of the protein will not result in cell death or lysis as often occurs in high-level cytoplasmic production of recombinant proteins (Hsiung *et al.*, 1989).

Current study focuses on the use of bacteriocin-release-protein (BRP) in the release of recombinant proteins. BRP is a small lipoprotein with 28 amino acid residues which is produced as a precursor with a signal peptide and then processed and translocated to the outside membrane of the cell (Luirink *et al.*, 1986). Controlled expression of the BRP has been used for the release of heterologous proteins from *E. coli* into the culture medium. Examples of proteins released by this system are the human growth hormone (Kato *et al.*, 1987), human immunoglobulin G-Fc (Kitai *et al.*, 1988) and the human growth hormone targeted by the OmpA signal peptide (Hsiung *et al.*, 1989). In this study, BRP system was used to release the recombinant protease from *B. stearothermophilus* F1 into the *E. coli* culture medium.

A wide variety of *Bacillus* species secrete serine proteases into the external medium. *Bacillus* serine proteases have their best-known application in detergent powders. To best meet the alkaline conditions in detergents, serine proteases with a highly alkaline pH optimum are preferred. Rahman *et al.* (1994) isolated an alkaline thermophilic *Bacillus* strain that produced a thermostable alkaline serine protease, with an optimal temperature at 85°C and an optimal pH value of 9.0. It has half-lives of 3.5 h at 85°C, 24 min at 90°C, and is stable between pH values 7.0 - 10 for 24 h at 70°C without loss of activity.

Hence, this research was carried out with the following objectives:

1. To clone the protease gene from *B. stearothermophilus* F1 and determine its nucleotide sequence.
2. To extracellularly express the alkaline protease from *B. stearothermophilus* F1 into a mesophilic host.
3. To partially characterize the recombinant protease.

CHAPTER II

LITERATURE REVIEW

Proteases

Introduction

Proteases are degradative enzymes, which catalyze the total hydrolysis of proteins. They are an important group of enzymes both physiologically and commercially. The vast diversity of proteases, in contrast to the specificity of the action, has attracted more and more attention in attempts to exploit their physiological and biotechnological applications (Rao *et al.*, 1998).

Classification of Proteases

Currently, proteases are classified on the basis of three major criteria: (1) type of reaction catalyzed, (2) chemical nature of the catalytic site, and (3) evolutionary relationship with reference to structure (Barett, 1994). Proteases are grossly subdivided into two major groups, i.e. exopeptidases and endopeptidases, depending on their site of action. Based on the functional group present at the active site, proteases are further classified into four prominent groups; serine proteases, aspartic proteases, cysteine proteases, and metallo-proteases (Hartley, 1960).

Serine Proteases Serine proteases are characterized by the presence of a serine group site in their active site, usually a Ser-His-Asp catalytic triad. Serine proteases usually follow a two-step reaction for hydrolysis in which a covalently linked enzyme-peptide intermediate is formed with the loss of the amino acid or peptide fragment (Fastrez and Fersht, 1973). Based on their structure similarities, serine proteases have been grouped into 20 families, which have been further subdivided into about six clans with common ancestors (Barett, 1994). Another interesting feature of the serine proteases is the conservation of glycine residues in the vicinity of the catalytic serine residue to form the motif Gly-Xaa-Ser-Yaa-Gly (Brenner, 1988).

Serine proteases are recognized by their irreversible inhibition by 3,4-dichloroisocoumarin (3,4-DCI), L-3-carboxytrans 2,3-epoxypropyl-leucylamido (4-guanidine) butane (E.64), diisopropylfluorophosphate (DFP), phenylmethylsulfonyl fluoride (PMSF) and tosyl-L-lysine chloromethyl ketone (TLCK). Some of the serine proteases are inhibited by thiol reagents such as *p*-chloromercuribenzoate (PCMB) due to the presence of a cysteine residue near the active site. Serine proteases are generally active at neutral and alkaline pH, with an optimum pH value between 7 and 11. They have broad substrate specificities including esterolytic and amidase activity. Their molecular masses range between 18 and 35 kDa, while the serine protease from *Blakeslea trispora* has a molecular mass of 126 kDa (Govind *et al.*, 1981). The isoelectric points of serine proteases are generally between pH 4 and 6. Serine alkaline proteases that are active at highly alkaline pH represent the largest subgroup of serine proteases. The alkaline serine

proteases possess wide specificity to the type of peptide bond, and hydrolyse a wide range of proteins. The enzyme hydrolysed more than 70% of peptide bonds in the substrate molecule (Barett, 1994).

The optimal pH of serine alkaline proteases is around pH 10, and their isoelectric point is around pH 9. Their molecular masses are in the range of 15 to 30 kDa. Although alkaline serine proteases are produced by several bacteria such as *Arthrobacter*, *Streptomyces*, and *Flavobacterium* spp. (Boguslawski *et al.*, 1983), subtilisins produced by *Bacillus* spp remain the best known. Two different types of alkaline proteases, subtilisin Carlsberg and subtilisin Novo, have been identified (Rao *et al.*, 1998). Both subtilisins have a molecular mass of 27.5 k Da, but differ from each other by 58 amino acids. They have similar properties such as an optimal temperature of 60°C and an optimal pH of 10. Both enzymes exhibit a broad substrate specificity and have an active site triad made up of Ser21, His64, and Asp32.

Cysteine Proteases About 20 families of cysteine proteases have been recognized. The activity of all cysteine proteases depends on a catalytic triad consisting of cysteine and histidine. The order of Cys and His (Cys-His or His-Cys) residues differs among the families (Barett, 1994). Generally, cysteine proteases are active only in the presence of reducing agents such as cysteine.

Cysteine proteases catalyze the hydrolysis of carboxylic acid derivatives through a double-displacement pathway involving general acid-base

