

**EXPRESSION OF RECOMBINANT ACID PROTEASE (THERMOP SIN)
GENE FROM THERMOPLASMA VOLCANIUM**

**A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE
OF
MIDDLE EAST TECHNICAL UNIVERSITY**

BY

BİLSEV KOYUNCU

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY**

JANUARY 2006

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Canan ÖZGEN

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science

Prof. Dr. Semra KOCABIYIK

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science

Prof. Dr. Semra KOCABIYIK

Supervisor

Examining Committee Members

Assoc. Prof. Dr. Candan GÜRAKAN (METU, FDE) _____

Prof. Dr. Semra KOCABIYIK (METU, BIOL) _____

Assoc. Prof. Dr. Meral KENCE (METU, BIOL) _____

Assoc. Prof. Dr. Fatih İZGÜ (METU, BIOL) _____

Assist. Prof. Dr. Hasan KOYUN (Yüzüncü Yıl Unv.) _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Bilsev KOYUNCU:

Signature :

ABSTRACT

EXPRESSION OF RECOMBINANT ACID PROTEASE (THERMOPSIN) GENE FROM THERMOPLASMA VOLCANIUM

KOYUNCU, Bilsev

M.Sc., Department of Biology

Supervisor: Prof. Dr. Semra KOCABIYIK

January, 2006, 113 pages

Acid proteases, commonly known as aspartic proteases are degradative enzymes which catalyze the cleavage reaction of peptide bonds in proteins with a pH optimum in the acidic range (pH 3-4). Acid proteases have crucial roles in metabolism. Moreover, they are used in different fields of industry.

Thermophilic microorganisms, especially archaea, gain special interest because of their thermal stability for both fundamental and industrial researches.

Thermopsin is an extracellular acid protease and a member of A5 family of proteases. This thermophilic enzyme has no characteristic active aspartyl residue, is insensitive to pepstatin and no apparent sequence homology to other acid proteases and therefore represents a new class of acid proteases. Thermophilic archaeal strain *Thermoplasma volcanium* GSS1 (optimum temperature 55⁰C and pH 2.7) in the genome has a putative thermopsin gene encoding 998 amino acid enzyme.

In this study thermopsin gene from *Thermoplasma volcanium* was expressed in *E. coli* as fusion with 6xHis tag under the control of T5 transcription/translation system. Putative thermopsin gene from *Thermoplasma volcanium* was amplified by PCR method using two primer sets and cloned. A 3080 bp and a 3070bp PCR products were obtained by using TP1/TP2 primer set (thermopsin gene with the start codon) and TP1'/TP2 primer set (thermopsin gene missing start codon) respectively.

PCR amplified thermopsin genes pDrive and pUC18 vectors in *E. coli* TG1 were cloned using and then cloned genes were sub-cloned directionally into pQE triple vector set for expression. In these expression vectors, cloned genes are placed downstream of a 6XHis tag to produce an expression fusion. *E.coli* strains (M15[pREP4], SG13009[pREP4], and TG1) used as hosts.

Recombinant colonies screened by colony blot/hybridization method based on immunological detection of the expressed 6XHis tag fusion by Anti-His HRP conjugates which are specific for 6xHis tag, and DAB chromogenic substrate was used for colony blot procedure.

PCR amplified thermopsin gene containing 3080bp could not expressed in pQE30 and 31 vectors in TG1 strains. It is thought that pQE32 open reading frame can be true for thermopsin gene (3080bp).

Three expression constructs, pQE31-1, pQE31-4 and pQE31-6 plasmids containing PCR amplified 3070bp thermopsin gene were confirmed as true recombinant plasmids according to both colony blot hybridization result and restriction digestion profile the agarose gel.

Keywords: Expression, 6XHis tag, acid proteases, thermopsin, *Thermoplasma volcanium*.

ÖZ

THERMOPLASMA VOLCANIUM REKOMBİNANT ASİT PROTEAZ (TERMOPSİN) GENİNİN EKSPRESYONU

KOYUNCU, Bilsev

Yüksek Lisans, Biyoloji Bölümü

Tez Yöneticisi: Prof. Dr. Semra KOCABIYIK

Ocak, 2006, 113 sayfa

Aspatik proteazlar olarakta bilinen asit proteazlar, proteinlerin peptit bağlarının yıkımı reaksiyonlarını katalizleyen hidrolitik enzimlerdir (optimum pH 3-4). Asit proteazlar metabolizmada çok önemli görevlere sahiptirler. Aynı zamanda endüstrinin farklı alanlarında da kullanılmaktadırlar.

Termofilik mikroorganizmalar, özellikle archaea, yüksek ısılarda kararlılıklarını koruyabildikleri için hem temel bilimsel araştırmalarda hem de endüstriyel araştırmalarda ilgi çekici hale gelmiştir.

Proteazların A5 ailesinin bir üyesi olan termopsin, hücre dışı bir asit proteazdır. Bu ısıya dayanıklı enzim; asit proteazlara özgü aspartil gruplarının eksikliği ve diğer asit proteazlarla bir homolojisinin bulunmaması nedeniyle yeni bir asit proteaz grubu olarak düşünülmüştür.

Optimum büyüme sıcaklığı 55⁰C ve pH 2.7 olan termofilik bir archaea olan *Thermoplasma volcanium* GSS1 suşu bir asit proteaz enzimi üretmektedir.

Bu çalışmada termopsin geni ekspresyonu; genin 6XHis kuyruğuyla birleşerek T5 transkripsiyon/translasyon sistemi kontrolü altında gerçekleştirilmiştir.

Thermoplasma volcanium asit proteaz geni PCR methodu ile olası termopsin genini kapsayan TP1/TP2 ve TP1'/TP2 primer setleri kullanılarak çoğaltılmıştır. PCR sonucu 3080bp ve 3070bp PCR ürünleri elde edilmiştir. TP1' primeri genin başlangıç kodonunu içermemektedir.

PCR ile çoğaltılmış termopsin genini içeren rekombinant pDrive ve pUC18 vektörleri, *SalI*, *SphI* v *HindIII* restriksiyon enzimleriyle, pQE ekspresyon vektörlerine genin yönlü olarak klonlanması için kullanılmıştır. *E. coli* M15[pREP4], SG130009[pREP4] ve TG1 suşları konukçu mikroorganizmalar olarak kullanılmıştır.

Rekombinant kolonilerin bulunması için immunolojik tarama temeline dayanan koloni hibridizasyon tekniği; 6XHis kuyruğuna özel Anti-His HRP konjugat ve renkli DAB substratı kullanılarak uygulanmıştır.

Çalışma sırasında PCR ile çoğaltılan 3080bp termopsin geni pQE30 ve 31 vektörlerinde eksprese edilememiştir. pQE32 vektörü ekspresyon çalışması için denenecektir.

PCR ile çoğaltılan 3070bp termopsin genini içeren rekombinant pQE31-1, pQE31-4 ve pQE31-6 vektörlerinde genin ekspresyonu koloni blot hibridizasyon tekniğiyle belirlenmiş ve agaroz gel elektroforezi ile onaylanmıştır.

Anahtar kelimeler: Ekspresyon, 6XHis, asit proteaz, termopsin, *Thermoplasma volcanium*.

To My Family

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor Prof. Dr Semra KOCABIYIK for her valuable guidance, critical discussions, and continued advice throughout this study.

I am also thankful to my labmate Burçak DEMİROK for her friendship and support.

Finally, I would like to express my sincere gratitude to my parents Münevver KOYUNCU and Bayram KOYUNCU, my sister Elçin KOYUNCU and my aunt Müyesser ÖZKURT for their help, permanent support and encouragement throughout this study.

TABLE OF CONTENTS

PLAGIARISM.....	iii
ABSTRACT.....	iv
ÖZ.....	vi
DEDICATION.....	viii
ACKNOWLEDGEMENTS.....	ix
TABLE OF CONTENTS.....	x
LIST OF FIGURES.....	xiii
LIST OF TABLES.....	xvi

CHAPTER

1. INTRODUCTION	1
1.1 Aspartic Proteases.....	4
1.1.1 Physiological Functions of Aspartic Proteases.....	5
1.1.2 Industrial Applications of Aspartic Proteases.....	7
1.1.3 Structure and Physical Properties of Aspartic Proteases.....	11
1.1.4 Action Mechanism of Aspartic Proteases	14
1.2 Microbial Aspartic Proteases.....	15
1.2.1 Fungal aspartic Proteases.....	15
1.2.2 Yeast aspartic Proteases.....	18
1.2.3 Bacterial Aspartic Proteases.....	19
1.2.4 Viral Aspartic Proteases.....	20
1.2.5Archaeal Aspartic Proteaeses.....	21
1.3 Expression of Aspartic Proteases	22
1.4 Scope and Aim of the Study.....	23
2 MATERIALS AND METHODS.....	25
2.1 Materials.....	25

2.1.1 Chemicals, Enzymes and Kits.....	25
2.1.2 Buffers and solutions.....	26
2.1.3 Plasmid vectors, Molecular size Markers and Genomic DNA Sequence Data.....	26
2.2 Strain and Medium.....	26
2.2.1 Bacterial and Archaeal Strains.....	26
2.2.2 Growth and Maintenance of Microorganisms.....	27
2.3 Gene Manipulation Methods.....	27
2.3.1 Genomic DNA Isolation from <i>Thermoplasma volcanium</i>	27
2.3.2 Purification of Chromosomal DNA of <i>Thermoplasma volcanium</i>	28
2.3.3. Spectrophotometric analysis of DNA.....	28
2.4 Agarose gel Electrophoresis.....	29
2.5. PCR Amplification of Thermopsin Gene of Putative Acid Proteases.....	29
2.5.1 Design of PCR Amplification Primers.....	29
2.5.2 PCR Amplification.....	31
2.6 Isolation of DNA Fragments from Agarose Gel.....	31
2.7 Cloning of PCR Amplified Thermopsin Gene.....	32
2.7.1 Cloning of PCR Amplified Thermopsin Gene With TP1/TP2 Primer Set.....	32
2.7.2 Cloning of PCR Amplified Thermopsin Gene with TP1'/TP2 Primer Set.....	33
2.7.3 Sub-cloning of Putative Thermopsin Gene for Expression.....	33
2.8. Introduction of Recombinant Plasmids into Competent Cells.....	35
2.8.1 Preparation of Competent <i>E. coli</i> strains.....	35
2.8.2 Transformation.....	37
2.9 Colony Blotting and Hybridization.....	37
2.10 Isolation of Plasmid DNA from Recombinant colonies.....	39
2.11 Computer analysis of Gene and Protein Sequence.....	41
3. RESULTS.....	42
3.1 Gene Manipulation.....	42
3.1.1 Genomic DNA Isolation of <i>Thermoplasma volcanium</i> .	42

3.1.2 Restriction Maps of 3080bp and 3070bp cloned	
Thermopsin Gene Fragments.....	43
3.1.3 PCR Amplifications and Cloning of PCR Amplified Fragments.....	45
3.1.3.1 Cloning of PCR Amplified Thermopsin Gene with TP1/TP2 Primer Set and selection of Recombinant Colonies.....	45
3.1.3.2 Characterization of Recombinant Plasmids by Restriction Digestion.....	51
3.1.3.3 PCR Amplification and Cloning of Thermopsin Gene with TP1'/TP2 Primer Set.....	55
3.1.3.4 Cloning of PCR Amplified 3070 bp Thermopsin Gene and Selection of Recombinants.....	55
3.1.3.5 Characterization of Recombinant Plasmids.....	59
3.1.4 Sub-cloning of Thermopsin Gene for Expression	62
3.1.4.1 Sub-cloning of Recombinant Thermopsin Gene from pDriveBSTP'-5 Plasmid to pQE Expression Vectors.....	65
3.1.4.1.1 Colony Blot Hybridization.....	68
3.1.4.2 Sub-cloning of Recombinant Thermopsin Gene from pUC18 Plasmid to pQE Expression Vectors.....	74
3.1.4.2.1 Colony Blot Hybridization.....	75
3.1.4.2.2 Characterization of Recombinant Plasmids.....	77
3.2 Computer Based Analysis of Thermopsin Gene.....	79
4. DISCUSSION.....	86
5. CONCLUSION.....	92
REFERENCES.....	93
APPENDICES.....	103
A. Buffers and Solutions.....	103
B. Plasmid Vectors.....	105
C. Molecular Size Marker.....	108
D. Nucleotide Sequence of Thermopsin.....	109
E. Multiple Sequence Aligment of Archaeal Thermopsin Genes.....	110

LIST OF FIGURES

FIGURES

Figure 1.1 Chymosin activity.....	9
Figure 1.2 Pepsin-like and retroviral aspartic protease structures.....	12
Figure 1.3 17 conserved water molecules in the crystal structure of Rhizopus-pepsin	13
Figure 1.4 Stereoview of firemen's grip of rhizopuspepsin.....	13
Figure 1.5 General acid-base catalysis of aspartic proteases.....	16
Figure 2.1 Calibration curve for DNA molecular size marker.....	30
Figure 2.2 Schematic representation of vectors.....	34
Figure 2.3 Schematic representation of cloning target sequence into QIAexpress vectors.....	36
Figure 2.4 The principle of color development in colony blot hybridization.....	40
Figure 2.5 Detection of positive expression clones by colony blotting hybridization.....	40
Figure 3.1 Genomic DNA of <i>Thermoplasma volcanium</i>	43
Figure 3.2 Schematic representation of pDrive SH/10/8 plasmid.....	45
Figure 3.3 Recombinant plasmid isolation.....	46
Figure 3.4 Restriction digestion of recombinant thermopsin gene.....	47
Figure 3.5 Restriction digestion profile of cloned thermopsin gene in pDrive vector.....	48
Figure 3.6 <i>SphI</i> and <i>SalI</i> digested pUC18 vector.....	49
Figure 3.7 Gel extraction of thermopsin gene fragment.....	50
Figure 3.8 Plasmid analysis by agarose gel electrophoresis.....	51
Figure 3.9 Schematic representation of thermopsin gene insertion into the pUC18 vector	53
Figure 3.10 Restriction digestion of recombinant pUC18-BST1-5 plasmid.....	54

Figure 3.11 Restriction digestion of recombinant pUC18-BST1-5.....	54
Figure 3.12 Schematic representation of restriction digestion of recombinant pUC18-BST1-5 plasmid	56
Figure 3.13 PCR amplification of thermopsin gene (3070bp).....	57
Figure 3.14 Schematic Representation of ligation of PCR amplified thermopsin...58	
Figure 3.15 Selection of true recombinant colonies.....	58
Figure 3.16 Restriction digestion of putative recombinant plasmids to confirm the cloning of thermopsin gene.....	60
Figure 3.17 Schematic representation restriction digestion of Recombinant pDrive vector.....	63
Figure 3.18 Restriction Digestion of recombinant plasmids.....	64
Figure 3.19 Schematic representation of pQE 30-31-32 expression vectors.....	65
Figure 3.20 Restriction digestion of pQE vectors with Sall for linearization.....	66
Figure 3.21 Agarose gel isolation of DNA fragments.....	66
Figure 3.22 Schematic representation of sub-cloning of thermopsin to pQE vector.....	67
Figure 3.23 Schematic representation of expression constructs.....	67
Figure 3.24 Colony blot hybridization results.....	68
Figure 3.25 Plasmid isolation	69
Figure 3.26 Plasmid digestion with SphI and Sall.....	69
Figure 3.27 Plasmid digestion.....	70
Figure 3.28 Plasmid digestion	70
Figure 3.29 Plasmid digestion with Sall for linearization.....	71
Figure 3.30 Seventh colony blot hybridization results	71
Figure 3.31 Plasmid isolation	72
Figure 3.32 Plasmid digestion with Sall.....	73
Figure 3.33 Plasmid digestion.....	73
Figure 3.34 HindIII cut recombinant plasmids.....	74
Figure 3.35 Agarose gel extraction of plasmids.....	75
Figure 3.36 Schematic representation of sub-cloning of thermopsin gene from pUC18 to pQE30-31 vectors	76
Figure 3.37 Colony blot hybridization results.....	77

Figure 3.38 Plasmid isolation.....	77
Figure 3.39 Plasmid digestion with HindIII.....	78
Figure 3.40 Plasmid digestion with Sall/HindIII.....	78
Figure 3.41 Restriction map of thermopsin structural gene.....	79
Figure 3.42 KEGG alignment data.....	81
Figure 3.43 Gene localization of Thermopsin in Different archaea species	82
Figure 3.44 Predicted secondary structure of Thermopsin gene.....	83
Figure 3.45 Phylogenetic tree	85
Figure B.1 pDrive vector.....	105
Figure B.2 pUC18.....	106
Figure B.3 pQE 30-31-32 vectors.....	107

LIST OF TABLES

TABLES

Table 2.1.	TP1/TP2 Primer set design.....	30
Table 2.2.	TP1'TP2 Primer set design.....	31
Table 3.1.	Restriction map of 3080bp cloned thermopsin gene.....	44
Table 3.2.	Restriction map of 3070bp cloned thermopsin gene.....	44
Table 3.3	The cut site positions of restriction enzymes in 3080bp PCR product of thermopsin gene.....	52
Table 3.4.	The cut site positions of some restriction enzymes in 3070bp PCR product.....	59
Table 3.5.	Cutter and non-cutter enzymes of thermopsin gene.....	80
Table 3.6	Scores of multiple sequence alignment of thermopsin genes.....	84

CHAPTER I

INTRODUCTION

Proteases which are also called proteinases or peptidases catalyze the peptide bonds hydrolysis reactions. These enzymes have important role in both physiological and industrial processes. They are involved in developmental progress like; cell growth and migration, tissue arrangement and morphogenesis. Protein catabolism, blood coagulation, inflammation, pharmacologically active peptides from precursor proteins, activation of zymogens, secretory protein transportation and release of hormones are some of essential and important roles of proteases. They can induce tumor growth and metastasis because of relationship with cell growth. Secretory proteases generally hydrolyze long peptide chains into smaller oligopeptides for routine absorption by cell while intracellular proteases play critical roles in the metabolic regulation of cell. Besides these physiological functions proteases have different industrial applications mainly detergent and food industry. Recently they are used in leather treatment and new generation pharmaceuticals. Proteases are also used for fundamental scientific researches such as structure-function relationships, peptide synthesis and protein sequencing (Rao *et al.*, 1998).

There are two major criteria for proteases classification: Firstly proteases are classified on the basis of action site. They are divided into two major groups, i.e., exopeptidases and endopeptidases. Exopeptidases cleave the peptide bond from the amino or carboxy termini of the protein and they are subdivided into two different groups: amino (N) proteases act at free N termini of the polypeptide chain, whereas carboxy (C) proteases act at C terminals of the polypeptide chain. Endopeptidases cleave the inner peptide bonds of proteins. Secondly, other proteases classification is based on their active site functional groups. According

to this principle (based on MEROPS peptidase database) proteases are classified into 4 groups, serine proteases, metallo proteases, cysteine proteases, and aspartic proteases. On the other hand, there are a few proteases which do not have special class characteristics so they do not fit into any class of proteases e.g., ATP-dependent proteases (Rao *et al.*, 1998).

Serine proteases are widespread enzymes among viruses, bacteria and eukaryotes. Their differential characteristic is the functional serine group found in the active site. There are 40 families of Serine proteases according to MEROPS peptidase database. Although they have different substrate specificity, they exhibit chemical and structural similarities. The catalytic triad, which is constructed by aspartate, histidine, and serine residues, is one of the common properties of serine proteases. Negatively charged group of aspartate (electrophile) enables the electron transfer through hydrogen bonds and the polarizable imidazole ring to the Serine oxygen so this oxygen becomes more powerful nucleophile (Hess., 1971). Histidine residue acts as a base into the catalytic triad of serine proteases. Another similarity of serine proteases is conserved glycine residues involved in the Gly-Xaa-Ser-Yaa-Gly motif around the catalytic triad (Rao *et al.*, 1998). Most of the serine proteases are active at neutral and alkaline pH (optimum pH between 7 and 11). They catalyze the hydrolysis of amide and ester bonds of proteins (Blow., 1971). Serine proteases can be inhibited by 3, 4-dichloroisocoumarin, phenylmethylsulfonyl fluoride (PMSF). Their molecular weights change between 18 and 35 kDa. Chymotrypsin, subtilisin, trypsin are some examples of the best known serine proteases. They are usually used in detergent and leather industries.

Metalloproteases are characterized by the requirement for a divalent metal ion for their activity. This type of proteases is produced by eukaryotes and bacteria. There are 54 different metalloprotease families according to MEROPS peptidase database. They can be grouped into different clans based on the amino acid which is in vicinity of the metal-binding site. Another classification based on the action specificity of metalloproteases. According to this classification metalloproteases can be divided into four major groups; neutral, alkaline, *Myxobacter I* and

Myxobacter II. EDTA inhibits all the metalloproteases because it holds the divalent cations. Their molecular weights change between 14 and 60 kDa. Metalloproteases are generally active at the alkaline pH values, i.e. pH 7 and 9. Collogenases which are isolated from higher organisms and bacterial thermolysin can be the representative enzymes of metalloproteases. Metalloproteases have the important role in tissue morphogenesis, differentiation and wound healing process. Furthermore they can be the new target for treatment of cancer and arthritis (Rao *et al.*, 1998).

Cysteine proteases are another group of proteases and they are found in both eukaryotes and prokaryotes. There are 59 cysteine proteases family according to MEROPS peptidase database. Cysteine proteases have catalytic dyad which is containing His and Cys residues into their active sites. They are also called cathepsins but cysteine cathepsins term is important to differentiate them from other serine or aspartic cathepsins. They are synthesized as zymogens and after maturation of enzymes their molecular mass are between 25-35 kDa (Berdowska., 2004). Generally HCN and cysteine are required for the activation of cysteine proteases. Except some lysosomal cysteine proteases, they usually have neutral pH optima. They are sensitive to sulfhydryl agents such as PCMB but they have resistance against the DFP and metal chelating agents. Papains, cathepsin B, L, H are the most known cysteine proteases (Rao *et al.*, 1998). They are involved in zymogens and pro-hormone activation, MHC-II mediated antigen presentation, bone remodeling, keratinocytes differentiation, hair follicle cycle, reproduction and apoptosis. Cysteine proteases can also be used as cancer markers (Berdowska., 2004). Additionally, they are known as senescence-specific proteases in *Arabidopsis sp* (Beers *et al.*, 2004).

Aspartic proteases are the last group of proteases. They are also called aspartyl or acid proteases. They are produced from eukaryotes and prokaryotes suggesting that acid proteases have vital roles in higher organisms and microorganisms. There are 12 acid protease families according to MEROPS peptidase database. They have aspartic acid residue at the catalytic center and they are active at acidic

pH conditions. Their molecular weights change between 30-45 kDa. They are usually susceptible to DAN and EPNP in the presence of copper ions. Pepstatin is also known acid protease inhibitor (Rao *et al.*, 1998). Acid proteases will be discussed in details following parts for being major subject of this thesis. Acid proteases' evolutionary development, biological functions, industrial importance, action mechanism and structure will be helpful for understanding the aim of the study.

1.1 Aspartic Proteases

Aspartic proteases are the most familiar groups among the proteases. They are found from higher organisms like human, and plants, to prokaryotes like bacteria, archaea and viruses. Aspartic proteases use a pair of aspartate residues to perform the protein cleavage reaction. Pepsin can be the representative enzyme for aspartic proteases. Pepsin was studied firstly by J. R. Green at 1898 and by E.H. Starling at 1906. At that time pepsin was popular subject because it was recognized to be the major catalytic agent of gastric digestion in mammals and birds. Furthermore pepsin was firstly characterized by T. Schwann (1836) and then E.von Brücke and C. A. Pikelharing purified pepsin (1861) (1896). These studies are building blocks of aspartic protease researches and they enabled the new inventions in this field. Nowadays aspartic proteases are classified into 12 families (A1-A2-A3-A5-A6-A8-A9-A11-A21-A22-A24-A26) according to MEROPS peptidase database. On the other hand aspartic proteases can be divided into two major groups; pepsin-like and retroviral enzymes. These enzymes can be isolated from five major sources. (i) Several types of gastric enzymes (pepsin, chymosin) can be isolated from stomach. (ii). Lysosomes are another source of aspartic proteases like Cathepsins. (iii). Renin can be isolated from kidney and sub-maxillary gland. (iv). Plants also produce aspartic proteases. (v) Lastly aspartic proteases can be isolated from microorganisms (Chitpinyol and Crabbe, 1997).

1.1.1 Physiological Functions of Aspartic Proteases

Aspartic proteases have different important physiological functions. The best studied aspartic protease is pepsin. Pepsin is gastric proteinase active at acidic pH values (pH 1-5) and formed by partial proteolysis of their inactive zymogen, the pepsinogen (Fruton, 1971). Gastric proteases catalyze hydrolysis of large proteins into smaller peptides and amino acids so they facilitate their absorption. So aspartic proteases are involved in the process of keeping cells alive (Rao *et al.*, 1998). A1 family members Pepsin A and Pepsin C, which are produced by *Homo sapiens*, are secreted into extracellular environment such as blood and stomach (Ostermann *et al.*, 2004).

Pepsin like enzymes resembles pepsin in several respects. One of the pepsin like protease is rennin (chymosin). Renin can be obtained from the fourth stomach of the calf. Renin has milk clotting activity and additionally regulates the blood pressure and like pepsin A and C rennin secreted into the extracellular environment (Fruton, 1971, Ostermann *et al.*, 2004).

Other pepsin-like enzymes are Cathepsin D and E. On the contrary cathepsin D and E remain intracellular and they are secreted into intracellular parts. They are synthesized as zymogene. Cathepsin D is a lysosomal protease and Cathepsin E is a component of endoplasmic reticulum or endosomal compartments associated vesicles (Ostermann, 2004). Generally they have important missions about protein turnover. Coopman *et al.* (2005) reviewed that Cathepsin D is involved in the cancer progression and metastasis. It also plays an important role in lysosomal cell death pathway. It is found that Cathepsin D is over-expressed from 2-to 50- fold in most breast cancer tumors compared to normal mammary glands (Coopman *et al.*, 2005). New researches reveal that procathepsin D acts as an autocrine mitogen in breast cancer cell-line (Fusek and Vetvicka, 1994). The second aspartic cathepsin is Cathepsin E. It has different physiological roles including the biogenesis of the vasoconstrictor peptide endothelin and the antigen/invariant chain processing. In addition it is thought that cathepsin E might have a role in

neurodegeneration associated with brain ischemia and aging (Ostermann *et al.*, 2004).

BACE (beta-site amyloid precursor protein converting enzyme) is a transmembrane aspartic protease and also called β -Secretase. It is involved in the generation of amyloid plaques from the amyloid precursor protein (APP) in Alzheimer's disease. BACE 2 is a homologous aspartic protease of BACE. Both β -Secretase and BACE 2 encoding genes are localized in the Down Syndrome (DS) critical region of chromosome 21. APP over-expression is thought to be responsible for senile plaques formation in DS patients that exhibit Alzheimer-like dementia. Over-expression of BACE 2 was shown in primary skin fibroblast derived from DS patients (Barbiero *et al.*, 2003; Citron, 2004).

In plants aspartic proteases play roles into different processes such as germination, differentiation, wound healing, senescence, programmed cell death and lastly defence system. CDR1 is an aspartic pepsin-like proteases and their over-expression causes disease resistance to *Pseudomonas sp.* Furthermore CDR1 plays a role in activation of wound related hormones (Hoorn and Jones, 2004). Plant senescence is based on an internally programmed degeneration leading to death. The research about broccoli florets reveals that aspartic proteases are found in post-harvesting stage of senescence (Wang *et al.*, 2004). Aspartic protease of cowpea (*Vigna unguiculata*) is detected in senescing leaves and stems but not in roots. The increased level of aspartic protease could be related to the remobilization of nitrogen from senescing tissues to reproductive sinks (Carvalho *et al.*, 2004). Palma *et al.* reviewed that these enzymes are required for germination and development of seeds. Cathepsin D-like aspartic-proteases are found in ungerminated seeds, during germination and in all living cells of barley grain. An aspartic protease has also induced systemic wound response protein in tomato leaves (Palma *et al.*, 2002). Both seeds and green malt contain aspartic proteases, they are required for dormant and germination stages (Jones, 2005). Carnivorous pitcher plants (*Nepenthes*) produce aspartic proteases within the digestive fluid of insect-trapping organs (Schaller, 2004).

Plasmepsins are the aspartic proteases of the malaria parasite; *Plasmodium falciparum* and other *Plasmodium sp.* Parasite degrades host hemoglobin to obtain free amino acids for protein synthesis so plasmepsins are essential for the survival of the parasite. There are different families of plasmepsin and plasmepsin 4 is localized into the food vacuole. This enzyme found in all *Plasmodium sp.* infecting man (Dame *et al.*, 2003).

Candida albicans is an opportunistic fungal pathogen that causes severe systemic infections in immunosuppressed individuals such as AIDS patients. The secreted aspartic proteases (SAPs) are one of the virulence factors of *Candida albicans*. SAPs are not found only *Candida albicans* but also found in *C. tropicalis*, *C. parapsilosis* and *C. guilliermondii* (Calderone and Fonzi, 2001).

Retroviral proteases (retropepsin) have been found in different viruses such as Human Immunodeficiency Virus, Rous Sarcoma Virus, Avian Myeloblastosis Virus, Simian Immunodeficiency Virus, Human T-cell Leukemia Virus type I (HTLV I). Proteolytic processing is essential for retrovirus life cycle and invasion of host cells (Shuker *et al.*, 2003).

1.1.2 Industrial Applications of Aspartic Proteases

Besides the crucial physiological functions, aspartic proteases are also very important for industrial applications.

The most important field of aspartic protease usage is cheese making industry. Aspartic proteases have milk clotting activity. The first indigenous protease in milk is reported in the end of the nineteenth century by Babcock and Russell.

The best known milk clotting enzyme is chymosin which is the synonym of rennin. This enzyme is firstly obtained from the fourth stomach of the calf and characterized by Foltmann in 1966 (Fruton, 1971). Chymosin exhibits a high ratio of milk clotting activity (MC) to proteolytic activity (PA). This ratio is also called

R factor. There are three major types of rennin used in cheese making industry. First one is rennin extracted from the abomasums of suckling ruminants; second one is rennin prepared from microbial sources, and the last is recombinant rennin (Poza *et al.*, 2004).

Aspartic proteases cleave the specific peptide bonds (Phe 105- Met 106 bond) to generate *para-κ-casein* and macropeptides since the initiation of cheese making process (Figure 1.1).

Increasing demand for cheese led to search for alternative milk coagulant sources researches, besides calf rennet. Microbial sources became quite attractive in recent years.

Different fungi can produce milk clotting aspartic proteases. *Endothia parasitica*, *Mucor miehei*, *Mucor pussilus*, *Mucor varians* var. *Pispek*, *Mucor mucedo*, *Rhizomucor miehei* are some examples of these fungi Among these *Mucor* species are preferred for cheese making due to their high MA and low PA. ‘Mucor rennins’ is a representative term of *Mucor sp.* aspartic proteases (Preetha and Boopathy, 1997; Fernandez-Lahore *et al.*, 1998).

Milk coagulating enzymes are also produced by *Bacillus subtilis* and *Bacillus cereus* (Arima *et al.*, 1970)

Pharmaceutical industry is another important field that aspartic proteases are used. It is known that some protozoa, most of the pathogenenic fungi and retroviruses require aspartic proteases for their life cycle and invasion strategies.

Malaria is thought as one of the most serious infectious disease around the world. It is estimated 40% of world population at risk of this infection. There are four major species of the malaria parasite that is *Plasmodium falciparum*, *P. vivax*, *P.malariae*, *P. ovale* but among these parasites *P. falciparum* is responsible for more than 95% of malaria related mortality and morbidity. *Plasmodium sp.* use hemoglobin as a nutrient. Hemoglobin is hydrolyzed by plasmepsin (aspartic

proteases) Today at least 10 aspartic protease genes are identified but two of them Plasmepsin I and II are studied in details. Studies show that plasmepsins can be the target of anti-malarial drugs. Pepstatin A is a natural inhibitor of aspartic proteases. It has hydroxyethylamine core and statine derived core and researchers try to create new inhibitors derived from pepstatin. Novel reverse-statine type isoesters (R, R)-III and (S, S)-III, where replacement of the N-terminus of the hydroxyethylamine core with a C-terminus isoester have shown promising activity against Plasmepsin I and II (Dahlgren *et al.*, 2003).

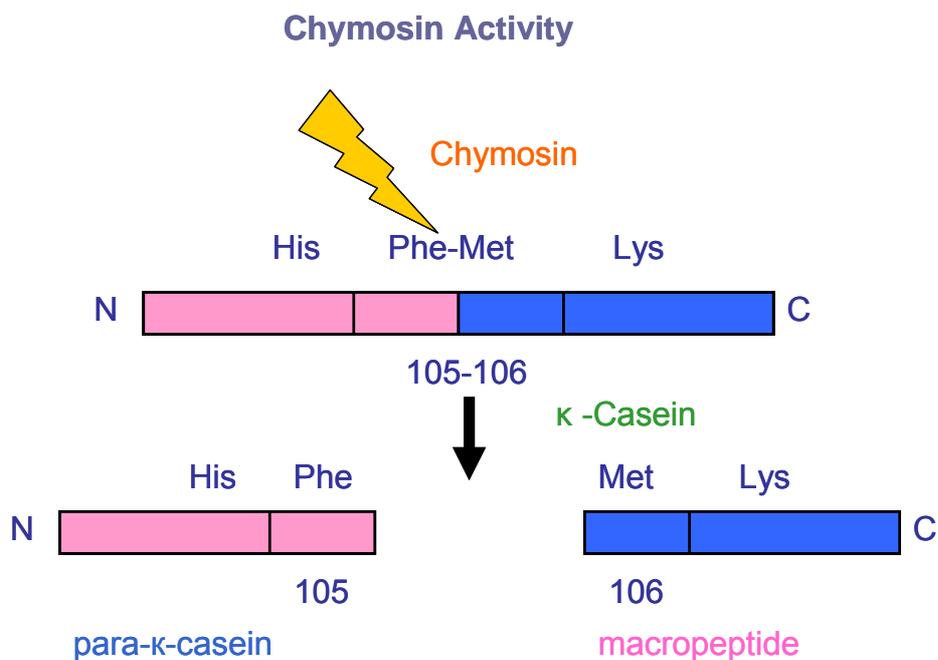


Figure 1.1: Chymosin activity. Chymosin cleaves the kappa casein’s peptide bond between Phe 105 and Met 106 and produces para-kappa casein and macropeptides.

Candida albicans causes systemic infections in immunosuppressed individuals and these infections increase morbidity and mortality. Secreted aspartic proteases (SAPs) are one of the virulence factors of *Candida* species, which became attractive targets for new generation drugs. After recognition of the SAPs being inhibited by Pepstatin A, Pepstatin A was chosen as a modal inhibitor (Backman and Danielson, 2003).

HTLV-I and HIV encode a virus specific aspartic protease (PR) responsible for processing the gag and gag-pro-pol poly proteins leading to proliferation of the retrovirus. HTLV-I PR is one of the major targets to develop the specific anti-HTLV-I agents. Nowadays new retroviral PR inhibitors are developed, named KNI compounds based on the concept of 'substrate transition-state mimic'. They inhibit the replication of HIV and HTLV-I. Experimental investigations are continuing (Maegawa *et al.*, 2004).

Aspartic proteases are also used in leather industry. Leather industry and environment can be described as the two sides of a coin. Harmful waste products are one of the disadvantages of traditional leather industry. To overcome this problem new enzyme based approaches are developed. Leather making processes have different stages and different classes of proteases are used for each stage. Aspartic proteases are used in post tanning and acid bating steps of leather making. *Aspergillus usami* is an ideal bating agent for sheep pelt. Pepsin extracted from bovine or porcine stomach can be added at pickling or after chrome tanning to give improved yield and soft leathers at low costs (Thanikaivelan *et al.*, 2004).

Additionally, aspartic proteases are widely used in genomic and proteomic studies because of their cleavage site specificity. Structure-function relationship studies allow for the construction of new enzymes with desired functions. The limited change in activities by introducing a new catalytic mechanism shows the possibility of converting a protease from one class to another, the e.g., pepsin

catalytic dyad is changed with serine protease's catalytic triad (Tanaka and Yada, 2004).

In conclusion aspartic proteases have a great potential to be used in different industrial fields especially cheese making, pharmaceutical and leather industries and furthermore they are important tools for genomic and proteomic studies.

1.1.3 Structure and Physical Properties of Aspartic Proteases

Aspartic proteases can be divided into two groups which are called pepsin like and retroviral proteases. Both pepsin-like and retroviral aspartic proteases use an Asp dyad to hydrolyze peptide bonds.

Pepsin-like proteases contain two asymmetric lobes called α and β monomers. Catalytic Asp dyad (Asp 32 and Asp 215, pepsin numbering) has been located at the lobe interface and a flap made up of a β -hairpin covering the peptide substrates (Casella *et al.*, 2004), (Figure 1.2.a).

On the other hand retroviral aspartic proteases have β homodimers. Catalytic Asp dyad (Asp 25 and Asp 25', HIV I protease numbering) is located on two loops at the monomer interface and where two β hairpins cover the active site (Dunn *et al.*, 2002), (Figure 1.2.b). One monomer contains four structural elements and; a hairpin, a wide loop which is carried on catalytic Asp residue, an α -helix and lastly second hairpin. Symmetric monomer is formed duplication of these four structural elements.

Besides the structural elements, water molecules are also integral part of aspartic proteases. They facilitate the stabilization of protein folding and play roles in function of enzymes. Protein hydration play crucial role in biological process. Although, there are 17 most common water molecules that are conserved in aspartic proteases, some of them have more important roles (Figure 1.3). Catalytic water molecule Wat 507 related with catalytic Asp residues. Two aspartates are

oriented toward each other around the pseudointerlobe diad axis in a complicated hydrogen-bonding network known as ‘firemen’s grip’. Central water molecule Wat 507 and two loops interaction is formed by the help of firemen’s grip (Chitpinyol and Crabbe, 1997). Other conserved water molecules Wat 513 and Wat 502 are placed in the catalytic site of aspartic proteases. They enable the rigidity of the active site geometry (Prasad and Suguna, 2002), (Figure 1.4).

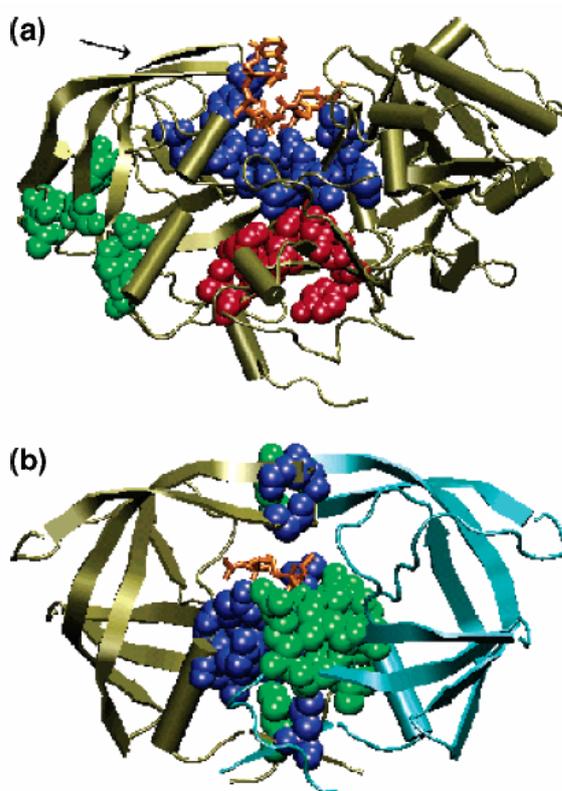


Figure 1.2: Pepsin-like (BACE) and retroviral aspartic protease (HIV I protease) structures are shown in complex with model substrate (orange sticks). (a); Blue color represents β hairpin, green color represents β monomer, and red color represents α monomer. Black arrow denotes the flap position. (b); Green color represents β monomers and blue color represents β hairpins (Cascella *et al.*, 2004).

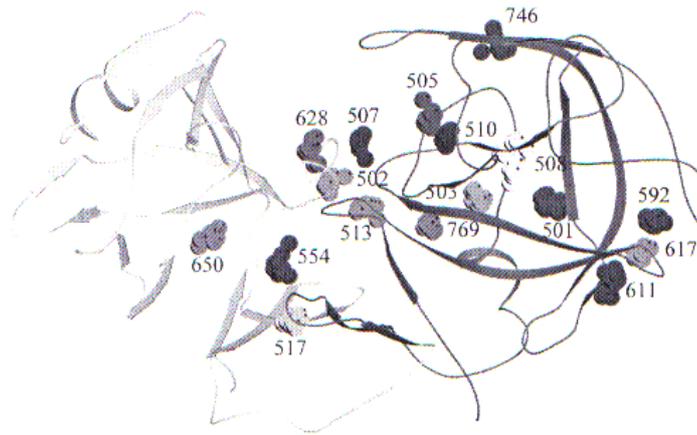


Figure 1.3: The most common 17 conserved water molecules in the crystal structure of Rhizopus-pepsin. Figures have been generated using the program MOLSCRIPT originally by Kraulis in 1991(Prasad and Suguna, 2002).

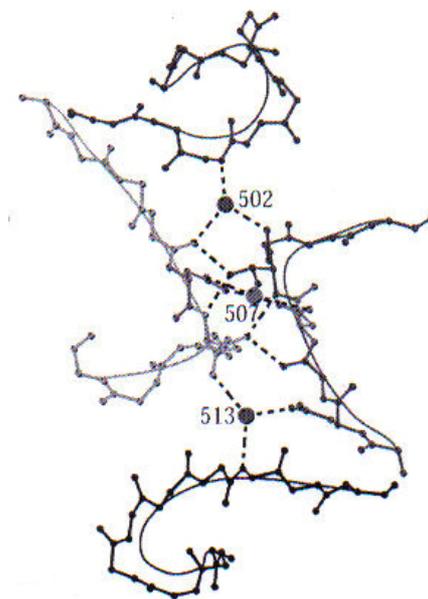


Figure 1 4: Stereoview of firemen’s grip of rhizopuspepsin. Catalytic water molecules Wat 502, Wat 507, Wat 513 which stabilize the aspartic protease active site geometry in the Rhizopuspepsin (Prasad and Suguna, 2002)

The catalytic Aspartate residues contain a conserved motif Asp-Xaa-Gly in both the N- and C-terminal lobes of the protease. Most of the acid proteases, Xaa is Ser or Thr residue, but in retropepsins Xaa is Ala residue

Aspartic proteases have pH optima between pH 3 to 4. Their isoelectric points change from pH 3 to 4.5 according to presence of isoenzymes or autodegradation products. Their molecular masses are in the range of 30 to 45 kDa. Pepstatin is a pentapeptid and naturally produced by *Streptomces* strains. Pepstatin is also known as aspartic protease inhibitor and it is widely used in pharmaceutical industry as model inhibitor. Diazoacetylnorleucinemethyl ester (DAN) inhibits most aspartic protease in the presence of cupric ions. Another inhibitor is 1, 2-epoxy-3-(p-nitrophenoxy) propane (EPNP) (Chitpinityol and Crabbe, 1997).

1.1.4 Action Mechanism of Aspartic Proteases

Pepsin can be the representative enzyme of aspartic proteases. Pepsin is synthesized as inactive pepsinogen. Mature and active pepsin exists after the removal of 41 amino acid residues from pepsinogen by the enzymatic cleavage reaction. At neutral and alkaline pH conditions pro-sequence of the pepsin is required for binding and stabilizing the active sites between two lobes by the help of electrostatic, hydrophobic interactions and hydrogen bonding. At acidic pH conditions acidic residues are protonated and this disturbs the electrostatic interactions with positively charged amino acid residues on the pro-sequence. These conformational changes cause the intramolecular cleavage of zymogens and resulting the existence of mature enzymes. Activation of pepsinogen is done by intermolecular or intramolecular activation. Between pH 4 and 5 intermolecular cleavage dominates, while at lower pH activation tends to be intramolecular (Campos and Sancho, 2003)

The pK_a values have crucial role of action mechanism of aspartic proteases. The first aspartic acid residue of active site is a part of the sequence Ile-Val-Asp-Thr-Gly-Thr-Ser-Leu, and has a pK_a about 5. Second aspartic acid residue is contained

in the sequence Ile-Val-Asp-Thr-Ser-Gly-Ser-Ser-Asn and has a pK_a about below 3 (Al-Janabi *et al.*, 1972). Deprotonated aspartate acts as general base, accepting a proton from H_2O , forming OH^- in the transition state. Other aspartate acts as general acid donates a proton, facilitating formation of tetrahedral intermediate. In general acid-general base mechanism the catalysis is initiated by protonation of the carbonyl oxygen of the substrate by a proton from Asp 215 (general base), followed by the nucleophilic attack on the carbonyl carbon of substrate of Asp by hydroxide ion generated from water after donation of its proton to Asp 32 (general acid). The first protonation cause the formation of the tetrahedral intermediate. Second protonation of intermediate's nitrogen atom from the catalytic carbonyl group of Asp 215 led to the dissociation of tetrahedral intermediate (Polgar 1987), (Figure 1.5). In the nucleophilic mechanism the ionized carboxyl attacked the carbonyl of peptides and protonated one offered a proton to the nitrogen atom to help it to leave (Jiang *et al.*, 2005).

1.2 Microbial Aspartic Proteases

Microorganisms represent an excellent source of aspartic proteases because of their biochemistry and susceptibility to genetic manipulations. They are also easy to handle in laboratories. Their short generation time is another advantage for seeing the effects of genetic manipulations in new generations.

1.2.1 Fungal Aspartic Proteases

Fungal aspartic proteases usually have the similar properties of pepsin (A1 family) or belong to the A4 family which contains only fungal proteases. Their molecular mass is approximately 40 kDa and their optimal pH is generally around pH 3-4. Like other proteases they are synthesized as inactive preproteins with approximately 27-60 amino acids.

Catalytic mechanism of aspartic proteinases

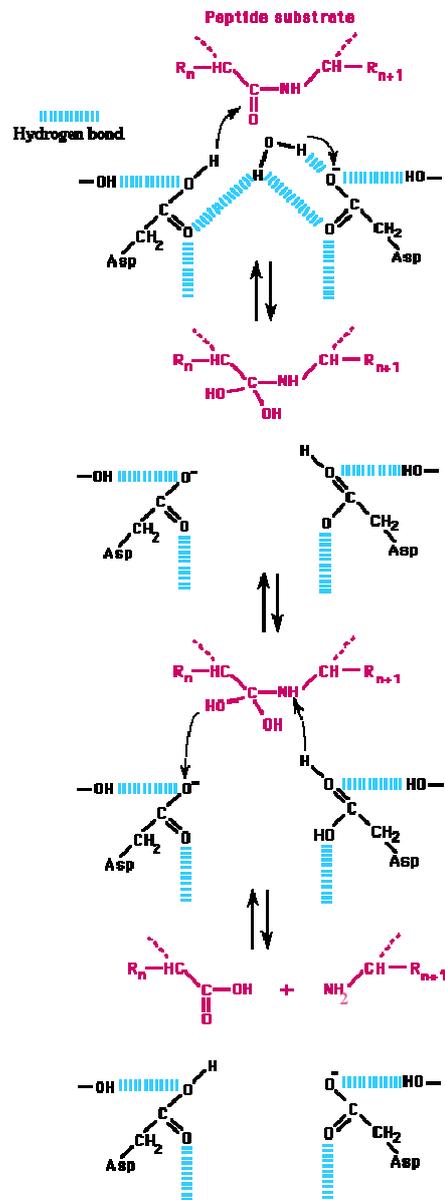


Figure 1.5: The schematic representation of general acid-base catalysis of aspartic proteases.

Mucor miehei and *Mucor pusillus* produce extracellular aspartic proteases which have milk clotting activity. *Mucor species* are preferred by the industry to produce milk coagulation because of its R factor (milk clotting/ proteolytic activity), high specificity for peptide bonds in casein and good quality of cheese (Pessela *et al.*, 2004). These enzymes are called 'Mucor rennins'. *Mucor miehei* aspartic protease has maximal activity at pH 5.5 and its molecular weight is estimated 29-30 kDa (Sardinas 1972). Other *Mucor* species also produce aspartic proteases such as *Mucor varians var. Pisppek*, *Mucor mucedo* *Mucor bacilliformis*.

Endothiapepsin is secreted by the fungus *Endothia parasitica*. As the result of high milk clotting activity; this enzyme can be used in cheese making industry (Barkhold, 1987).

Rhizomucor miehei is another fungus which secretes an aspartic protease. The enzyme has maximal activity for proteolysis at pH 4.1 and for milk clotting at pH 5.6. Enzyme's molecular mass is 20 kDa and also inhibited by pepstatin (Preetha and Boopathy, 1997).

Rhizopus oryzae is a member of fungi and produce extracellular aspartic protease. Monomeric enzyme's molecular mass is 34 kDa. It shows maximal activity at pH 5.5 and acts optimally 60⁰C. It is inhibited by pepstatin (Kumar *et al.*, 2005). *Rhizopus nives* secreted an aspartic protease has optimal activity at pH 3 (Horiuchi *et al*, 1988).

Aspartic protease which is produce by *Aspergillus oryzae* is encoded by pepA gene. This enzyme is used for production of Japanese traditional baverage sake. Nunokawa *et al.* 1955 reported that aspartic protease hydrolyzes protein in steamed rice to liberate peptides. *Aspergillus fumigatus* produce two aspartic proteases (Reichard *et al*, 1995). Pep 1 is one of them and very similar to *Aspergillus niger*'s enzyme which is also called Aspergillopepsin A (Berka *et al.*, 1990). Second of them is called Pep 2 and present in the cell wall of the *Aspergillus fumigatus*.

Trichoderma strains are important as biocontrol agents against a wide set of phytopathogen fungi. *Trichoderma harzianum* encodes pap A an aspartic protease. pI value of the enzyme varied from 4.3 to 4.9 (Delgado-Jarana *et al.*, 2002; Suarez *et al.*, 2005).

Amantia muscaria (Nehls *et al.*, 2001), fungal pathogens *Botrytis cinerea* (Monteau *et al.*, 2003), *Fusarium poae* (Pekkarinen *et al.*, 2000), thermostable fungus *Penicillium duponti* (Hashimoto *et al.*, 1972) are other secreted aspartic proteases.

1.2.2 Yeast Aspartic Proteases

Yeast aspartic proteases contain signal glycosylphosphatidylinositol (GPI) anchoring which locates the proteins in the plasma membrane. YPS3 in other words yeast aspartic protease 3 is one of the aspartic protease produced by *Saccharomyces cerevisiae*. YPS3 is the membrane bound protein and localized in a post-ER compartment of the yeast secretory pathway and has a potential role in zymogen activation processes. Yap3 has the pH optima around pH5-5.5 (Ledgerwood *et al.*, 1996; Olsen *et al.*, 1999).

Candida albicans is the major human pathogen. They produce secreted aspartic proteases that are called Saps. Saps hydrolyze immune system components such as IgG heavy chains, α_2 -macroglobulin, C3 complement system protein. Thus this commensal microorganism became a pathogen especially for immunosuppressive patients like AIDS and cause life-threatening systemic infections. Saps are known as virulence factor of *Candida species*. Nowadays it is known that at least 10 genes encode Sap proteins. These genes are expressed different stages of the pathogenesis and they have special roles (Kumamoto and Vinces, 2005). *Candida tropicalis* (Sapt), *Candida parapsilosis* (Kuhn *et al.*, 2004) (Sapp) and *Candida lusitanae* produce Saps and these enzymes are potential target for new generation drugs. The research of Pichova *et al* results show that Sap2p, Sapt1p, Sapp1p and Sap1p are inhibited by ritanovir and saquinavir. These inhibitors are

derived from the general aspartic protease inhibitor pepstatin A and they are used clinically (Pichova *et al.*, 2001).

AXP1 is expressed by the yeast *Yarrowia lipolytica* against the carbon, nitrogen and sulfur starvation at acidic conditions (Gonzales-Lopez *et al.*, 2002). *Neosartorya fischeri* is a heat resistant mold and secretes 45 kDa aspartic protease and optimum production conditions are determined by Wu and Hang (Wu and Hang, 2000).

1.2.3 Bacterial Aspartic Proteases

Bacterial aspartic proteases also have conserved sites at the catalytic center. ~Hydrophobic –Hydrophobic –Asp-Thr/Ser-Gly which together with a further ~Hydrophobic- Hydrophobic-Gly~ motifs and they are encoded by only one gene. It is reported that *Escherichia coli* and *Haemophilus influenzae* produce aspartic proteases (Hill and Phylip, 1997).

A unique carboxyl peptidase insensitive to the papstatin isolated from *Pseudomonas sp.* Enzyme is the first example of prokaryotic enzyme of this class. It contains 587 amino acids and molecular mass of mature protein is 43 kDa (Oda *et al.*, 1994).

Type II signal peptidase (SPaseII) remove signal peptides from lipid-modified preproteins of *B. subtilis* and *E. coli*. Type IISPases belong to a novel family of aspartic proteases.

Type 4 prepilins are secreted by a wide range of bacterial species and required for type 4 pilus formation, toxin and other enzyme secretory gene transfer and biofilm formation. Type 4 prepilin peptidase represents a novel family of bilobal aspartic protease. *Vibrio cholera* secretes this enzyme (LaPointe and Taylor, 2000).

1.2.4 Viral Aspartic Proteases

According to MEROPS peptidase database all retroviral proteases are involved in A2 family of aspartic proteases. Leukemia virus, immunodeficiency virus, infectious anemia virus, mammary tumor virus are some examples of A2 family members. Retroviral proteases are encoded by a part of pol gene (Dunn *et al.*, 2002). Until now, only six aspartic proteases have been characterized from retrovirus. Observations show that retroviral proteases have conserved regions that stabilize the dimeric structure are located at the interface. Cleavage site region with the two Asp residues is the first region called region (a), and second one the four antiparallel β -sheets that cross-link the two subunits called region b. There are two flap regions into retroviral proteases because of the homodimeric structure (Cascella *et al.*, 2004). Retroviral proteases are essential for viral replication, and they are potential drug targets for treatment of viral diseases.

HTLVI protease molecular mass is 28 kDa and it is homodimeric aspartic protease. Although it has a lot of similarities other retroviral proteases, it exhibits distinct substrate specificity and different sensitivity to aspartic acid inhibitors. MES13-099 and DMP-323 are non-peptide inhibitors for HTLVI protease. Among the peptide derived inhibitors JG-365 has the best inhibition results (Shuker *et al.*, 2003).

HIVI is the causative agent of AIDS. HIV protease is responsible for the cleavage of *gag* and *gag-pol* polyproteins to yield capsid proteins and viral enzymes. It is a single chain protein and active aspartic protease presents only when two identical 99 amino acid domains join to generate a homodimer (Babé and Craik, 1997). There are seven HIV inhibitors indinavir, saquinavir, nelfinavir, ritonavir, amprenavir, lopinavir and atazanavir which are used for clinically (Ohtaka and Freire, 2005).

1.2.5 Archaeal Aspartic Proteases

FlaK is preflagellin peptidase of *Metanococcus voltae*. While FlaK has catalytic dyad of aspartic proteases, (Asp₁₈ and Asp₇₉) are catalytic residues, it does not contain conserved motif D(T/S)G (Lapointe and Taylor, 2000; Bardy and Jarrel, 2003). It is thought that FlaK is novel family of aspartic proteases according to MEROPS peptidase database.

Thermopsin is a thermostable aspartic protease isolated from the *archaea Sulfolobus acidocaldarius*. Thermopsin gene is cloned using pBluescript vector in *E.coli* XL1-Blue competent cells. Thermopsin is a single chain enzyme. Enzyme contains 299 amino acid residues. It has pH optima at pH2 and maximal activity at 90⁰C. The nucleotide sequence showed that the thermopsin structure is encoded by 1020 nucleotides. In the deduced protein sequence, there are 41 amino acid residues (including Met) preceding the NH₂-terminal position of thermopsin. Thermopsin does not have the sequence similarity with other aspartic proteases so it is suggested that thermopsin should be a new class (A5 family) of aspartic proteases. It prefers the hydrophobic residues of substrates like pepsin for cleavage and hydrolyzes Leu-Val, Leu-Tyr, Phe-Phe, Phe-Tyr, and Tyr-Tyr peptide bonds. Maximum activity of thermopsin is recorded at 75 ⁰C and pH 2 with K_m and k_{cat} values under these conditions of 5.3x10⁻⁵M and 14.3 s⁻¹. While pepstatin completely inhibit thermopsin, other aspartic protease inhibitors DAN and EPNP inhibits thermopsin with nonspecific reactions slowly. Furthermore thermopsin can be competitively inhibited by urea, acetamide, and phenylalanineamide. Although there are similarities between thermopsin and pepsin such as size, optimal pH; thermopsin has different structure and active site (Fusek *et al.*, 1990). The most important difference is the absence of the characteristic aspartic protease catalytic dyad (two aspartate residues). There are 11 potential N-glycosylation sites on each thermopsin molecule. The molecular weight estimated from gel filtration (45000 Da) is larger than the calculated from the sequence (36601 Da), suggesting that thermopsin is glycosylated at least some of these 11 sites (Lin and Tang, 1990) It is suggested that thermopsin is a new

class of aspartic proteases which is called A5 family (MEROPS peptidase database).

1.3 Expression of Aspartic Proteases

Expression of recombinant proteins usually attempted in *E.coli* strains as a host because *E.coli* has rapid generation time, a lot of vectors and reagents for cloning and expression are available for it and high amount of expression product can be obtained. In the developing area of heterologous production, yeasts offer a number of advantages as expression systems for complex proteins of eukaryotes. As unicellular organisms, they retain the advantage of bacteria in ease of manipulation and growth capacity (Madzak *et al.*, 2004). Viral expression systems are other alternatives for protein expression studies. For example, baculovirus expression system was used for the expression of thermopsin gene from *Sulfolobus acidocaldarius* and Vaccinia virus expression system was used for the production of recombinant human cathepsin D (Lin *et al.*, 1992; Demoz *et al.*, 2006).

Heterologous protein expression is widely used in both biological and industrial researches. There are some proteins which facilitate the soluble expression and refolding of recombinant proteins: maltose binding proteins (MBP), thioredoxin (Trx), glutathione-S transferase (Boross *et al.*, 2005). *Boophilus micropulus* aspartic protease called BYC, was expressed as fusion with thioredoxin protein (Leal *et al.*, 2006). Histidine tags are used as a fusion partner of aspartic proteases and it facilitate the selection and purification of the recombinant fusion protein with affinity chromatography (Hochuli *et al.*, 1987). Human pro-BACE-1 (β -secretase) was expressed as a fusion protein with His tag (Bruinzeel *et al.*, 2002)

A 35kDa thermopsin protein was expressed as a heterologous fusion protein as pepsinogen-thermopsin (PP-TH), procathepsin-D (PCaD), rhizopus pepsinogen (RP-TH) to prevent the degradation, under the control of T7 promotor in *E.coli* strains. In this study pET-3B vector was used for the construction of expression

vector resulting in 16 extra amino acids on pepsinogen N-terminus. Recombinant pET-3b-PP vector transfected into *E.coli* strain BL21(DE3). The thermopsin gene was modified with oligonucleotide-directed mutagenesis with the addition of a starting codon before residue -14, which is immediately preceded by a *Bgl*III site. The ligation of the *Bgl*III fragment of thermopsin gene to the unique *Bam*HI site of the pepsinogen gene in pET-3b-PP resulted in a pepsinogen –thermopsin fusion in which the correct reading frame was preserved. PCaD gene was modifying with PCR to introduce a *Pst*I site at at the 5' end of the PCaD gene for cloning into the expression vector and a *Hind*III site at 3' end of the PCaD gene for a fusion to the thermopsin gene. Simultaneously other PCR was performed to generate a *Hind*III site at 5' end of the thermopsin gene and at the same time, installed two Asp-Pro residues between PCaD and TH. This sequence would permit the cleavage of two peoteins in fusion by mild acid hydrolysis. Fusion gene with a *Bam*HI site at both ends was cloned into the *Bam*HI site of the pET vector for expression. RP-TH fusion gene was cloned to pET-3xb. A honeybee mellitin signal peptide was placed in front of the thermopsin coding sequence in order to facilitate secretion of recombinant thermopsin. Mel-TH gene was cloned into pBlueBac for expression. The best expression results were obtained from pepsinogen-thermopsin fusion. It was shown that unglycosylated thermopsin had thermostability similar to that of the native (glycosylated) enzyme. The expression level of glycosylated recombinant thermopsin from the baculovirus system was poor. Moreover in baculovirus/insect cell expression systems preliminary results were promising in that thermopsin can be synthesized as an active enzyme (Lin *et al.*, 1992).

1.4 Scope and Aim of the Study

Aspartic proteases are also called acid proteases have vital role in metabolism and besides its physiological functions it is very important for industrial application and structure-function relationship studies. Cheese making and pharmaceutical industries are the fields that aspartic proteases are widely used. Stability under extreme conditions of the enzyme is the most important criteria that determine the

enzyme quality for industrial applications. Thus thermostable enzymes can be used in harsh industrial process. They are also resistant to contamination and solvents.

Thermopsin is one of the thermostable aspartic protease from thermoacidophilic archaeobacterium *Thermoplasma volcanium* which optimally grows at 60⁰C and pH 2.7. In this study it was aimed, for the first time, to heterologous expression of thermoacidophilic thermopsin gene from *Thermoplasma volcanium* in mesophilic host *E. coli* as a 6XHis tag fusion protein. Also, amino acid sequence analysis from the structural and functional points of view was performed.

CHAPTER II

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals, Enzymes and Kits

During the experiments all used chemicals were molecular biology grade. Agar agar hochrein, Trypton (pepton from casein), Sodium chloride (NaCl), Tris H₂NC (CH₂OH)₃, Sodium hydroxide pellets pure (NaOH), D(+)-Glucose(C₆H₁₂O₆), Yeast Extract granulated, Magnesium sulphate heptahydrate extra pure (MgSO₄·7H₂O), Potassium dihydrogen phosphate crystalized extra pure (KH₂PO₄) were from MERCK, Darmstad, Germany. Tween 20 (Polyoxyethylene sorbitan monolaurate), Triton X-100 (t-Octylphenoxypolyethoxyethanol), Phytigel (gellan gum; agar substitute gelling agent), Glycerol (glycerin), Dodecylsulfat natriumsalz (C₁₂H₂₅NaO₄S), Clacium chloride dihydrate (CaCl₂·2H₂O), Ammonium sulfate ((NH₄)₂SO₄), Agarose (low melting point), Ampicillin sodium salt (D (-)-α Aminobenzylpenicillin), Kanamycin (Kan A) mono sulfate (from *Streptomyces kanamyceticus*) were purchased from Sigma Chemical Co.; St. Louis. Yeast Extract was also purchased from DIFCO; Detroid, USA. Ethanol was from Reidel de Häen.

Restriction endonucleases; *SphI* (*PaeI*), *Sall*, *EcoRI*, *EcoRV*, *PstI*, *BamHI*, *HindIII* and their suitable buffers were from MBI Fermentas AB; Vilnius, Lithuania. Taq

polymerase, T4 ligase and their buffers were also from MBI Fermentas AB; Vilnius, Lithuania. Immobilon-NC Transfer Membrane used for screening of expression clones was from Millipore Corporation, USA.

PCR Cloning Kit, QIAExpressionist Type IV Kit, QIAGEN HRP Conjugate Kit were from QIAGEN Inc.; Valencia, USA. DNA Purification System purchased from Promega Corporation; Madison, WI, USA. BIO 101 Systems-Gene Clean Kit was used for DNA extraction from gel was from Q-BIOgene Corporation. Two different chromogenic substrates, BM blue POD substrate and DAB substrate used for detection of expressed thermopsin gene, were from Roche Applied Science; Indianapolis, USA. Other chromogenic substrate was from Roche Applied Science; Penzberg, Germany. All of the primers were synthesized by Applichem; Darmstadt, Germany.

2.1.2 Buffers and Solutions

Compositions of buffers and solutions used in this study are listed in Appendix A.

2.1.3 Plasmid vectors, Molecular Size Markers and Genomic DNA Sequence Data

The maps of cloning and sub-cloning vectors are illustrated in the Appendix B. DNA molecular size marker is shown in Appendix C. Nucleotide sequence of the thermopsin gene from *Thermoplasma volcanium* is given in Appendix D.

2.2 Strains and Mediums

2.2.1 Bacterial and Archaeal Strains

Thermoacidophilic archaea *Thermoplasma volcanium* GSS1 was used in this study as source organism.

Thermoplasma volcanium GSS1 (Type strain 4299) culture was purchased from DSM (Deutsche Sammlung von Microorganismen und Zellkulturen, GmbH, Braunschweig, Germany). For transformation studies different recipients were used. *E.coli* Strains M15 [pREP4] and SG13009 [pREP4] were from QIA-Expressionist Kit, QIAGEN and *E.coli* strain TG1 was from our laboratory collection were used as host microorganisms during cloning and sub-cloning procedures. TG1 genotype is: supE thi-1 Δ (lac-proAB) Δ (mcrB-hsdSM)5 (rK⁻ mK⁻) [F' traD36 proAB lac^qZΔM15]. Both of the strains derived from *E.coli* K12 and have the phenotype Nal^S, Str^S, Rif^S, Thi⁻, Lac⁻, Ara⁺, Gal⁺, Mtl⁻, F⁻, RecA⁺ Uvr⁺, Lon⁺.

2.2.2 Growth and Maintenance of Microorganisms

Thermoplasma volcanium was grown in Volcanium medium (pH 2.7) (Robb, 1995) which was supplemented with glucose (MERCK) and yeast extract (Oxoid) to final concentrations of 0.5% and 0.1% approximately 60⁰C. Cultures were renewed by subculturing routinely once a week.

E. coli M15 [pREP4] and SG13009 [pREP4] strains were grown on LB agar which was supplemented with Kanamycin and Ampicillin to final concentrations of 25 μg/ml and 100 μg/ml. *E. coli* strain was grown LB medium which was supplemented with Ampicillin to final concentration of 50 μg/ml. For routine uses these *E.coli* strains were grown on LB agar plates with described amount of antibiotics at 37⁰C. The cultures were renewed by subculturing once a month.

2.3 Gene Manipulation Methods

2.3.1 Genomic DNA Isolation from *Thermoplasma volcanium*

Genomic DNA of *Thermoplasma volcanium* was isolated essentially using the modified method of Sutherland *et al.* (1990). After cell lysis, NaClO₄ was added to obtain a final concentration of 1 M. Then, one volume

chloroform:isoamylalcohol (24:1, v/v) was added and the mixture was swirled on ice-water bath for 15 minutes until it was homogenous. After centrifugation at 11 000 rpm for 15 min. at 4°C, the aqueous phase was transferred into a new tube and two volumes of ice cold absolute ethanol was added. The precipitated DNA following incubation at -20°C for 1 hour, was collected by centrifugation at 12 000 rpm for 45 min., at 4°C. The supernatant was discarded and the pellet was washed with 70 % ethanol. After centrifugation the supernatant was poured and the remaining ethanol was removed by air drying. The DNA pellet was then dissolved in 1 ml of sterile double distilled (dd) H₂O, stored at - 20 °C, until use.

2.3.2 Purification of Chromosomal DNA of *Thermoplasma volcanium*

Isolated chromosomal DNA was purified as described by Sambrook *et al.* (2001). After removal of RNA by RNAase (Sigma Chemical Co., S. Louis, Missouri, USA) treatment, two successive (1/1, v/v) phenol:chloroform extractions were performed. The contaminating phenol was removed by ether extraction. The purified chromosomal DNA was precipitated with ethanol and washed with 70 % (v/v) ethanol. Resulting DNA pellet was air dried and dissolved in minimal volume of sterile ddH₂O.

2.3.3 Spectrophotometric Analysis of DNA

The amount of purified DNA was quantified as described by Sambrook *et al.* (2001). A series of dilutions (1/200-1/500) of DNA samples were prepared in sterile dd H₂O. The DNA concentrations of these samples were determined spectrophotometrically (Shimadzu Analytical Co. Kyoto, Japan) at 260 nm. The ratio between the readings at 260 nm and 280 nm (OD₂₆₀/OD₂₈₀) was calculated to estimate the purity of the nucleic acids in the samples. A pure solution of double stranded DNA solution at 50 µg/ml has an optical density of 1.0 at 260 nm and an OD₂₆₀/OD₂₈₀ ratio of 1.8 (Sambrook *et al.*, 2001).

2.4 Agarose Gel Electrophoresis

DNA fragments which were ligated into pQE vectors for sub-cloning were isolated from 1% (w/v) low melting point agarose gel after electrophoresis (Sigma Chemical Co., St. Louis, Missouri, USA). On the other hand PCR amplicons, DNA fragments after restriction digestion and DNA samples following any manipulation were analyzed by electrophoresis on 0.8% agarose gel (Applichem Co., Darmstadt, Germany) in 1X TAE buffer. DNA samples of 10-20 µl, mixed with 1/10 volume of 6X tracking dye, were applied to gel which was supplemented by ethidium bromide (0.5µg/ml). After analyzing of the samples by electrophoresis the bands were visualized with a UV transilluminator (Vilber Lourmat, Marne La Vallée Cedex 1, France) and gel photographs were taken using a gel imaging and documentation system (Vilber Lourmat, Gel Imaging and Analysis System Marne La Vallée Cedex 1, France). *EcoRI/HindIII* cut Lambda DNA (MBI Fermentas AB, Vilnius, Lithuania) was used as a size marker (see appendix C) to estimate the molecular weight of DNA bands after different manipulations. The molecular sizes of DNA fragments were determined by referring to calibration curves, which were obtained by plotting the log molecular weights of known marker fragments against migration distance on the gel. Calibration curve generated for *EcoRI/HindIII* cut Lambda DNA (MBI Fermentas AB, Vilnius, Lithuania) is shown in the Figure 2.1.

2.5 PCR Amplification of Thermopsin Gene of Putative Acid Proteases

2.5.1 Design of PCR Amplification Primers

A PCR amplification primer set was designed on the basis of predicted thermopsin gene sequence as reported by Kawashima *et al.*(1999). Forward (TP1) and reverse primer (TP2) sets flanked a region of 248052-251132 which covered complete thermopsin gene with some upstream and downstream sequences.

In addition second set of primer (TP1'and TP2) was designed according to recommendation of QIA-Expressionist Kit so that at the 5'-end of amplification

product start codon of the thermopsin structural gene was missing. The list of PCR primers is given in Table 2.1 and 2.2.

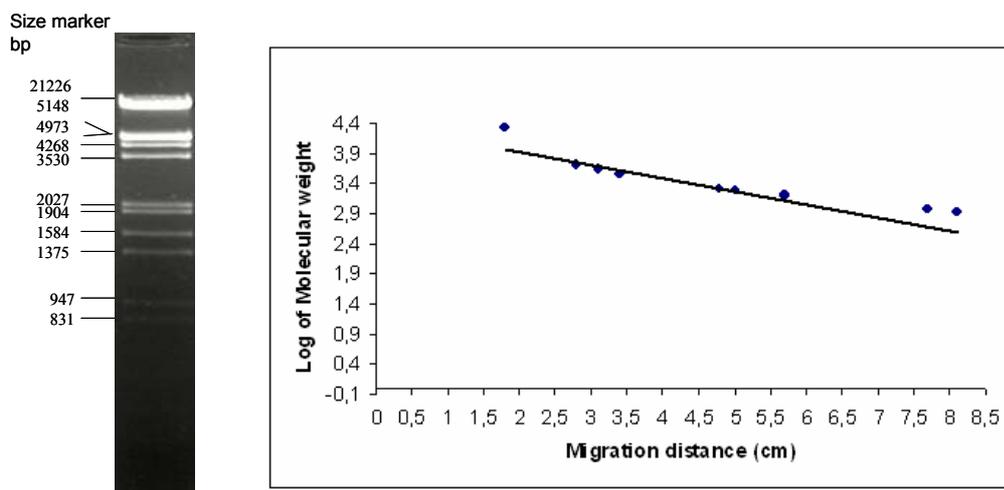


Figure 2.1: Calibration curves for *EcoRI* and *HindIII* cut Lambda DNA molecular size markers. (MBI Fermentas AB, Vilnius, Lithuania)

Table 2.1: Primers designed for the putative acid protease gene sequence of *Thermoplasma volcanium*.

Name of the primer	Sequence of the primer (5'-3')
TP1 Forward Primer	5'-GGAATTTTGGTAATATCTCTGATTACCG-3'
TP2 Reverse Primer	5'-GCCATAAAATCTATCTTTCCCC-3'

Table 2.2: Primers designed for the putative acid protease gene sequence of *Thermoplasma volcanium* following the recommendation of expression kit.

Name of the primer	Sequence of the primer (5'-3')
TP1' Forward Primer	5'-GTAATATCTCTGATTACCGTTACG-3'
TP2 Reverse Primer	5'-GCCATAAAATCTATCTTTCCCC-3'

2.5.2 PCR Amplification

PCR amplifications were carried out with TP1/TP2 and TP1'/TP2 primer sets using *Thermoplasma volcanium* genomic DNA as template. PCR reaction mixture (in a total volume of 100 μ l), 1X PCR buffer (composition of buffer; 750mM Tris-HCl pH 8.8 at 25 °C, 200mM (NH₄)₂SO₄, 0.1% Tween 20) was consisted of 200 μ M of each deoxyribonucleoside triphosphate (dNTP), 50-500 ng of the template DNA, 1.5 mM MgCl₂, 100 pmol each primer and 2.5 U of Taq DNA polymerase (MBI Fermentas AB, Vilnius, Lithuania). The amplification program of the thermal cycler (Techgene, Techne Inc. NJ. USA) started with the pre-heating step at 94°C for 5 minutes. This step was followed by 30 cycles of amplification (denaturation at 94°C for 1 minute, annealing at 55-58°C for 2 minutes, extension at 72°C for 3 minutes and final extension at 72°C for 10 minutes). PCR products (15 μ l) were analyzed by agarose gel (0.8% w/v) electrophoresis.

2.6 Isolation of DNA Fragments from Agarose Gel

PCR products (3080bp and 3070 bp) and linearized vector DNAs (pUC18, pQE 30-31-32 vectors) were run on the 1% (w/v) low melting point agarose gel (Sigma

Chemical Co., St. Louis, Missouri, USA) using a BIO-RAD mini-gel and Power Supply1000/500 system. BIO 101 Systems-Gene Clean Kit (Q-BIOgene Corporation) was used for gel extraction of fragments and vectors. Gel extraction was according to the Kit's instructions.

Firstly we determined the approximate volume of the gel slices (0,1g equals approximately 100µl) in 1.5ml microcentrifuge tubes. For dissolving the gel, we added 3 volumes of NaI. Tubes were incubated in a 55⁰C water-bath for 5 minutes, and then, GLASSMILK was added (5µl for 500µl volume). Then the mixture was incubated at room temperature for 5 minutes through mixing every 1 minute to allow binding of the DNA to the silica matrix. The tubes were centrifugated for 45 seconds at 13.000rpm in a microcentrifuge (Eppendorf minifuge) and the supernatant was discarded. For resuspension of the pellet, we added 500µl of NEW Wash solution and resuspended the pellet by pipetting back and forth. After resuspension, the glassmilk pellet was obtained by centrifugation and the washing step was repeated two more times. Following the last wash step pellet was air dried. For elution we added the minimum volume of sterile distilled water and incubated the tubes for 5 minutes at 55 ⁰C. After centrifugation of the sample for 60 seconds at 13.000rpm supernatant containing DNA was transferred put into a new sterile tube and stored at -20⁰C, until use.

2.7 Cloning of PCR Amplified Thermopsin Gene

There are two different cloning strategies for PCR amplified thermopsin gene.

2.7.1 Cloning of PCR Amplified Thermopsin Gene with TP1/TP2 Primer Set

Previously in our laboratory, PCR amplified 3080 bp Thermopsin gene has been cloned using the QIAGEN PCR Cloning Kit (QIAGEN Inc., Valencia, USA) into the pDrive vector. In the positive clone, PCR amplified thermopsin gene was placed opposite direction relative to LacZ promoter in pDrive vector. It was very

important detail for us because we had planned to use them for subcloning for expression. On the other hand, pQE 30-31-32 triple expression vector set (QIAGEN Inc., Valencia, USA) were so designed that N-terminus 6XHis tag fusions could be produced. In other words, 5' end of the ligated open reading frame should be in frame with 6XHis for expression to occur. To be able to ligate the 5' end of the gene downstream of the 6XHis tag, suitable sticky ends could be provided only after subcloning into pUC18 vector. To this, firstly, the thermopsin gene was excised from pDrive vector via *SphI* (*PaeI*) and *SalI* (MBI Fermentas AB; Vilnius, Lithuania) restriction enzymes. Secondly pUC18 vector was digested with the same enzymes. Both thermopsin gene (3080bp) and linearized pUC18 vector was isolated from gel via BIO 101 Systems-Gene Clean Kit (Q-BIOgene Corporation). After the gel extraction, thermopsin was ligated to pUC18 vector by T4 ligase and the recombinant plasmids were introduced into *E. coli* TG-1 cells by transformation.

Schematic representations and Multiple Cloning Sites of pDrive and pUC18 vectors are shown in Figure 2.2.

2.7.2 Cloning of PCR Amplified Thermopsin Gene with TP1'/TP2 Primer Set

PCR amplified putative acid protease gene (thermopsin) lacking start codon was cloned by using the QIAGEN PCR Cloning Kit (QIAGEN Inc., Valencia, USA). The PCR fragment, was obtained by second primer set, was ligated into pDrive vector according to the Manufacturer's instructions.

2.7.3 Sub-cloning of Putative Thermopsin Gene for Expression

QIAExpressionist Kit Type IV was used for expression of the cloned thermopsin genes. pQE triple vector set (pQE 30-31-32), which had three different open reading frame, were included in kit. These vectors provide high level expression of 6XHis-tagged proteins in *E. coli* based on their T5 promoter transcription-translation system.

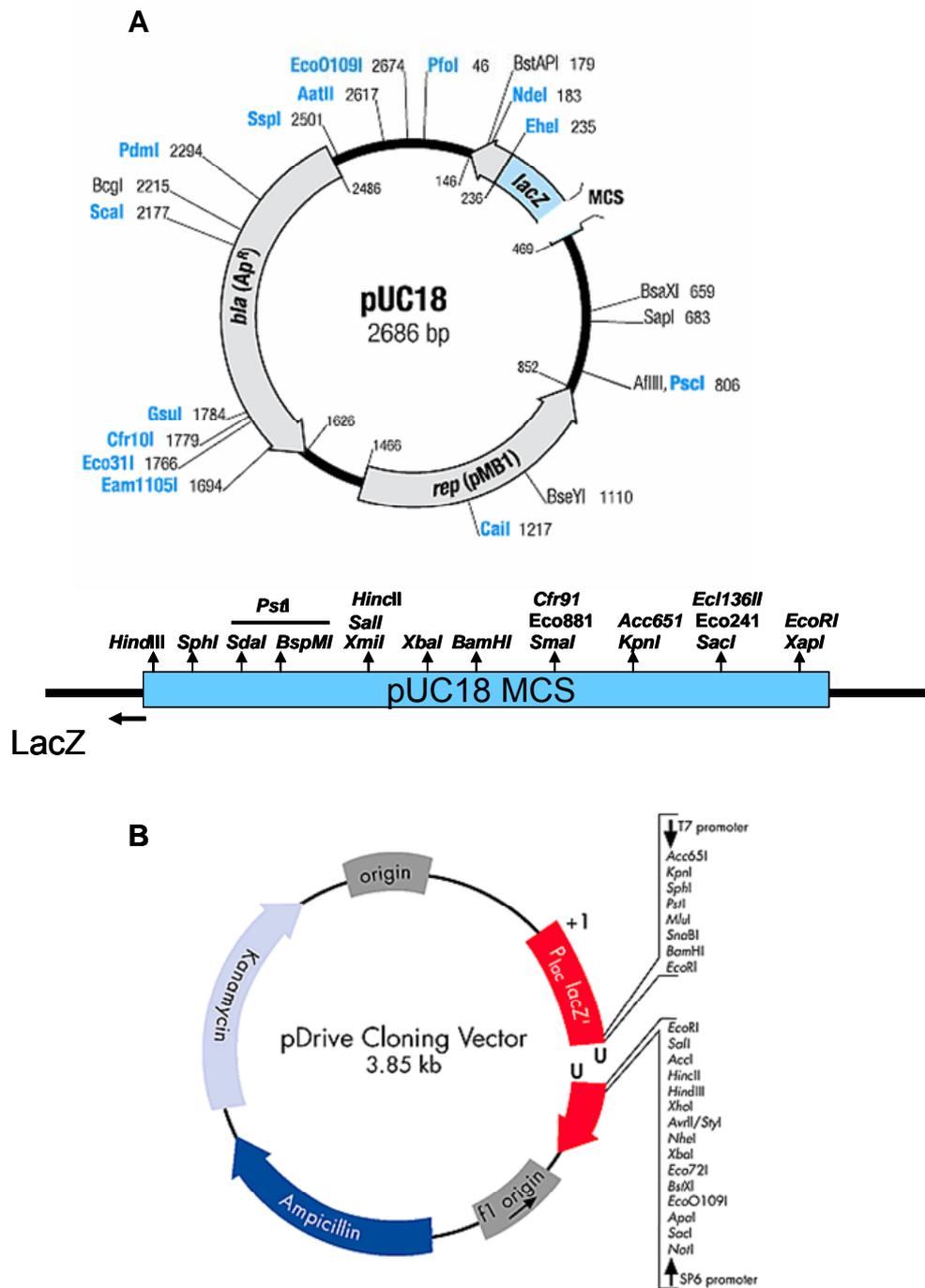


Figure 2.2: Schematic representation of vectors. A represents pUC18 and multiple cloning site of pUC18 vector and B represents pDrive vector.

Expression vectors from the QIAexpressed Kit were prepared by first dissolving them in TE (pH 8.0) buffer. The addition of 10µl TE to 5µg of plasmid DNA was done. A 2 µl aliquot was linearized using the appropriate restriction enzymes.

The recombinant pDrive vector containing 3070bp PCR product was digested via *SphI* and *SalI* restriction enzymes. pQE triple vector set (pQE30-31-32) were also digested with the same restriction enzymes. Excised cloned fragment (3070bp) and linearized vector DNAs were isolated by extraction from low-melting agarose gel.

Also, recombinant pUC18 vector containing 3080bp PCR product and pQE 30-31 was digested with *SalI* and *HindIII* restriction enzymes. The cloned 3080bp PCR fragment and linearized vectors were, then, isolated from low-melting agarose gel.

Thermopsin (both 3070bp and 3080bp PCR products) genes were ligated into pQE 30-31-32 vectors according to the Manufacturer's method using T4 ligase (MBI Fermentas AB; Vilnius, Lithuania) The general protocol for sub-cloning is schematized in Figure 2.4.

2.8 Introduction of Recombinant Plasmids into Competent Cells

2.8.1 Preparation of Competent *E.coli* Strains

Competent *E.coli* TG1 strains from our laboratory collection, M15 [pREP4] and SG13009 [pREP4] strains from QIAexpressionist kit, were prepared according to the modified method of Chung *et al.* (1989). *E.coli* TG1 strains were grown in unsupplemented LB broth, and *E.coli* M15 [pREP4], SG13009 [pREP4] cells were grown in LB broth which was supplemented with kanamycin (25µg/ml) overnight. All of the strains the volume of 0.2ml were inoculated into 20 ml LB medium and they incubated at 37⁰C through vigorous shaking (Heidolph UNIMAX1010, Heidolph Instruments GmbH, Kelheim, Germany). The growth was followed by measuring the optical absorbance of cell culture at 600 nm using

Cloning target sequence into QIAexpress vectors

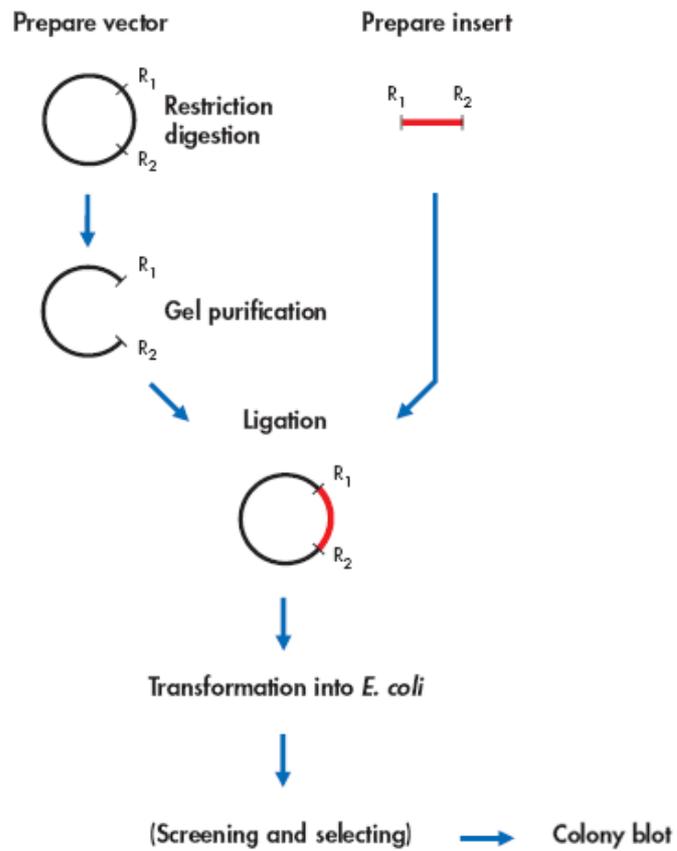


Figure 2.3: Schematic representation of cloning target sequence into QIAexpress vectors.

the Shimadzu UV-160A double beam spectrophotometer (Shimadzu Analytical Co. Kyoto, Japan). The cells in early log phase ($\sim 10^8$ cells/ml) were collected by centrifuging at 4000 rpm for 10 min. After discarding the supernatant, the pellet was dissolved in 1/10 TSS solution and distributed into eppendorf tubes in aliquots (100 μ l). Competent cells were stored at -80 °C until use.

2.8.2 Transformation

Competent *E.coli* cells (M15 [pREP4], SG13009 [pREP4], TG1) were removed from deep freezer and thawed on ice. The ligation mixture was added into the tube containing 100 μ l of competent cells, and the mixture was incubated for 30 minutes on ice-bath. Then the cell content was transferred into 0.6ml LB broth which was supplemented with 1M glucose. Mixtures were incubated at 37⁰C with vigorous shaking (Heidolph UNIMAX1010, Heidolph Instruments GmbH, Kelheim, Germany) for 1.5 hour. Control with no added DNA or ligation mix was also included into each assay. Different dilutions of the control and test samples were prepared and plated onto LB agar plates containing IPTG, X-Gal, and Ampicillin (in a final concentration of 50 μ g/ml) and incubated at 37⁰C overnight, to obtain single colonies. Recombinant colonies were detected by the blue-white colony screening method. Putative recombinant white colonies picked-up and inoculate on new LB agar plates containing relevant antibiotic. In sub-cloning experiments for expression the selective LB agar medium was prepared by including ampicillin, 50 μ g/ml (if TG1 competent cells are used) or ampicillin, 100 μ g/ml plus kanamycin, 25 μ g/ml (if M15[pREP4] and SG13009[pREP4] competent cells are used). The plates were incubated at 30⁰C, overnight. The colonies were screened by colony blotting and immunoscreening, as described below.

2.9 Colony Blotting and Hybridization

Colony blot hybridization method was used for positive expression colony screening because it was allowed simultaneous screening of transformants with

the correct coding fragment, expression levels, and in-frame translation of the 6XHis tag. This method was modified from the instruction manuals of QIAexpressionist Kit and Anti-His HRP Conjugate Kit (QIAGEN Inc., Valencia, USA).

Anti-Histidine conjugates are specific for histidine residues so they bind to expressed 6XHis tagged protein's histidine residues. These conjugates carry horse raddish peroxidase (HRP) enzyme and when the chromogenic substrates exist enzyme catalyzes the reaction resulting color development. So Anti-His HRP conjugates recognize and bind to only 6XHis tag protein and chromogenic substrate is used as reporter. Significant color development can be used for sign of expression.

After transformation plates were incubated overnight at 30⁰C until the colonies are about 1-2mm in diameter. After the removal of plates from the incubator, lids were opened slightly and allowed any condensation to dry. Dry and numbered nitrocellulose membrane (Millipore Cat. No: HATF08250 Filter type 0.45 um, white surfactant free, HATF 82 mm) was placed on the agar surface in contact with the colonies. Filter was then transferred to a fresh plate containing antibiotics and IPTG (final concentration of 250µM). The replica plates were incubated in inverted position for 4 hours at 37⁰C to induce expression. The master plates were also incubated to allow the colonies to regrow. A set of polystyrene dishes were prepared for colony lysis and binding of proteins to the filters. Each dish contained a sheet of 3MM (Whattmann Int. Ltd., Maidstone) paper soaked with one of the following solutions; SDS solution, denaturation solution, neutralization solution, 2XSSC (Salt Sodium Citrate). Nitrocellulose filters with colony blots were placed (colony side up) on top of the soaked filter papers and incubated at room temperature as follows: in SDS solution for 10 minutes, in Denaturation solution for 5 minutes, in neutralization solution for 5 minutes,(two times) and in 2XSSC for 15 minutes. This incubation steps were followed by washing step with TBS buffer for 10 minutes. Then nitrocellulose filter was transferred into a sterile petri dish containing blocking buffer and incubated for 1 hour in a shaking

incubator at room temperature. To prevent unspecific binding filter was firstly washed in TBS-Tween/Triton buffer and then in TBS buffer each was for 10 minutes. For immunological screening, filter was incubated in Anti-His HRP conjugate solution for 1 hour in a plastic bag through gentle shaking. This step was followed by washing in TBS-Tween/Triton buffer and then in TBS buffer each for 10 minutes. The filter was stained with the DAB chromogenic substrate of Anti-His-HRP conjugate until the signals were clearly visible. Lastly the reaction was stopped by rinsing the membrane once with sterile distilled water. The colors would be fade with time because the product formed when using HRP was particularly unstable so we photographed the filter as soon as possible. All these steps are schematized in Figure 2.3. In the Figure 2.4 it is shown that the color development principle of QIAExpression kit.

2.10 Isolation of Plasmid DNA from Recombinant Colonies

Plasmid isolation was carried out by using Wizard® Plus SV Minipreps DNA Purification System (Promega Corporation, Madison, WI, USA) according to the procedure provided by the manufacturer. The cultures (10ml) were centrifugated at 4,000 rpm for 15 min. After removal of supernatant the cell pellet was thoroughly resuspended in 250 µl Cell Resuspension Solution. The cells were lysed then by adding 250 µl Cell Lysis Solution and the proteins were digested with 10 µl Alkaline Protease. After incubation for 5 min at room temperature 350 µl Neutralization Solution was added and incubated on ice for an additional 1-5 minutes. The cell debris was pelleted by centrifuging at 13000 rpm for 10 minutes and the supernatant was transferred to Spin Column which was provided with the Kit and centrifuged at 13000 rpm to bind the DNA to the column. The bound DNA was washed twice with the Washing Solution and eluted with 100 µl Nuclease Free Water. Finally 100 µl of plasmid sample, at a concentration of 1.5 mg/ml, was obtained from the starting culture solution. Nearly 15 µl of the sample was run on the agarose gel to check the efficiency of purification and the remaining sample was stored at -20°C for further use.

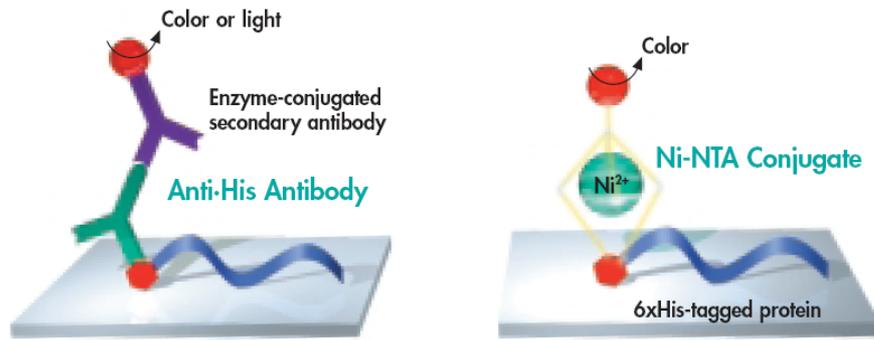


Figure 2.4: The principle of color development in Anti-His HRP conjugate kit. We used Anti-His HRP Conjugate and HRP specific DAB chromogenic substrates. (Figure is taken from QIAExpressionist Kit Manual)

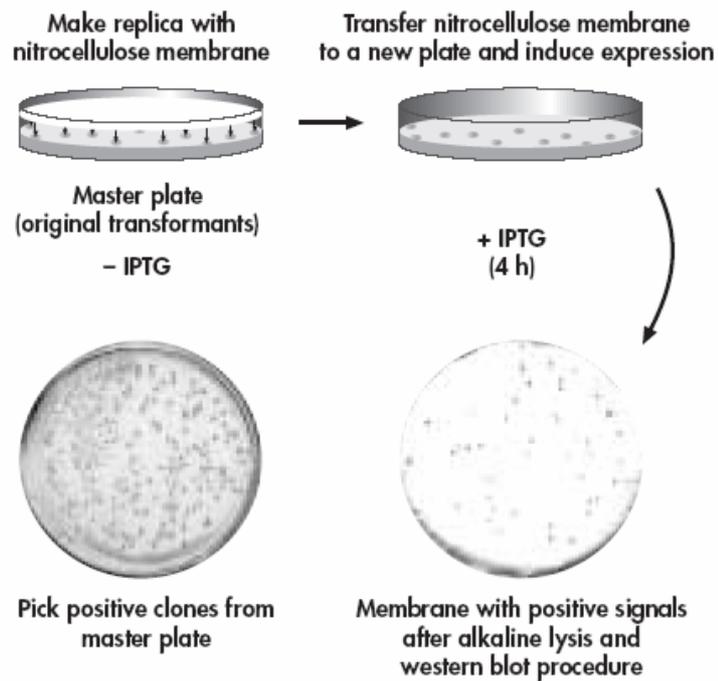


Figure 2.5: Detection of positive expression clones by colony blotting hybridization. (Figure is taken from QIAExpressionist Kit Manual)

2.11 Computer Analysis of Gene and Protein Sequence

NCBI (National Center for Biotechnology Information) nucleic acid sequence database was searched to obtain the putative acid protease (thermopsin) gene sequence of *Thermoplasma volcanium*.

ClustalW Version 1.8 program was used for the multiple sequence alignments of archaeal putative acid protease amino acid sequences.

Alternative alignment was performed by KEGG (Kyoto Encyclopedia of Genes and Genomes) database according to SW score.

Restrictionmapper version 3 program was used to map the sites for restriction endonucleases in the putative thermopsin gene sequence.

DNA sequence analysis and domain search was done by Pfam (Protein Families Database).

Thermopsin gene details and comparative analysis of gene localization among the archaea were obtained from IMG (Integrated Microbial Genomes) database.

The gene sequence details of thermopsin were also obtained from TIGR (The Institute for Genomic Research) database.

CHAPTER III

RESULTS

3.1 Gene Manipulation

3.1.1 Genomic DNA Isolation of *Thermoplasma volcanium*

Genomic DNA was isolated from *Thermoplasma volcanium* and it was used for the template DNA for thermopsin gene amplification by PCR method (Figure 3.1).

In each purification, isolated DNA samples were analyzed spectrophotometrically (Sambrook *et al.*, 2001). The absorbance readings at 260 nm and 280 nm were obtained to calculate the quantity and purity of DNA samples as described in Section 2.3.3. The concentrations of the samples varied over the range of 1.0-1.6 mg/ml.

OD₂₆₀/OD₂₈₀ ratios were between 1.7 and 1.8, which indicated a sufficient purity for the DNA samples to be used in gene manipulation. The eligibility of the purified DNA samples for gene manipulation experiments was also checked by restriction enzyme digestion (Figure 3.1).

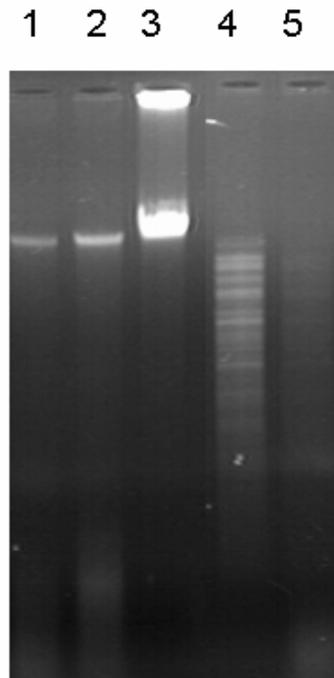


Figure 3.1: Genomic DNA of *Thermoplasma volcanium*. Lane 1 and 2, uncut DNA sample 1; lane 3, uncut DNA sample 2; lane 4, *EcoRI* cut DNA sample 2; lane 5, *EcoRI* cut DNA sample 1.

3.1.2 Restriction Maps of 3080bp and 3070bp cloned Thermopsin Gene Fragments

Recognition sites of different restriction enzymes on 3080bp and 3070bp PCR amplified fragments were determined by Restriction Mapper version 3 program. Restriction enzymes and cut positions and also noncutter restriction enzymes are shown in Table 3.1 and 3.2 for 3080bp and 3070bp cloned thermopsin gene fragments.

Table 3.1: Restriction map of 3080bp cloned thermopsin gene.

Noncutter Restriction Enzymes	Cutter Restriction Enzymes	Cut positions
<i>Ava</i> I	<i>EcoRV</i>	273
<i>Avr</i> II	<i>Pvu</i> II	2947
<i>Bcl</i> II	<i>Ssp</i> I	604
<i>Hind</i> III	<i>Bam</i> HI	1120
<i>Kpn</i> I	<i>Bgl</i> II	3004
<i>Mlu</i> I	<i>Bgl</i> I	869
<i>Not</i> I	<i>Bst</i> XI	217
<i>Pvu</i> I	<i>Sac</i> I	1827
<i>Sac</i> II	<i>Eco</i> RI	1150, 1722
<i>Sal</i> II	<i>Pst</i> I	1790, 2952
<i>Sma</i> I	<i>Ava</i> II	1048, 1707, 2587
<i>Sph</i> I	<i>Hae</i> II	1114, 1180, 1430
<i>Xho</i> I	<i>Acc</i> I	402, 1622, 1937, 2569, 2750

Table 3.2: Restriction map of 3070bp cloned thermopsin gene.

Noncutter Restriction Enzymes	Cutter Restriction Enzymes	Cut positions
<i>Ava</i> I	<i>EcoRV</i>	264
<i>Bcl</i> II	<i>Pvu</i> II	2938
<i>Hind</i> III	<i>Bam</i> HI	1111
<i>Kpn</i> I	<i>Bgl</i> I	860
<i>Mlu</i> I	<i>Sac</i> I	1818
<i>Pvu</i> I	<i>Eco</i> RI	1141, 1713
<i>Sal</i> II	<i>Pst</i> I	1781, 2943
<i>Sma</i> I	<i>Ava</i> II	1039, 1698, 2578
<i>Sph</i> I	<i>Acc</i> I	393, 1613, 1928, 2560, 2741

3.1.3 PCR Amplifications and Cloning of PCR Amplified Fragments

3.1.3.1 Cloning of PCR Amplified Thermopsin Gene with TP1/TP2 Primer Set and Selection of Recombinant Colonies

First primer set, TP1 and TP2, was designed referring to the predicted thermopsin gene sequence of *Tp. volcanium*, as described in the Materials and Methods. Forward primer (TP1) extended from the position 248052 to 248080 and reverse primer (TP2) extended from the position 251110 to 251132 and this PCR amplification had previously been done in our laboratory as well as cloning of the 3080bp PCR product to pDrive vector. Two recombinant clones were obtained. In the recombinant pDriveSH/10/26 and pDriveSH/10/8 plasmids, the PCR product was placed in reverse orientation relative to LacZ promoter in pDrive vector. Schematic representation of pDrive SH/10/8 plasmid is shown in Figure 3.2. Since, thermopsin gene did not have its own promoter and regulatory regions expression of thermopsin enzyme did not occur. To achieve the expression of cloned thermopsin gene and to construct an efficient expression construct QIAGENexpressionist Kit was used.

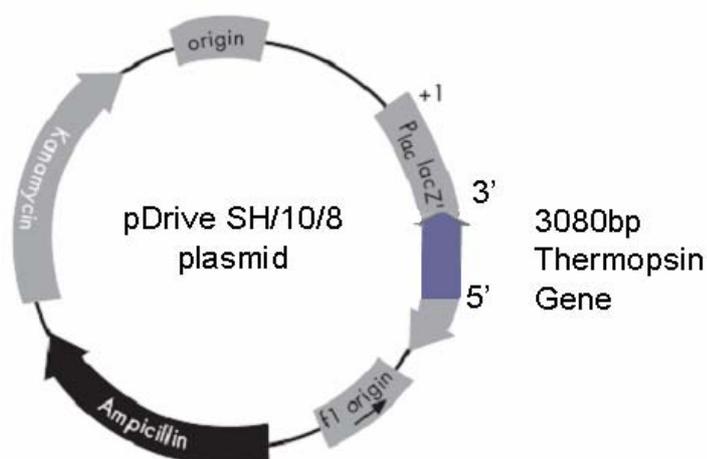


Figure 3.2: Schematic representation of pDrive SH/10/8 plasmid.

The QIAexpress system is based on selectivity and affinity of Ni-NTA (nickel-nitrilotriacetic acid) matrices for biomolecules which have been tagged with 6 consecutive histidine residues . There are several advantages to use this kit. For example, one step purification is allowed; tag does not interfere with the structure and function of recombinant protein and mild elution conditions can be used to isolate recombinant protein. 6XHis affinity tag should be placed N-terminus of the protein of interest.

Firstly, recombinant plasmids pDriveSH/10/26 and pDriveSH/10/8 were isolated by using Promega DNA purification Kit (Promega Corporation; Madison, WI). Agarose gel results of the plasmid DNA isolation are shown in Figure 3.3. Plasmids were digested with *Sph*I and *Hind*III restriction enzymes.

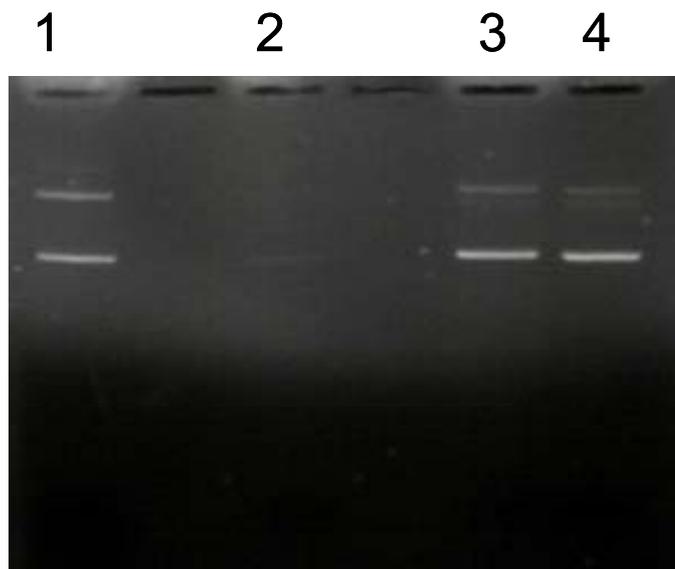


Figure 3.3: Recombinant plasmid isolation. Lane 1 and 2, recombinant pDriveSH/10/8 plasmid; lane 3 and 4, recombinant pDriveSH/10/26 plasmid. Promega DNA isolation kit was used for plasmid isolation experiments.

Thermopsin gene sequence does not have cut sites for *SphI* and *HindIII*, but there was on site for each in MCS of the pDrive vector. Therefore, this digestion released the cloned thermopsin gene from recombinant vectors. The digestion products were 3080bp PCR fragment and 3851bp linear pDrive vector (Figure 3.4 and 3.5.a). On the other hand, there is one restriction site for *BamHI* in the vector's MCS and one within the thermopsin gene (in position 1120) (Figure 3.4.b). Thus as expected two bands in lengths of 4971bp and 1960bp, were observed on the gel as *BamHI* digestion products (Figure 3.4 and 3.5.b).



Figure 3.4: Restriction digestion of recombinant thermopsin gene. Lane1, single digestion with *BamHI* (pDriveSH/10/26); lane2 (pDriveSH/10/26) and 4 (pDriveSH/10/8) double digestions with *SphI* and *HindIII* and lane 3, *EcoRI* and *HindIII* cut Lambda DNA molecular size marker (MBI Fermentas AB, Vilnius, Lithuania).

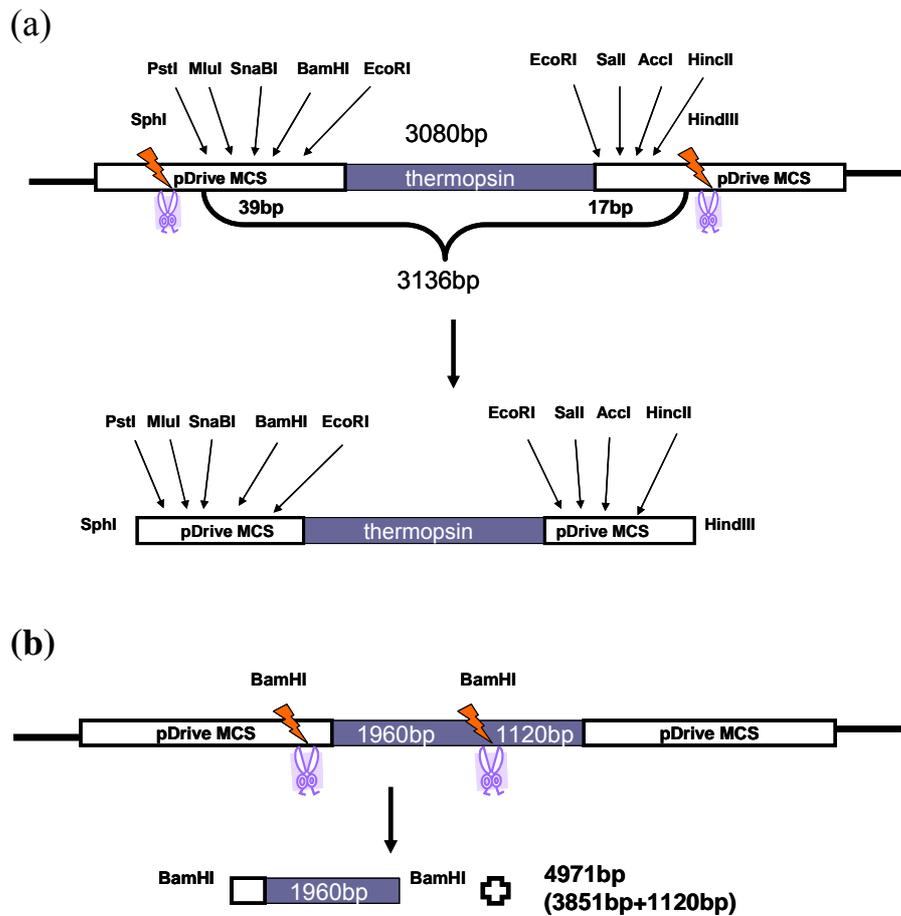


Figure 3.5: Restriction digestion profile of cloned thermopsin gene in pDrive vector. (a): Double digestion of recombinant pDrive SH/10/8 plasmids by *SphI* and *HindIII*. (b): Digestion of recombinant pDrive SH/10/8 plasmid by *BamHI*.

After the restriction analysis of recombinant plasmids we decided to use recombinant pDriveSH/10/8 plasmid for expression studies.

First we sub-cloned the 3080bp cloned thermopsin gene fragment into pUC18 vector. To achieve this, we digested both recombinant pDrive SH/10/8 plasmid and pUC18 vector with *SphI* and *Sall*. After digestion, pUC 18 vector, was

observed as a single band on agarose gel (Figure 3.6). Both of the restriction enzymes were non-cutter for thermopsin gene so we isolated the thermopsin gene fragment (3080bp) from agarose gel.

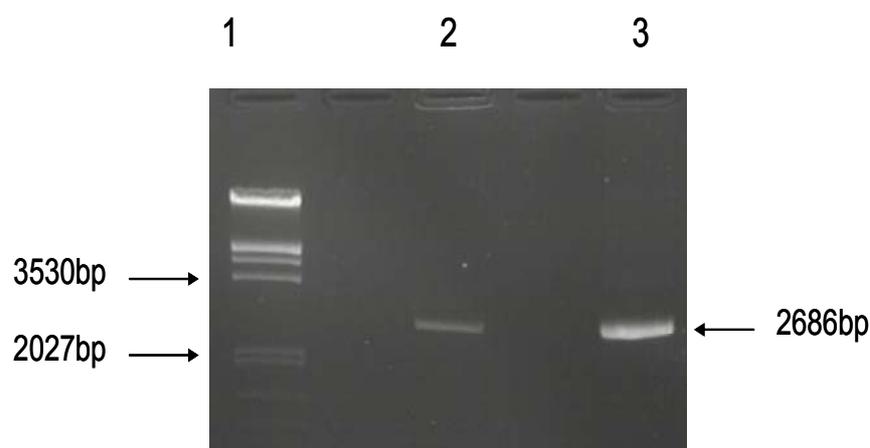


Figure 3.6: *SphI* and *SalI* digested pUC18 vector. Lane1, *EcoRI/HindIII* cut molecular size marker; lane2 and 3, *SalI* and *SphI* cut linear pUC18 vector.

The thermopsin gene isolated from gel was ligated to pUC18 through complementation of their *SphI* and *SalI* sticky ends and then they transferred into competent TG1 cells (transformation frequency: 8×10^6 transformants/ μg vector DNA). This directional cloning would place 5' end of the thermopsin gene downstream of the LacZ promoter of the pUC18 vector.

The selection of recombinant colonies was performed according to blue-white colony screening method. Since cloning was directional one, all white colonies were expected to be recombinants. Selected putative recombinant colonies were analyzed by plasmid isolation and restriction enzyme digestion to confirm the cloning of thermopsin gene fragment.

When the competent cells were transformed by pDrive Cloning Vector after ligation reaction, the transformation efficiency was 4×10^5 transformants/ μg vector DNA and recombination efficiency was 3×10^4 transformants/ μg vector DNA.

The 1 recombinant plasmid, pUC18-BST1 from the only white colony (Figure 3.7), proved to be a true recombinant by restriction analysis.



Figure 3.7: Plasmid analysis by agarose gel electrophoresis. Recombinant pUC18-BST-1 vector.

There is just one cut site for *Hind*III on the MCS of the pUC18 vector DNA, while there is not any site on the thermopsin gene. Digestion with *Hind*III linearized the recombinant pUC18-BST1, which was observed as a 5776bp fragment (2686bp vector+3080bp gene fragment) on the gel (Figure 3.8). Cloning of thermopsin gene in to pUC18 vector is schematized in figure 3.9 (a) and (b).

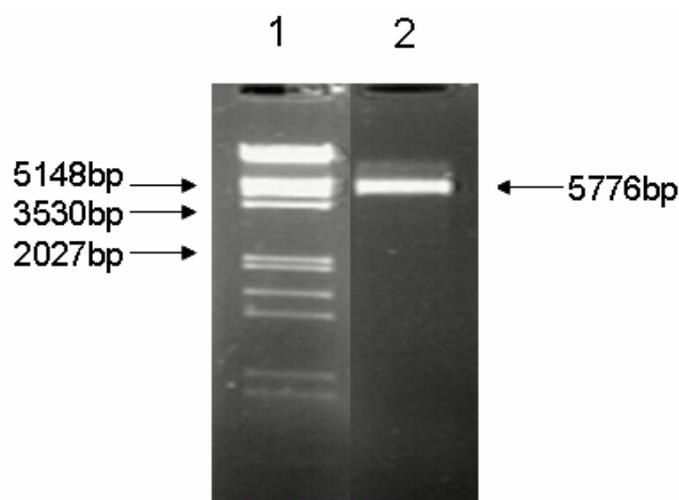


Figure3.8: Restriction digestion of recombinant pUC18-BST1 plasmid DNA with *HindIII*. Lane1, *EcoRI/HindIII* cut molecular size marker; lane2, linear recombinant pUC18-BST1.

3.1.3.2 Characterization of Recombinant Plasmids by Restriction Digestion

Restriction enzymes were determined by the help of Restrictionmapper program and shown in Table 3.3. After identification of the pUC18-BST1-5 plasmids as the true recombinant, we digested the plasmids with *SalI* and *HindIII* for further confirmation.

Both of the enzymes were non-cutter for thermopsin gene but their cut sites were available in the flanking sequences (Figure 3.9.b). So with these enzymes we could obtain two fragments in sizes of 3080bp (the insert) 2686 bp (vector) (Figure3.10). We have also digested the recombinant plasmid with 2 more enzymes, *BamHI*, and *EcoRI*. Single digestion with *BamHI* gave rise to 3 bands on agarose gel which are in lengths of 1120 bp, 1960 bp and 2686 bp since there

was single cut site for *Bam*HI in the cloned gene sequence (at position 1120bp), as well as one for each of the flanking MCSs of the vector (Figure3.9.b and Figure 11). Thermopsin gene was placed between the two *Eco*RI cut sites in the MCS of pDrive; and also there are two internal cut sites for *Eco*RI at position 1150 and 1722 of thermopsin gene (Figure 3.12). Moreover there was one more cut site for *Eco*RI in the MCS of pUC18 vector. After digestion of pUC18-BST1-5 with *Eco*RI 5 fragments were obtained (50bp, 2686bp, 1358bp, 1150bp and 572bp). Restriction digestion results on agarose gel were shown in Figure 3.11. Restriction sites and results also were summarized in Table 3.3 and Figure 3.12. Restriction digestion analysis results show correlation with the expected results Table 3.3.

Table3.3: Table of the cut site positions of restriction enzymes in 3080bp PCR product of thermopsin gene determined by restriction digestion and gel electrophoresis.

Cutter Restriction Enzymes	Cut positions
<i>Eco</i> RV	273
<i>Pvu</i> II	2947
<i>Bam</i> HI	1120
<i>Sac</i> I	1827
<i>Eco</i> RI	1150, 1722

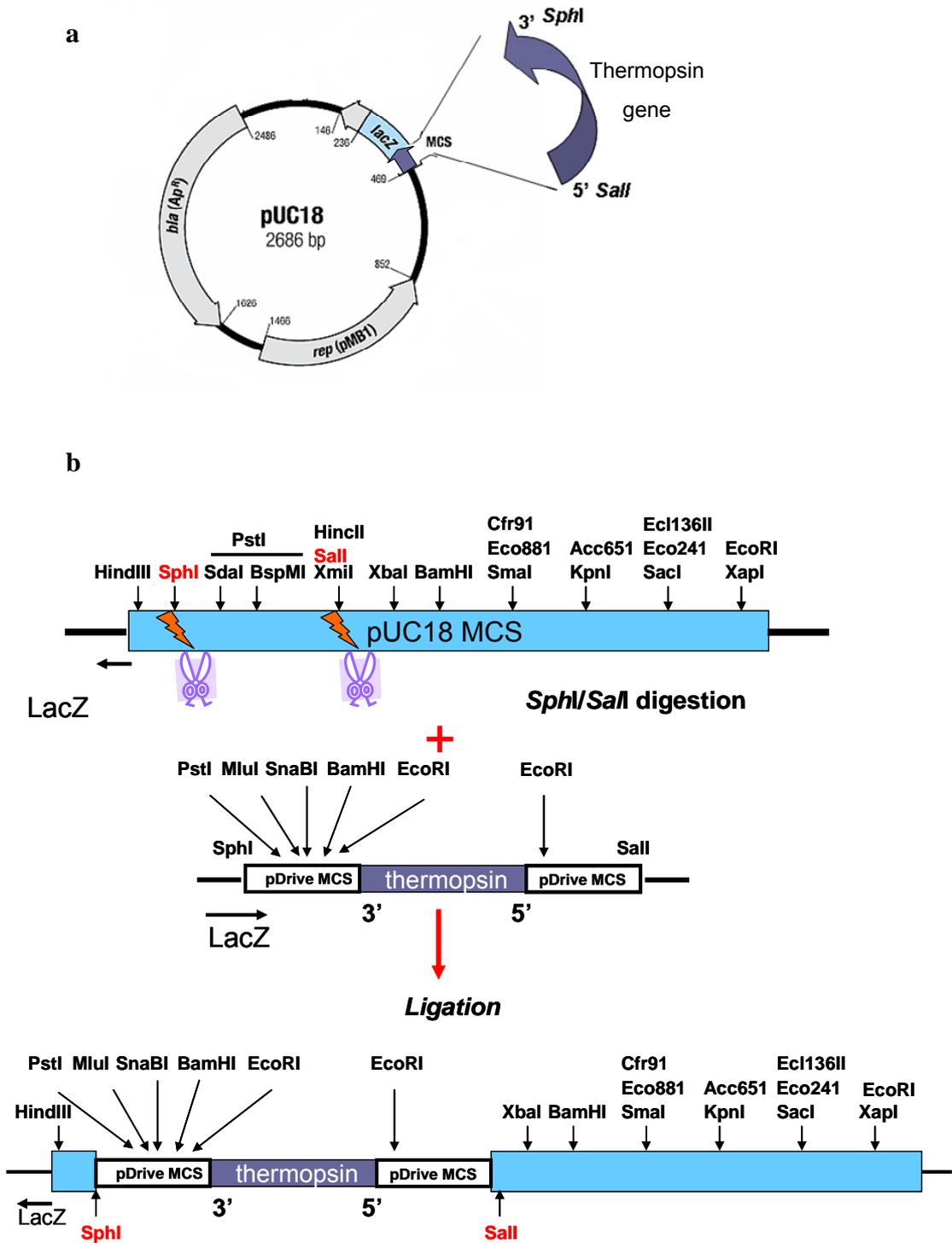


Figure 3.9 (a): Schematic representation of thermopsin gene insertion into the pUC18 vector. **(b):** Ligation scheme of thermopsin gene to pUC18 vector DNA at *SphI* and *SalI* sites.

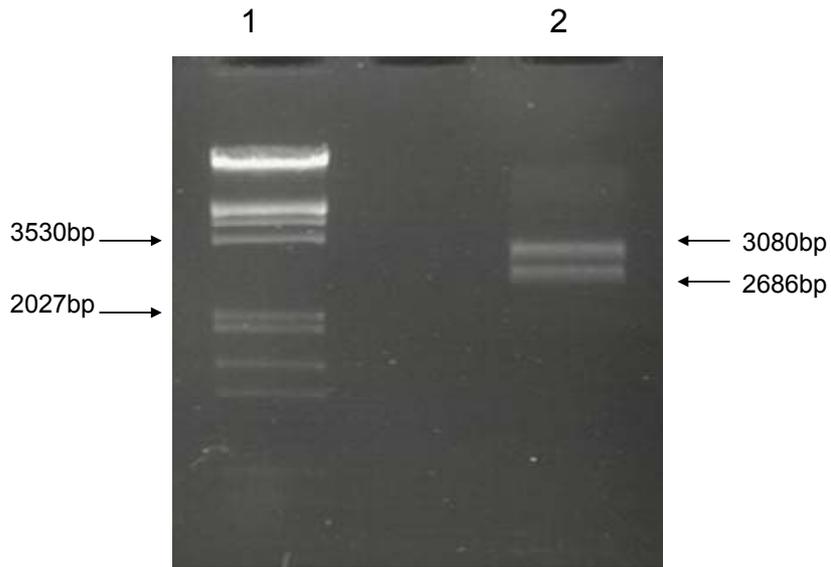


Figure3.10: Restriction digestion of recombinant pUC18-BST1-5 plasmid with *HindIII* and *SalI*. Lane1, *EcoRI/HindIII* cut molecular size marker; lane2, *SphI* and *SalI* cut plasmid resulted 3080bp thermopsin gene and 2686bp pUC18 vector.

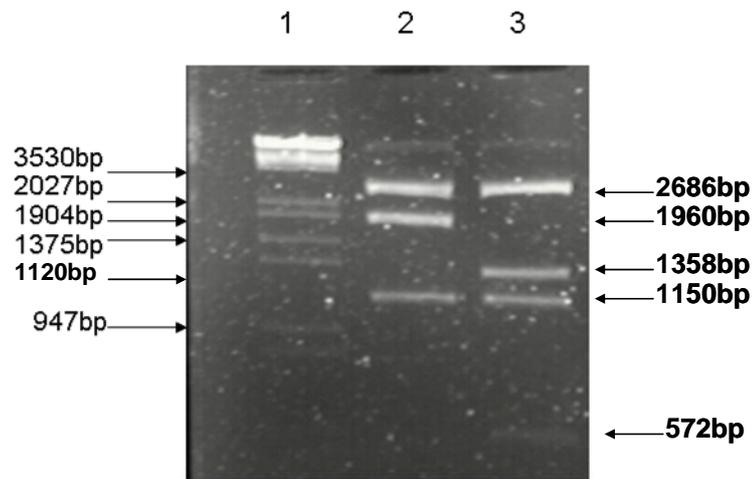


Figure3.11: Restriction digestion of recombinant pUC18-BST1-5. Lane1, *EcoRI/HindIII* cut molecular size marker; lane2, *BamHI* cut; lane3, *EcoRI* cut.

After this cloning step we planned subcloning of 3080bp thermopsin gene into pQE vector for expression by fusion of the thermopsin gene at the 5' end with the 6XHis tag .

3.1.4.1 PCR Amplification and Cloning of Thermopsin Gene with TP1'/TP2 Primer Set

The primer set TP1'/TP2 was designed according to the QIAexpress Kit instructions. Initiation of translation at internal start sites can occur when a ribosome binding consensus sequence (Shine-Dalgarno sequence) is present 5' to an internal ATG or GTC codon. To produce an efficient N-terminal His tag fusion translation should be from the start codon of 6XHis tag. Forward primer (TP1') extending from the position 248071 to 248095 and reverse primer (TP2, which was same as the reverse primer of the first primer set) extending from the position 251110 to 251132, were designed to conform this requirement. PCR amplification was carried out at two different annealing temperatures, 55⁰C and 58⁰C. We obtained clear single band at each annealing temperature representing approximately 3070bp PCR product which was consistent with the expectation (Figure 3.13).

3.1.4.2 Cloning of PCR Amplified 3070bp Thermopsin Gene and Selection of Recombinants

PCR amplification with the TP1' and TP2 resulted in a 3070 bp fragment containing thermopsin structural gene missing initiation codon at 5' end. In cloning we have used Qiagen PCR Cloning Kit (QIAGEN PCR Cloning Kit, QIAGEN Inc. Valencia, USA).

Figure 3.14 shows scheme for ligation of the PCR amplified thermopsin gene to pDrive vector.

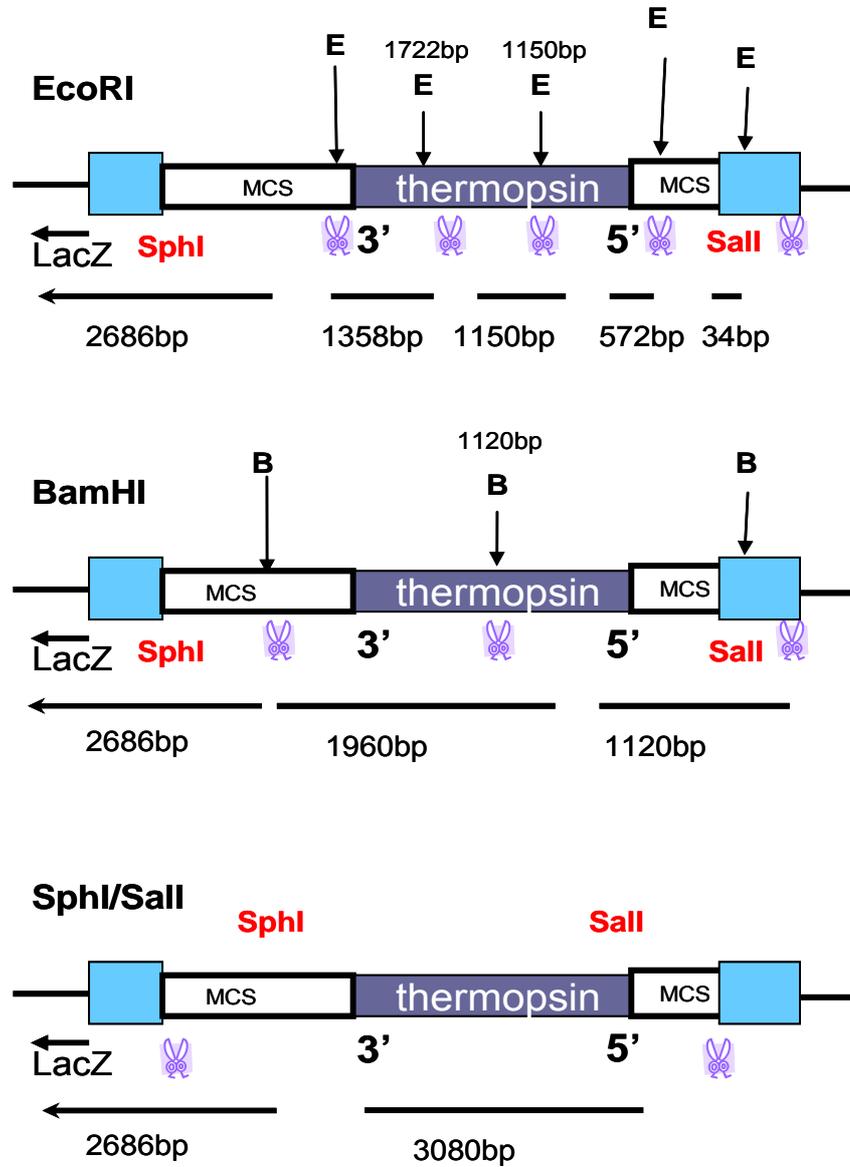


Figure3.12: Schematic representation of restriction digestion patterns of recombinant pUC18-BST1-5 plasmid with; *EcoRI*, *BamHI*, *SphI* and *Sall* as revealed by fragment size determination on agarose gel.

When cloning was carried out with Qiagen Cloning Kit and *E. coli* TG-1 competent cells were used as recipient cells in the transformation experiments. The transformation efficiency of these competent cells was calculated as 8×10^6 transformant/ μg vector DNA when prepared.

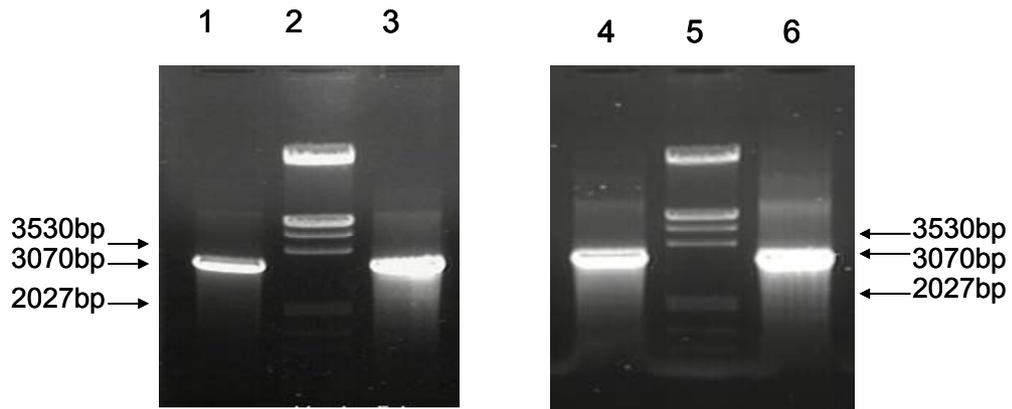


Figure 3.13: PCR amplification of thermopsin gene (3070bp) from *Thermoplasma volcanium* at different annealing temperatures. Lane 1 and 2, PCR amplification at 55°C ; lane 4 and 6, PCR amplification at 58°C ; lane 2 and 5, *EcoRI*/*HindIII* cut λ -DNA size marker.

The selection of recombinant colonies was according to blue/white screening.

When the competent cells were transformed by pDrive Cloning Vector after ligation reaction, the transformation efficiency was 2×10^4 transformants/ μg vector DNA and recombination efficiency was 8×10^2 transformant/ μg vector DNA.

A total of 18 white colonies were obtained in a given cloning experiment. Plasmid DNAs were isolated from 10 of these putative recombinant colonies, to check for the presence of cloned thermopsin gene (Figure 3.15).

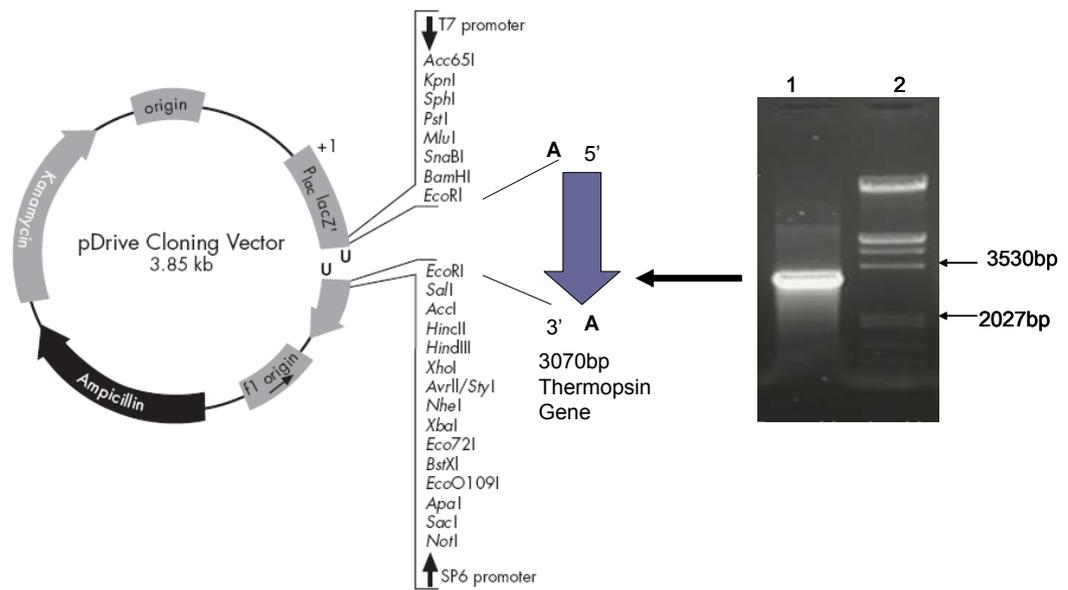


Figure 3.14: Schematic representation of ligation of PCR amplified thermopsin gene (3070bp) to pDrive vector. Lane1, 3070bp PCR product; lane 2, *EcoRI/HindIII* cut molecular size marker (λ DNA).

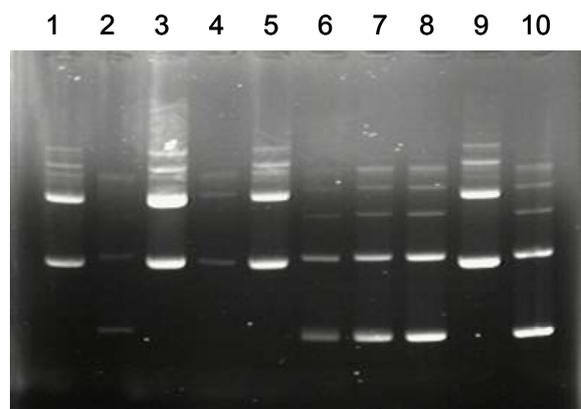


Figure3.15: Putative recombinant plasmids isolated from white colonies by Promega DNA Isolation Kit.

3.1.4.3 Characterization of Recombinant Plasmids

According to the data from Restrictionmapper version3 program, restriction map of the PCR amplified 3070bp thermopsin gene is shown in table 3.4.

Restriction digestions were performed to find out true recombinant pDrive vectors which would contain 3070bp PCR product. Firstly, out of ten putative recombinant plasmids were selected to be digested with *SphI* and *SalI* restriction enzymes. Both of the enzymes were non-cutter for thermopsin gene but they had one cut site in the pDrive vector each in one of the flanking MCS. Thus, this double digestion is expected to yield two bands on agarose gel, representing pDrive vector (3851bp) and PCR product (3070bp) (Figure3.16).

Table 3.4: The cut site positions of some restriction enzymes in 3070bp PCR product. These predicted cut site positions were obtained by Restriction digestion and gel analysis of the fragment.

Cutter Restriction Enzymes	Cut positions
<i>EcoRV</i>	264
<i>PvuII</i>	2938
<i>BamHI</i>	1111
<i>SacI</i>	1818
<i>EcoRI</i>	1141, 1713

The results showed that out of 5 plasmids checked with restriction digestion, four (pDriveBSTP'-1, pDriveBSTP'-3, pDriveBSTP'-5, pDriveBSTP'-9) proved to be recombinants. Due to unexpected digestion profile pDriveBSTP'-8 is not a recombinant plasmid. pDriveBSTP'-1, pDriveBSTP'-3, pDriveBSTP'-5 and pDriveBSTP'-9 plasmids were selected for further characterization studies.

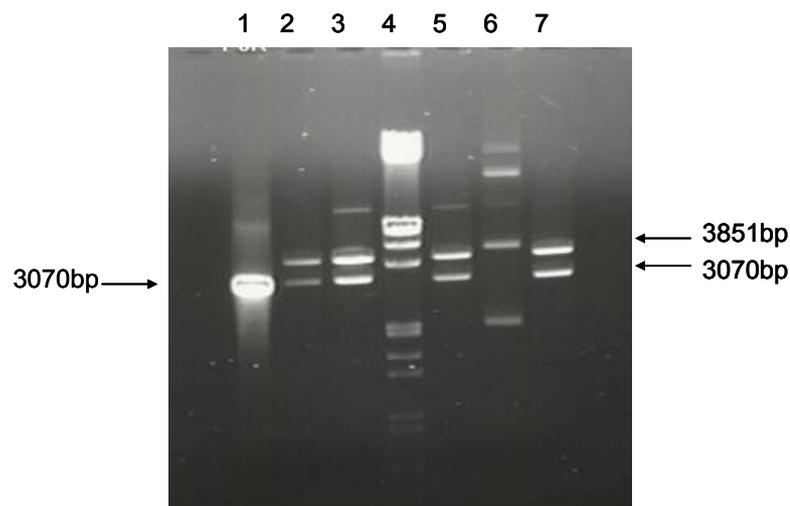


Figure 3.16: Restriction digestion of putative recombinant plasmids to confirm the cloning of thermopsin gene. Screening for the true recombinant was done by *SphI* and *SalI* double digestion of pDriveBSTP'-1 (lane2), pDriveBSTP'-3 (lane3), pDriveBSTP'-5 (lane5), pDriveBSTP'-8 (lane6), pDriveBSTP'-9 (lane7). Lane1, 3070bp PCR product, lane4, *EcoRI/HindIII* cut molecular size marker (λ DNA).

pDriveBSTP'-1, pDriveBSTP'-3, pDriveBSTP'-5 and pDriveBSTP'-9 plasmids were digested with different restriction enzymes. In restriction analysis, first enzyme used was *PvuII*. *PvuII* had an internal cut site at position 2938 in thermopsin gene and two cut sites at positions 54 and 517 in pDrive vector (Figure 3.17). We expected to see three bands on agarose gel in length of 3388bp, 3150bp, and 256bp, if insertion is in forward direction. For pDriveBSTP'-3 we saw bands with appropriate sizes. But, just based on this result we could not decide about the thermopsin gene is direction. *SacI* was the second restriction enzyme used in analysis which had an internal cut site at position 1818 in thermopsin gene and also one at MCS of pDrive vector (Figure 3.17). According to direction of insert there were two possibilities: if thermopsin gene placed in the

same direction with *LacZ* promoter, there would be two bands in sizes of 1252bp and 5669bp (Figure 3.17). If thermopsin gene was placed in the reverse direction relative to *LacZ*, there would be two bands in sizes of 1818bp and 5103bp. After the digestion of pDriveBSTP'-3 and pDriveBSTP'-1 plasmids with *SacI* two fragments in the sizes of 1818bp and 5103bp, were observed on the gel (Figure 3.18) However, digestion of pDriveBSTP'-5 and pDriveBSTP'-9 plasmids produced two fragments in the sizes of 1252bp and 5669bp. Other restriction enzyme used is *BamHI* which had an internal cut site at position 1111 in thermopsin gene and also one in MCS of pDrive vector (Figure 3.17). When, pDriveBSTP'-3 and pDriveBSTP'-1 plasmids were digested with *BamHI* two bands in the sizes of 1959bp and 4962bp were visualized on agarose gel indicating reverse orientation of the insert in these plasmids with respect to *LacZ* (Figure 3.18). pDriveBSTP'-5 and pDriveBSTP'-9 plasmid digestion with *BamHI* two bands in the sizes of 1111bp and 5810bp were visualized on the gel which shows that insertion is in the forward direction (Figure 3.18). *EcoRV* had an internal cut site at position 264 in thermopsin gene, and one cut site on the vectors MCS (Figure 3.17). The digestion of pDriveBSTP'-5 plasmid with *EcoRV*, yielded a linearized plasmid DNAs in 6921bp length equal to the, sum of the insert and vector and probably cut site on the vector was lost during ligation (Figure 3.19). *EcoRI* had internal cut sites at positions 1141 and 1713 in thermopsin gene and also at two cut positions in the flanking MCS, of pDrive vector (Figure 3.17) we expected to see four bands in the sizes of 3851bp, 1357bp, 1141bp, 572bp regardless the insertion direction (Figure 3.17). Indeed we saw three bands at the expected positions when the plasmids pDriveBSTP'-5 and pDriveBSTP'-9 were cut with *EcoRI* (Figure 3.18).

The restriction digestion profiles obtained by size determination of the fragments on agarose gel was in good agreement with the restriction map derived from Restrictionmapper Version3 program (Table 3.1 and 3.4)

Based on the results of the restriction analysis we decided that; thermopsin gene is placed reverse direction with respect to *LacZ* in the pDriveBSTP'-3 and

pDriveBSTP'-1 plasmids. On the other hand, thermopsin gene placed in the same direction as *LacZ* region in the pDriveBSTP'-5 and pDriveBSTP'-9 plasmids. For expression with QIA Expression Kit it was required that the 5' end of the thermopsin gene should be placed downstream (3' end) of the 6xHis tag in the expression vectors. Since pDriveBSTP'-5 plasmid had the suitable restriction sites for sub-cloning of the thermopsin gene to pQE vector set in the appropriate position, this recombinant plasmid was used in the expression studies.

3.1.5 Sub-cloning of Thermopsin Gene for Expression

Expression experiments were carried out using by QIAexpression Kit (QIAGEN Inc.; Valencia, USA). Expression kit contained the expression vector set pQE 30-31-32, host strains *E.coli* M15[pREP4] and SG130009[pREP4] strains. These host strains contain pREP4 plasmid which was required for regulation of expression by lacI protein. The extremely high transcription rate initiated at the T5 promoter can only be efficiently regulated and repressed by the presence of high levels of the lac repressor protein. This repressor protein can be used in *trans* or *cis* to the gene to be expressed. In *trans* system, the host strains contain the low-copy plasmid pREP4 which contains kanamycin resistance gene and expresses lac repressor protein encoded by the *lac I* gene (Farabaugh, 1978)

The pREP4 plasmid is derived from pACYC and contains p15A replicon. Multiple copies of pREP4 are present in the host cells so high levels of lac repressor protein exists and binds to the operator sequences and tightly regulates recombinant protein expression.

High level expression of 6Xhis tagged proteins in *E. coli* using pQE vectors is based on the T5 promoter transcription-translation system. pQE vectors (pQE 30-31-32) are derived from pDS family of plasmids (Bujard *et al.*, 1987). They have three different open reading frame and 6Xhis tag coding sequence 5' to the cloning region (Figure 3.19).

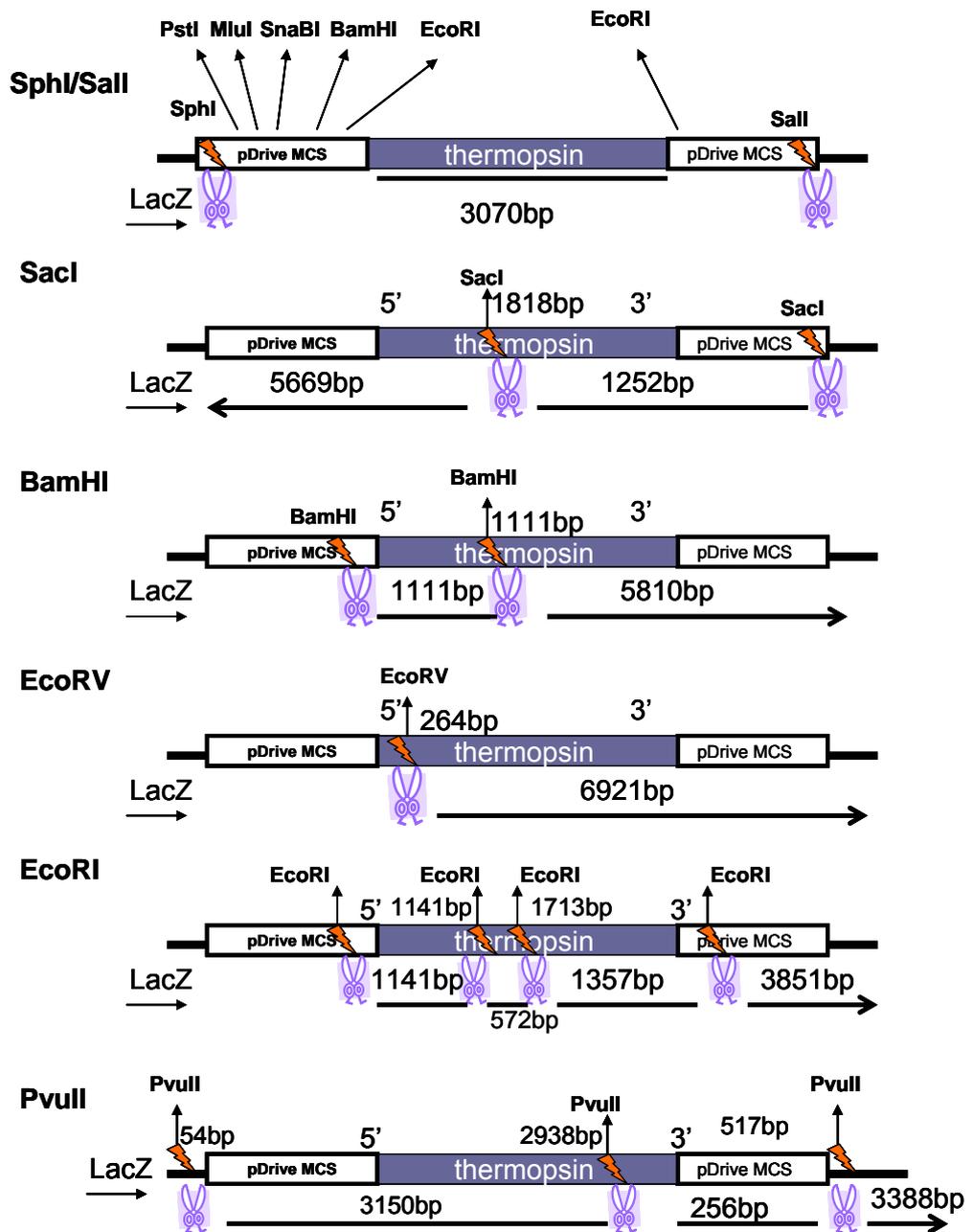


Figure3.17: Schematic representation of restriction digestion of recombinant pDriveBSTP'-5 vector. *SphI/SalI*, *SacI*, *BamHI*, *EcoRV*, *EcoRI*, *PvuII* cut positions and plasmids after digestions.

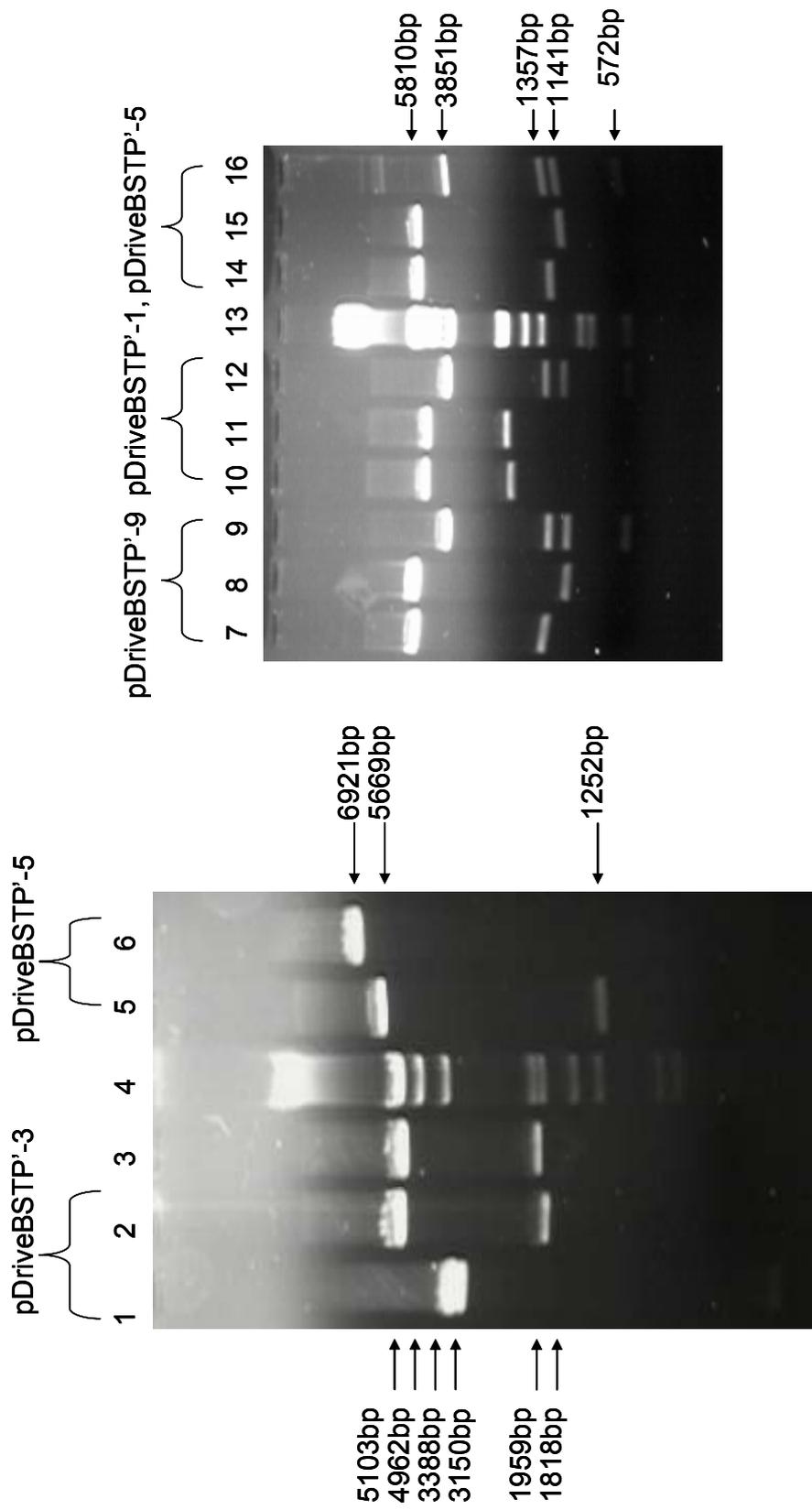


Figure 3.18: Restriction digestion of recombinant pDrive vectors. Lane 1 *PvuII*; lane 2, 5, 7, 10 and 14 *SacI*; lane 3, 8, 11 and 15 *BamHI*; lane 6, *EcoRV*; lane 9, 12 and 16 *EcoRI* digestions. Lane 4 and 13, *EcoRI/HindIII* cut molecular size marker.

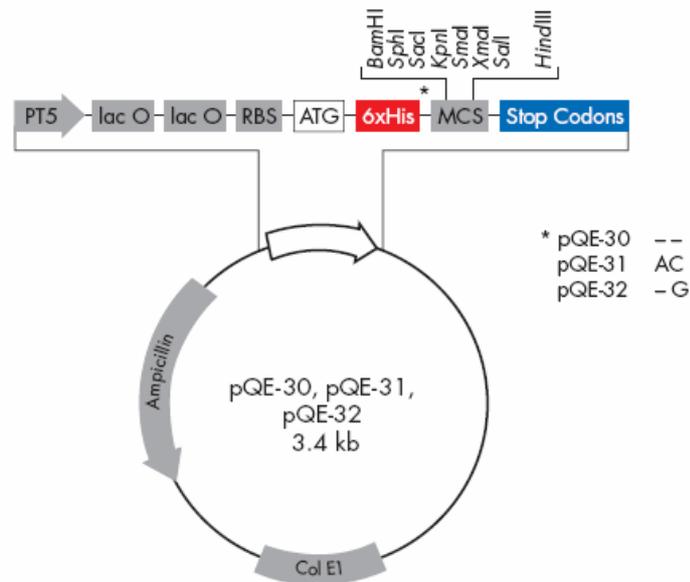


Figure 3.19: Schematic representation of pQE 30-31-32 expression vectors.

3.1.5.1 Sub-cloning of Recombinant Thermopsin Gene from pDriveBSTP'5 Plasmid to pQE Expression Vectors

Sub-cloning of the thermopsin gene from pDriveBSTP'-5 plasmid to pQE triple vector set was performed according to QIAexpression kit instructions. The first step was to find out proper restriction enzymes that would enable the ligation of 5' end of the thermopsin gene to 3' end of 6XHIS tag. For this purpose, *SphI* and *SalI* enzymes were suitable. Double digestion with these enzymes excised thermopsin gene from pDriveBSTP'-5 which was inserted into expression vectors cut with the same enzymes. To this, pQE vector set was digested with *SalI* and *SphI* which had single cut site at MCS of pQE vectors (Figure 3.20). Three pQE vectors and recombinant pDriveBSTP'-5 plasmid were double digested separately and the digestion mixtures were run on low melting point agarose gel. Then cut vector DNAs (3464 bp) and cloned 3070bp thermopsin fragments were extracted from gel (Figure 3.21). Directional ligation and subcloning procedures are schematized in Figure 3.22 and 3.23. Ligation mixture was transformed into three different *E.*

coli host strains, M15[pREP4], SG130009[pREP4] and TG1 strains. Competent cell frequencies with SG130009, M15 and TG1 were 6×10^7 , 3×10^7 , 1.5×10^9 , respectively.

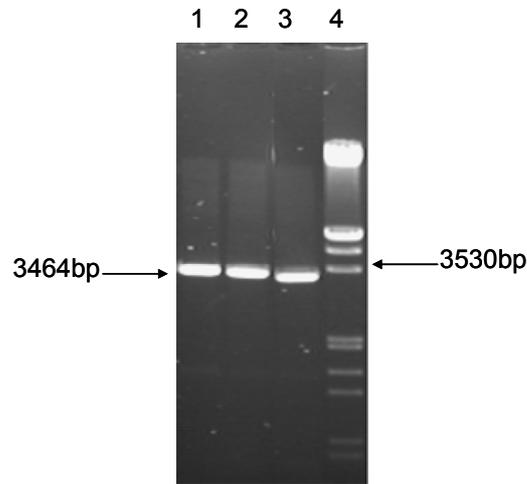


Figure 3.20: Restriction digestion of pQE vectors with SalI for linearization. Lane1, pQE 30 vector; lane2, pQE31 vector; lane3, pQE32 vector; lane4, *EcoRI/HindIII* cut molecular size marker (λ DNA).

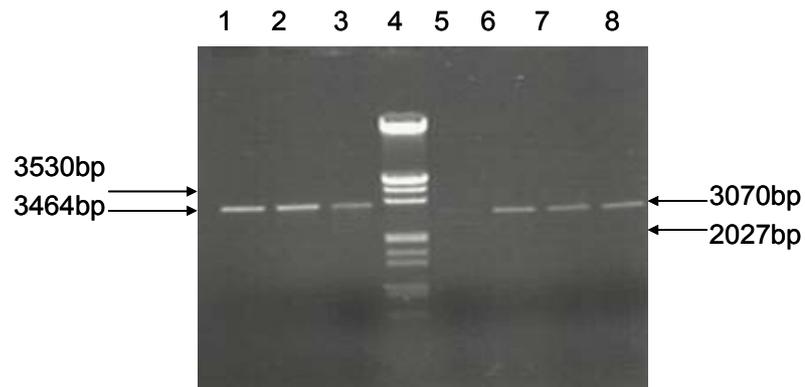


Figure 3.21: Isolation DNA fragments from agarose gel. Lane1, cut pQE30; lane2, cut pQE31; lane4, cut pQE32; lane5-6-7-8, excised thermopsin gene fragment (3070bp); lane4, *EcoRI/HindIII* cut molecular size marker (λ DNA).

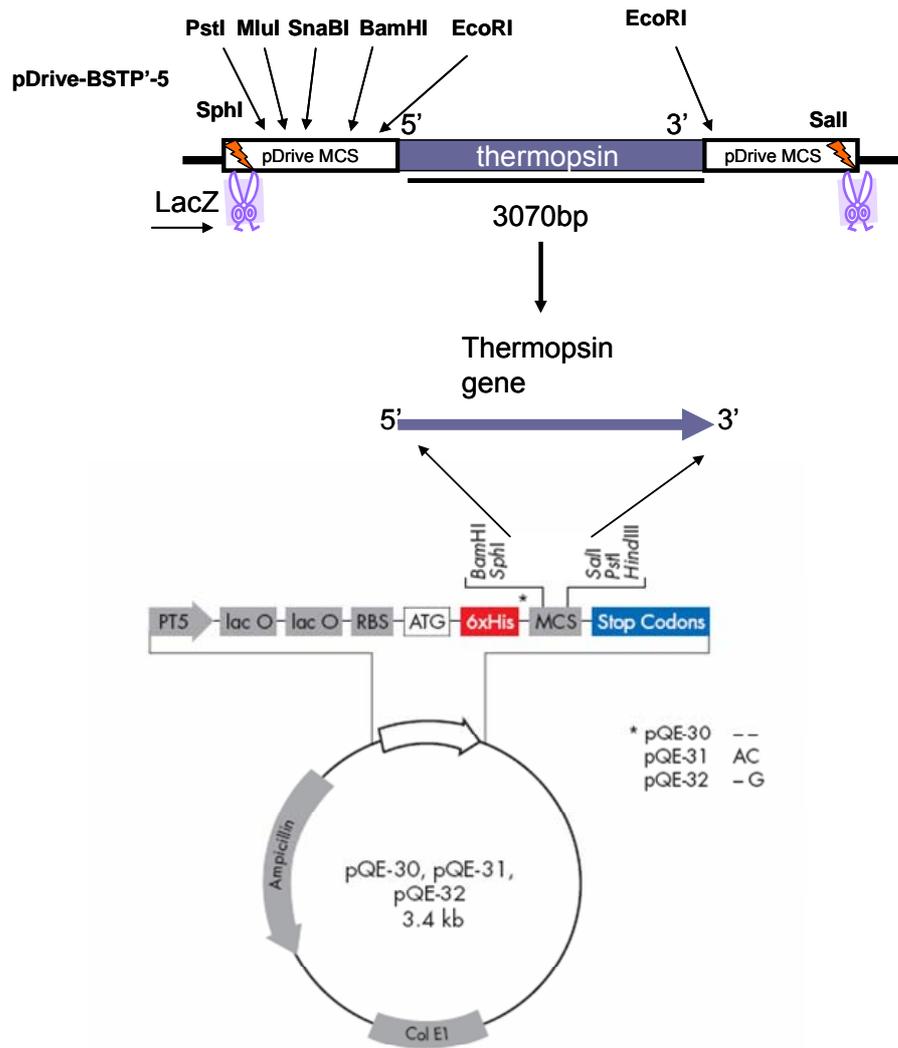


Figure 3.22: Schematic representation of sub-cloning of thermopsin to pQE vector.

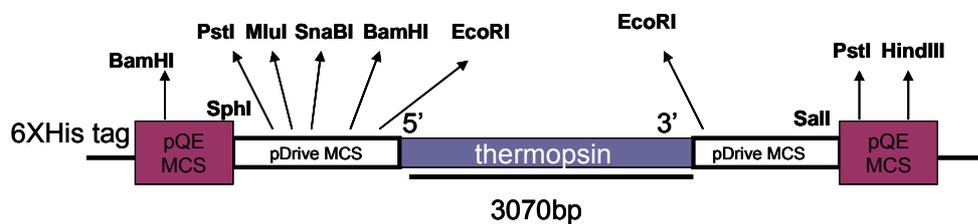


Figure 3.23: Schematic representation of expression constructs in which 5' end of the thermopsin gene placed downstream 6XHis tag.

3.1.5.1.1 Colony Blot Hybridization

pQE vectors are not suitable to be used for any color based screening method to find out recombinant colonies. In such cases colony blot hybridization method might have some advantages. This method allows simultaneous screening of transformants with the correct coding fragment, expression levels, and in frame translation of the 6XHis tag. We carried out 7 colony blot hybridization experiments for the recombinant clone.

Colony blot hybridization results with transformants of M15 recipient, indicating that 3070bp cloned thermopsin gene could be expressed in-frame in the pQE31 vector. While strong signals on the membrane blots were obtained with pQE31 ligation products (e.g., membrane #3 and #4, in Figure 3.24) there were no signal or very weak if any, with pQE30 and pQE32 ligation products (e.g., membrane #2, in Figure 3.24).

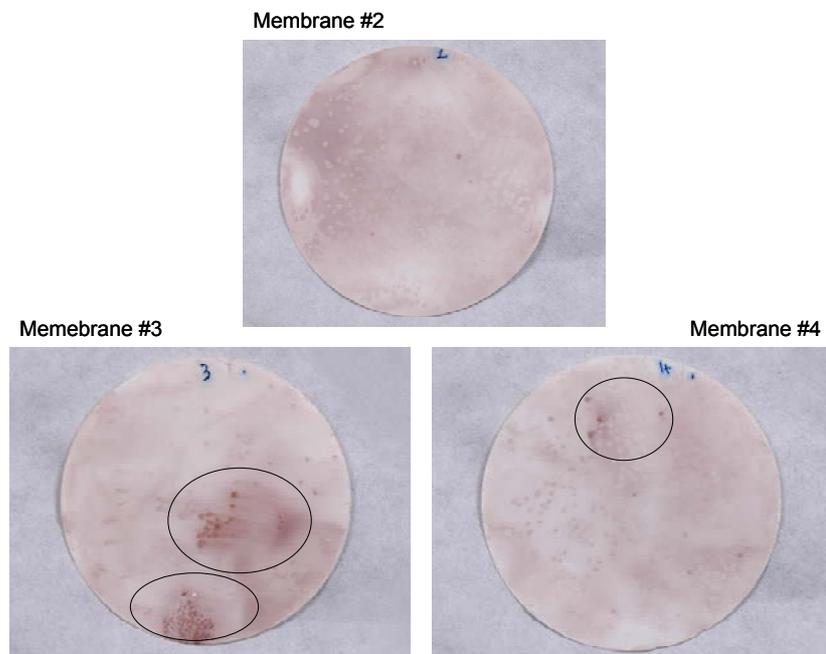


Figure 3.24: Colony blot hybridization results. Membrane#2, pQE32 colonies, membrane#3 and #4 blots of pQE31 colonies. Positive signals were in circles.

About 30 colonies were picked up from master plates, from same relative positions as compared to the positive signals on the membrane blots. Plasmids from these clones (Figure 3.25) were screened for the presence of the cloned thermopsin gene, through *SphI/SalI* digestion (Figure 3.26-29). However non of the plasmid digestion profile indicated the presence of thermopsin gene in the digested plasmids.

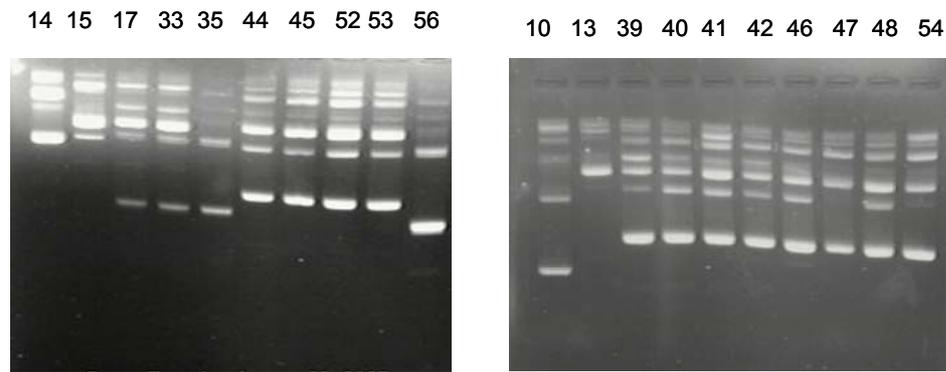


Figure 3.25: Plasmid isolation from 20 putative plasmids in the fifth colony blot hybridization experiment.

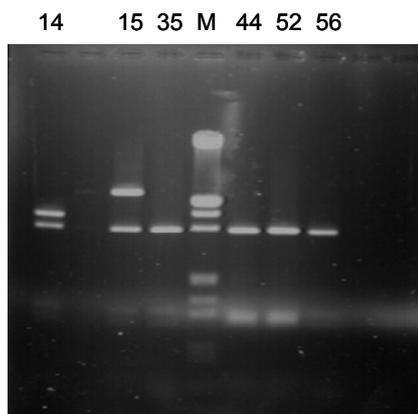


Figure 3.26: Plasmid digestion with *SphI/SalI*. Plasmid samples 14, 15, 35, 44, 52 and 56 are cut with *SphI* and *SalI*. M: *EcoRI/HindIII* cut molecular size marker.

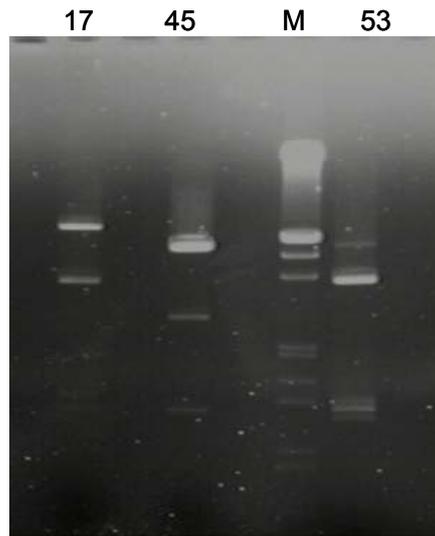


Figure 3.27: Plasmid digestion with *SphI* and/or *SalI*. Plasmid sample 17 was digested with *SalI*, sample 45 and 53 were digested with *SphI/SalI*. M: *EcoRI/HindIII* cut molecular size marker (λ DNA).

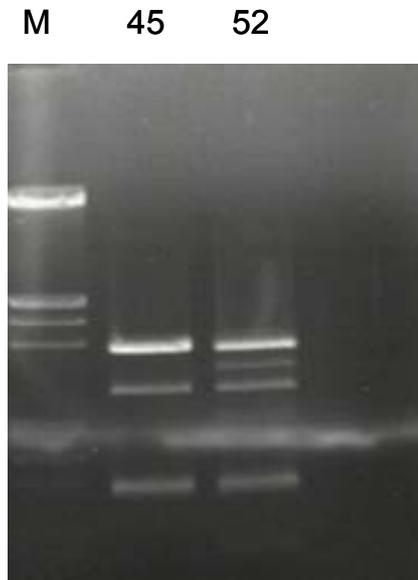


Figure 3.28 Plasmid digestion with *SphI/HindIII*. Plasmid samples 45 and 52 were cut with *SphI/HindIII*. M: *EcoRI/HindIII* cut molecular size marker (λ DNA).

13 39 41 M 47 48 54

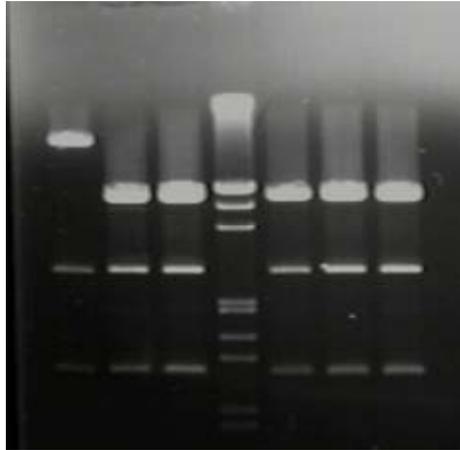


Figure 3.29: Plasmid digestion with *SalI* for linearization of the plasmids. Plasmid samples 13, 34, 41, 47, 48 and 54 were cut with *SalI*.

Then colony blot/hybridization with the transformants of TG-1 recipient cells with pQE31 recombinants were performed. In the colony blot several strong positive signals were obtained (Figure 3.30).

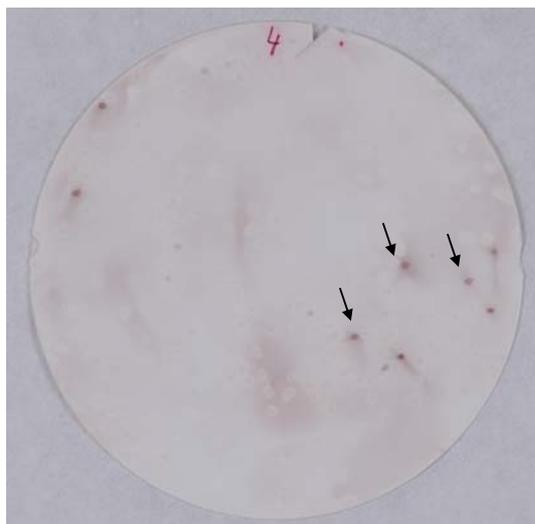


Figure 3.30: Seventh colony blot hybridization results. Arrows represent recombinant expression colonies.

Eight colonies were picked-up from the master plates, which produce the positive signals on the membrane blots. Their plasmids were isolated (Figure 3.31) and analyzed by restriction digestion and gel electrophoresis. Restriction enzyme *SalI* had a single cut site on the MCS of the pQE31 vector DNA, but a non-cutter of thermopsin gene. We expected to see clear single band in the length of 6534bp, equal to sum of the pQE vector (3464bp) and the thermopsin gene (3070bp). After *SalI* digestion only with three colonies we obtained expected result by gel electrophoresis (Figure 3.32). These three plasmids were also digested with *SphI/SalI*. These enzymes cut at MCS's flanking the inserts. The expected result of this digestion is the excision of the thermopsin gene (3070bp) from pQE31 vector (3464bp). Digestion profiles obtained with all three recombinant plasmids in good correlation with the expected results (Figure 3.33).

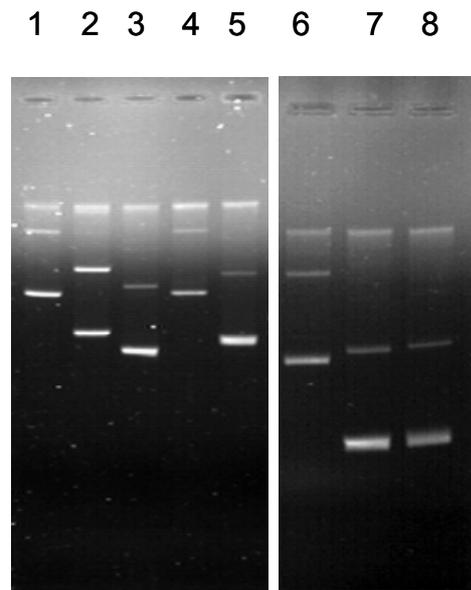


Figure3.31: Plasmid isolation from putative positive clones based on colony blot/hybridization results. Lane1, pQE31-1; lane 4, pQE31-4; lane6, pQE31-6 and other plasmids seems to be non-recombinant.

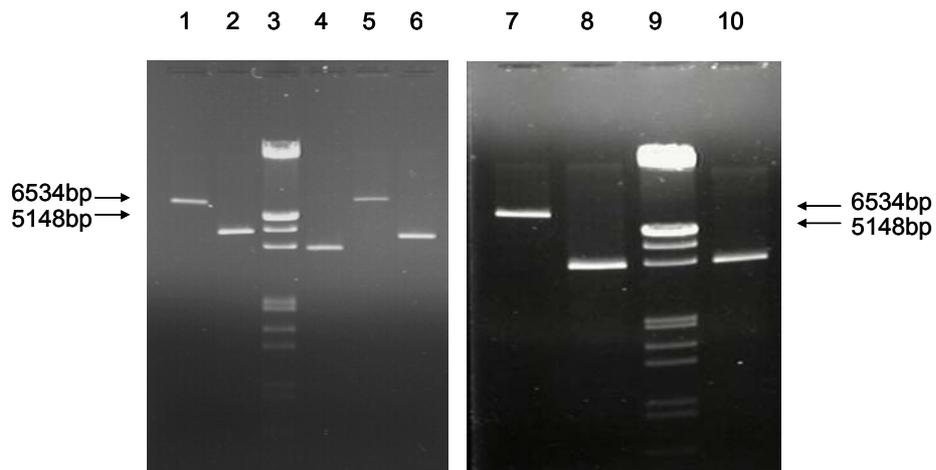


Figure 3.32 Digestion of recombinant plasmids with *SalI*. Lane1, pQE31-1; lane5, pQE31-4; and lane7, pQE31-6 were the possible recombinants. Lane3 and 9 *EcoRI/HindIII* cut molecular size marker (λ DNA).

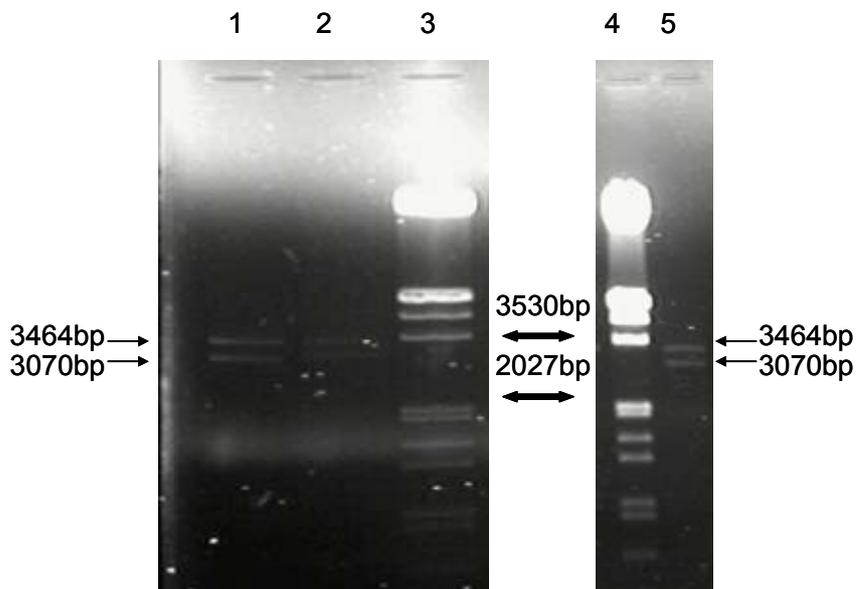


Figure 3.33 Plasmid digestion with *SphI/SalI*. Lane1, pQE31-1; lane2 pQE31-4; lane5, pQE31-6; lane3 and 4, *EcoRI/HindIII* cut molecular size marker (λ DNA).

The recombinant plasmids pQE31-1, pQE31-4, pQE31-6 were further characterized by digestion with *Hind*III. *Hind*III was non-cutter for thermopsin gene but it had single cut site in MCS of pQE vector so we expected to see single band in the size of 6534bp, and this was confirmed by agarose gel electrophoresis result (Figure 3.34).

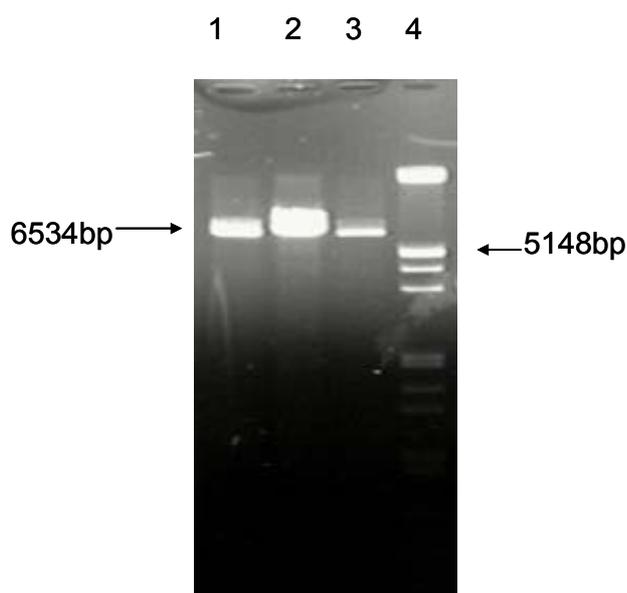


Figure 3.34: *Hind*III cut recombinant plasmids. Lane1, pQE31-1; lane2, pQE31-4; lane3, pQE31-6 and lane4, *Eco*RI/*Hind*III cut molecular size marker.

3.1.5.2 Sub-cloning of Recombinant Thermopsin Gene from pUC18-BST1 Plasmid to pQE Expression Vectors

Firstly pQE vectors (pQE30 and pQE31) and recombinant pUC18BST1 plasmids were digested with *Hind*III and *Sal*I. These enzymes were non-cutter for thermopsin gene and they had cut sites in MCS of both pQE and pUC18 vectors. After digestion released thermopsin gene and cut pQE vectors were isolated from low melting point agarose gel (Figure 3.35). Thermopsin gene and pQE vectors

were ligated and then they were transferred into *E.coli* TG1 competent cells. The schematic representation of sub-cloning of thermopsin gene from pUC18 vector to pQE vectors was shown in Figure 3.36.

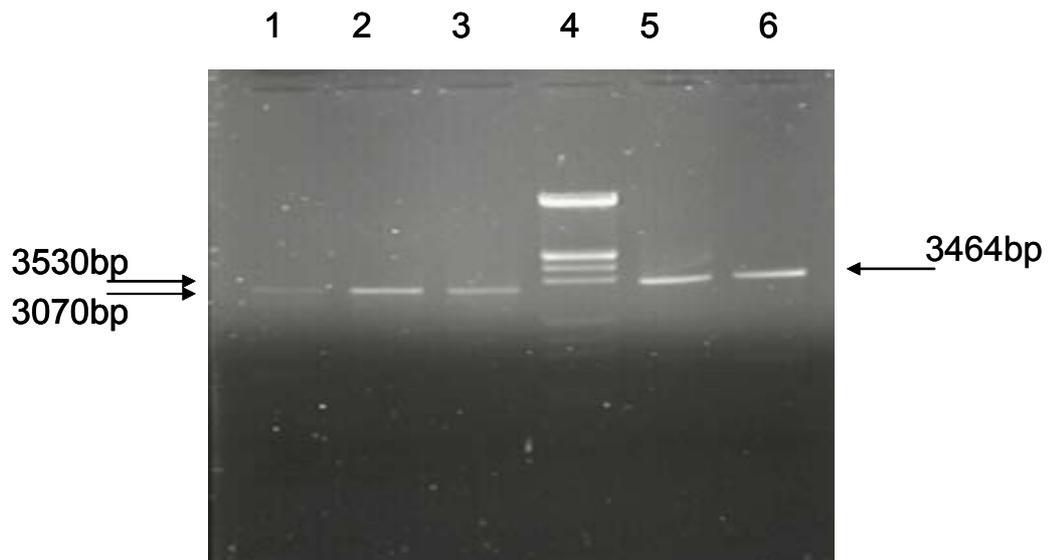


Figure 3.35: Agarose gel extraction from *SaII* and *HindIII* digestions. Lane1, 2 and 3: cloned thermopsin fragments; lane5:cut pQE30 vector; lane6:cut pQE31 vector; lane4: *EcoRI/HindIII* cut molecular size marker (λ DNA).

3.1.5.2.1 Colony Blot Hybridization

Recombinant colonies were selected by the colony hybridization method. We detected some expression signals on the membrane with the blots of recombinant pQE31 plamids. Some weak signals were observed on the membrane containing blots of recombinant pQE30 clones (Figure 3.37)

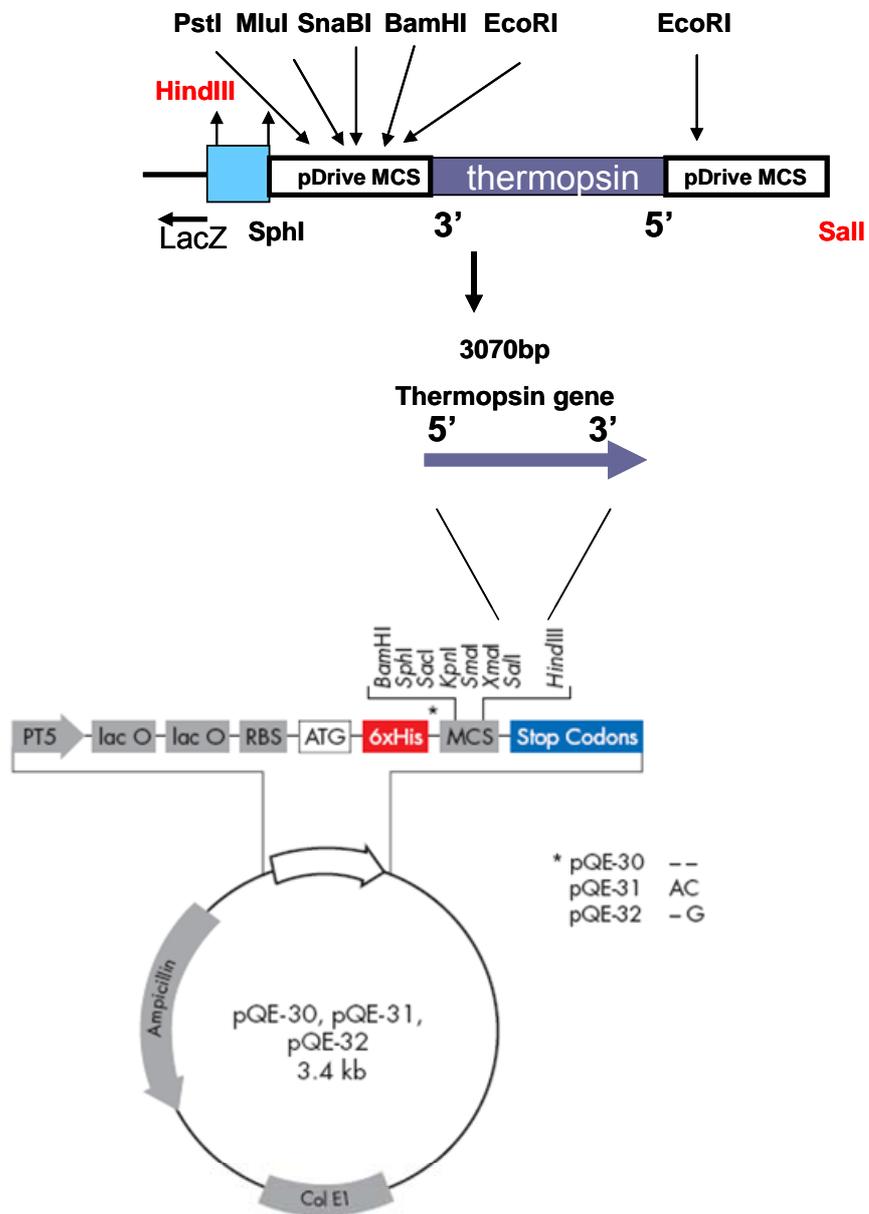


Figure 3.36: Schematic representation of sub-cloning of thermopsin gene from pUC18 to pQE30 and 31 vectors.

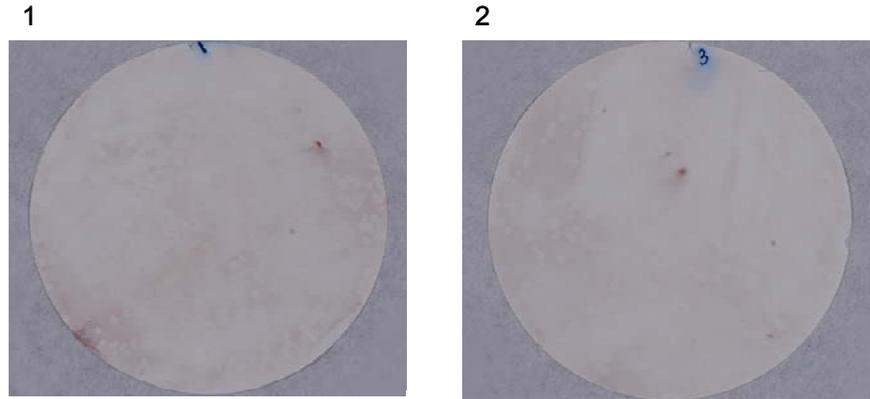


Figure 3.37: Results of colony blot hybridization experiment. First picture represents membrane 1 and second picture represents membrane 3.

3.1.4.2.2 Characterization of Recombinant Plasmids

We isolated 12 putative recombinant colonies for characterization (Figure 3.38). After the isolation of plasmids they were digested with *Hind*III (Figure 3.39) and then only one plasmid seemed to a possible recombinant which was in the length of 6534bp. Then this plasmid was double digested with *Sal*I/*Hind*III. We expected to see two bands in the length of 3070bp and 3464bp but unfortunately we saw two bands with estimated size of 4200bp and 3500bp (Figure 3.40).

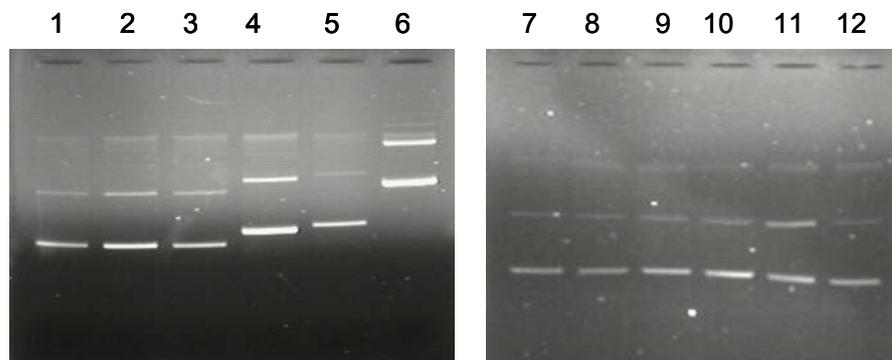


Figure 3.38: Plasmid isolation of putative expression colonies.

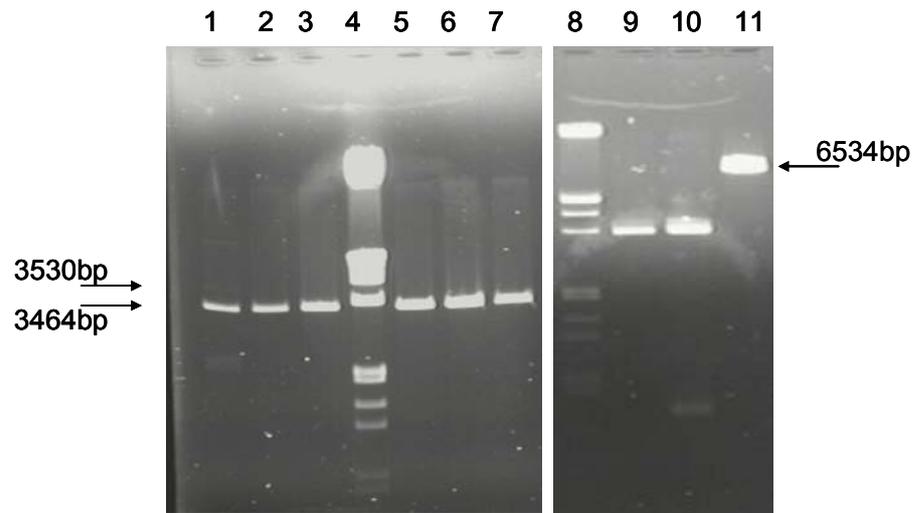


Figure 3.39 Plasmid digestion with the *HindIII* for linearization. Lane1, plasmid 7; lane2, plasmid 8; lane3, plasmid 9; lane5, plasmid 10; lane6, plasmid11; lane7, plasmid 12; lane9,plasmid 1; lane10, plasmid 4; lane11, plasmid 6; lane 4 and 8, *EcoRI/HindIII* cut molecular size marker.

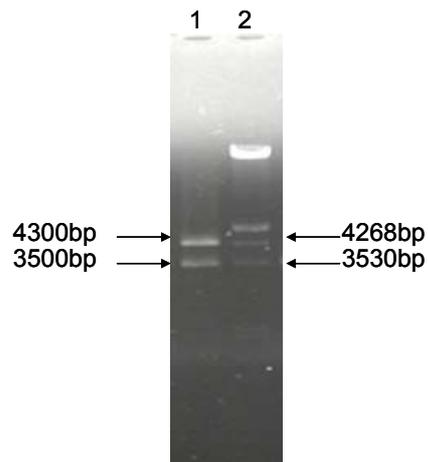


Figure 3.40: Plasmid digestion with *SalI/HindIII*. Lane1, plasmid 6 and lane2, *EcoRI/HindIII* cut molecular size marker.

3.2 Computer Based Analysis of Thermopsin Gene

Thermopsin gene details from TIGR database are as follows: *Thermoplasma volcanium* is an archaea evolved in euryarchaeota family. *Thermoplasma volcanium* thermopsin gene's primary locus name is TVN0237 and TIGR locus name is NT01TV0271. The gene length is 2997bp (coordinates 248059 to 251055) and protein length is 998 amino acid. Estimated molecular weight of the putative protein is 109.8 kDa and estimated pI value is 5.1682. Thermopsin gene is predicted GC content is 40.55%.

Thermopsin structural gene restriction digestion map was derived by using NEB cutter version2 program (figure 3.41). The cutter and non-cutter restriction enzymes and their cut positions are shown in the table 3.5.

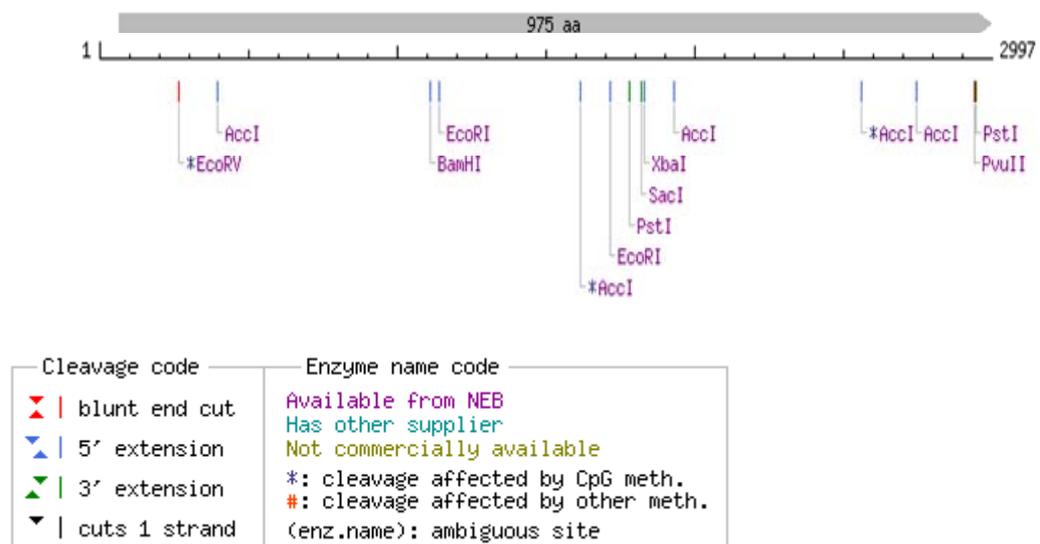


Figure 3.41: Restriction map of thermopsin structural gene.

Table 3.5: Cutter and non-cutter enzymes for structural thermopsin gene.

Noncutter Restriction Enzymes	Cutter Restriction Enzymes	Cut positions
<i>Ava</i> I	<i>EcoRV</i>	267
<i>Bcl</i> II	<i>Pvu</i> II	2941
<i>Hind</i> III	<i>Bam</i> HI	1114
<i>Kpn</i> I	<i>Xba</i> I	1830
<i>Mlu</i> I	<i>Sac</i> I	1821
<i>Pvu</i> I	<i>Eco</i> RI	1144, 1716
<i>Sal</i> I	<i>Pst</i> I	1784, 2946
<i>Sma</i> I	<i>Acc</i> I	393, 1613, 1928, 2560, 2741
<i>Sph</i> I		

Thermopsin gene alignments were done by KEGG Alignment Program which uses alignment score called Smith-Waterman (SW) score (Figure 3.42). The SW algorithm relies on the property of the scoring system in which the cumulative score for a path will decrease in region of poorly matching sequences. The basic principle of calculating SW score based on ignorance of the non-homolog sequence with query. So algorithm does not need to compute alignments with non-homologous proteins (Itoh *et al.*, 2005). When the score drops to zero extension of the path is terminated and a new one can begin. There can be many individual paths bounded by regions of poorly matching sequence the one with the highest score is reported as the optimal local alignment (Baxevasis and Ouelleette, 2001). *Thermoplasma volcanium* hypothetical membrane bound protein (TVG0154235), *Thermoplasma acidophilum* thermopsin precursor (Ta1403), thermopsin related protein (Ta0167) and thermopsin precursor (Ta0741) local alignment scores have high similarity values with *Thermoplasma volcanium* thermopsin gene (TVG0251055).

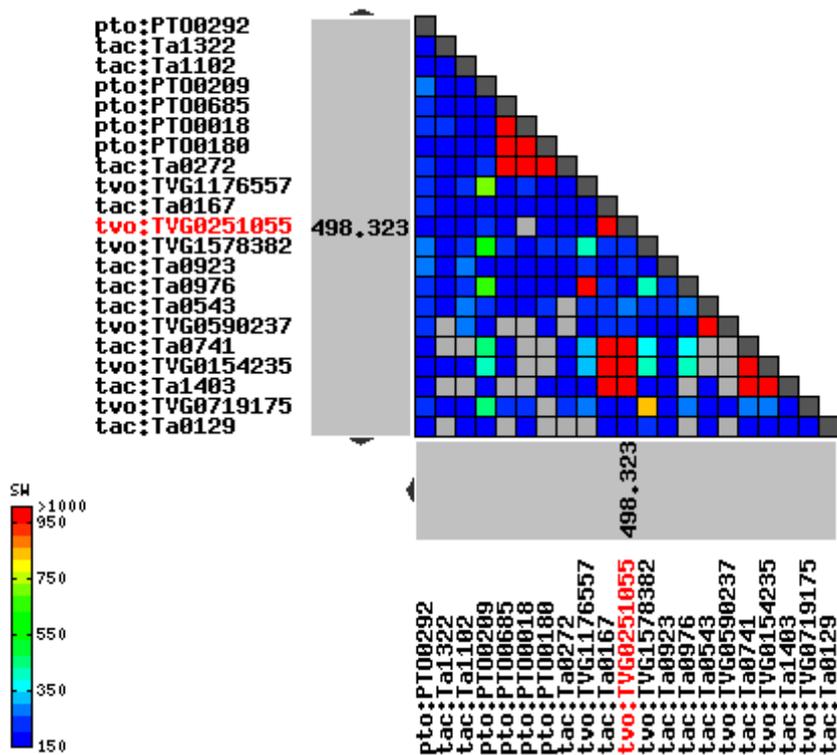


Figure 3.42: KEGG alignments data according to SW score. Red boxes represent high local alignment scores.

Gene neighboring data was collected for different *Archaea* strains, for the thermopsin gene, to know the genes in the neighborhood and to understand the gene location-strain relationship (Figure 3.43).

Thermoplasma acidophilum and *Thermoplasma volcanium* putative thermopsin gene localizations were highly similar which are placed between NADH oxidase and citrate synthase genes.

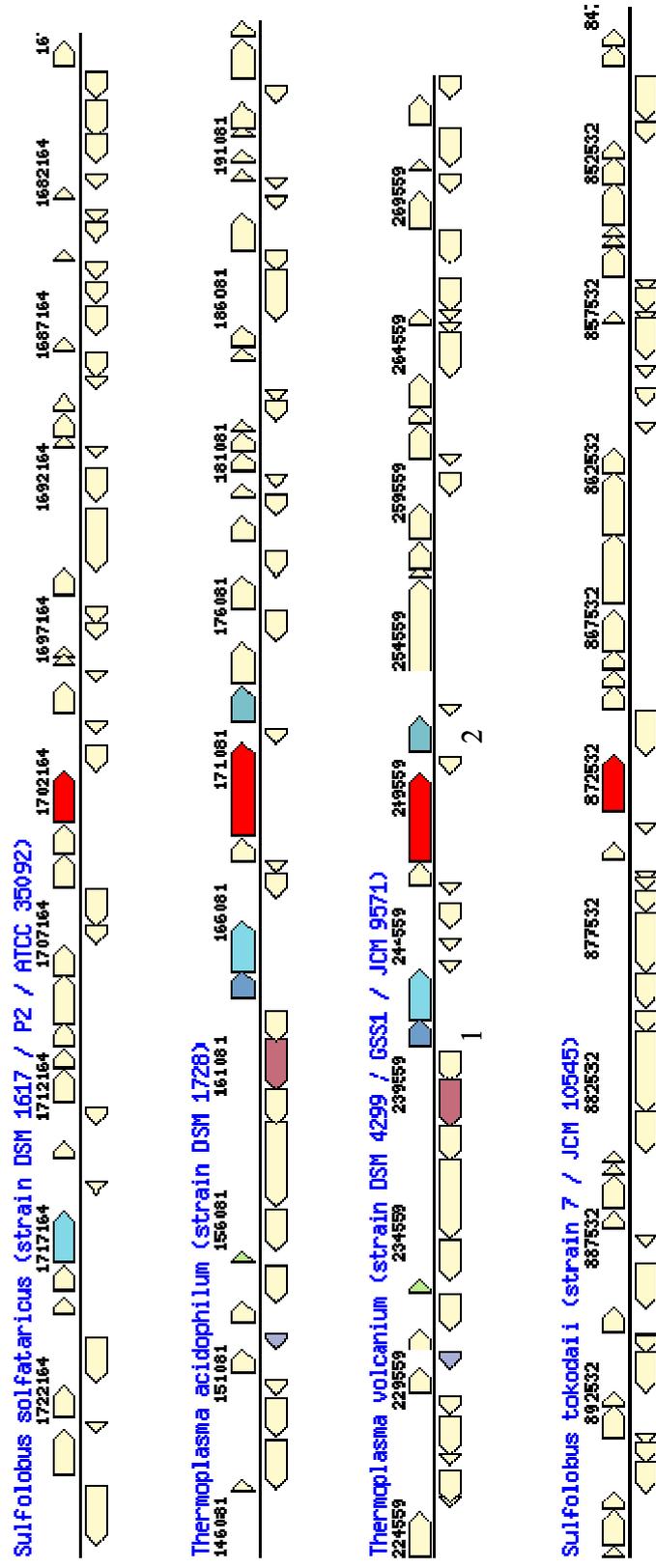


Figure 3.43: Gene neighboring data from *Ip. volcanium* genome analysis. Red arrows represent Thermopsisin gene. Genes marked as 1 are NADH oxidase and 2 are citrate synthase.

PSIPRED is a secondary structure prediction method, incorporating two feed-forward neural networks which perform an analysis on output obtained from PSI-BLAST (Position Specific Iterated - BLAST) (Altschul *et al.*, 1997). Figure 3.44 presents output from Version 2.0 of PSIPRED (from TIGR database). PSIPRED Versions 2.0 includes a new algorithm which averages the output from up to 4 separate neural networks in the prediction process to further increase prediction accuracy.

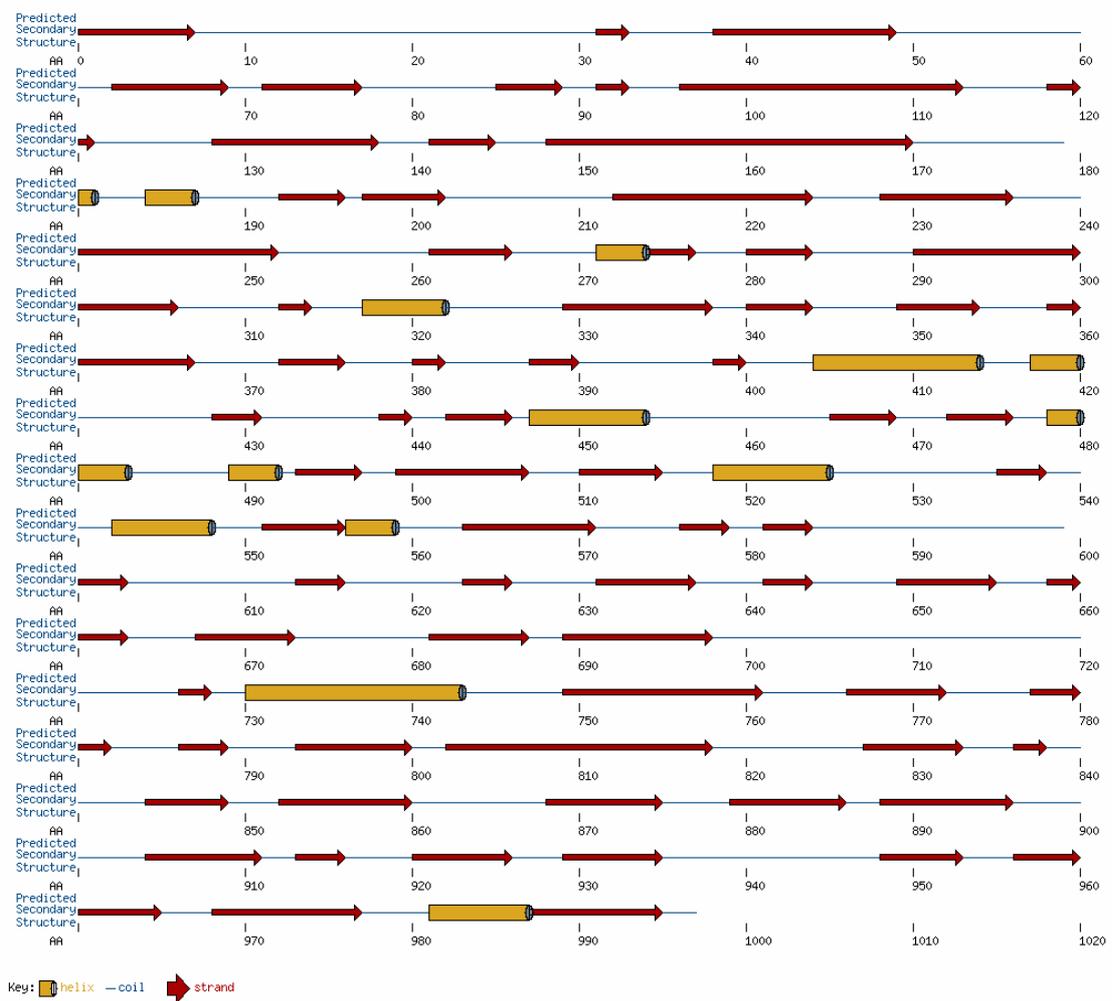


Figure 3.44: Predicted secondary structure of Thermopsin gene. Yellow color represents helix structure, blue color represents coil structure and red color represents strands.

Multiple sequence alignments of different *Archaea* strains for thermopsin gene were done by ClustalW Version 1.8 program. Alignments results are shown in Appendix E. Table 3.6 shows the alignments scores and Figure 3.45 shows phylogenetic tree of these strains.

Table 3.6: Scores of multiple sequence alignment thermopsin from different *Archaea*

	TA1	TA2	FA	PT	SA1	SA2	ST	SS	TA3	TV
TA1										
TA2	40									
FA	51	36								
PT	51	38	60							
SA1	16	15	20	16						
SA2	16	14	15	12	37					
ST	11	11	12	13	31	27				
SS	15	17	16	16	42	48	32			
TA3	45	28	45	44	18	21	11	17		
TV	38	44	35	36	20	17	12	19	30	

TA1: *Thermoplasma acidophilum* 1

TA2: *Thermoplasma acidophilum* 2

FA: *Ferroplasma acidermanus*

PT: *Picrophilus torridus*

SA1: *Sulfolobus acidocaldarius* 1

SA2: *Sulfolobus acidocaldarius* 2

ST: *Sulfolobus tokadaii*

SS: *Sulfolobus solfataricus*

TA3: *Thermoplasma acidophilum* 3

TV: *Thermoplasma volcanium*

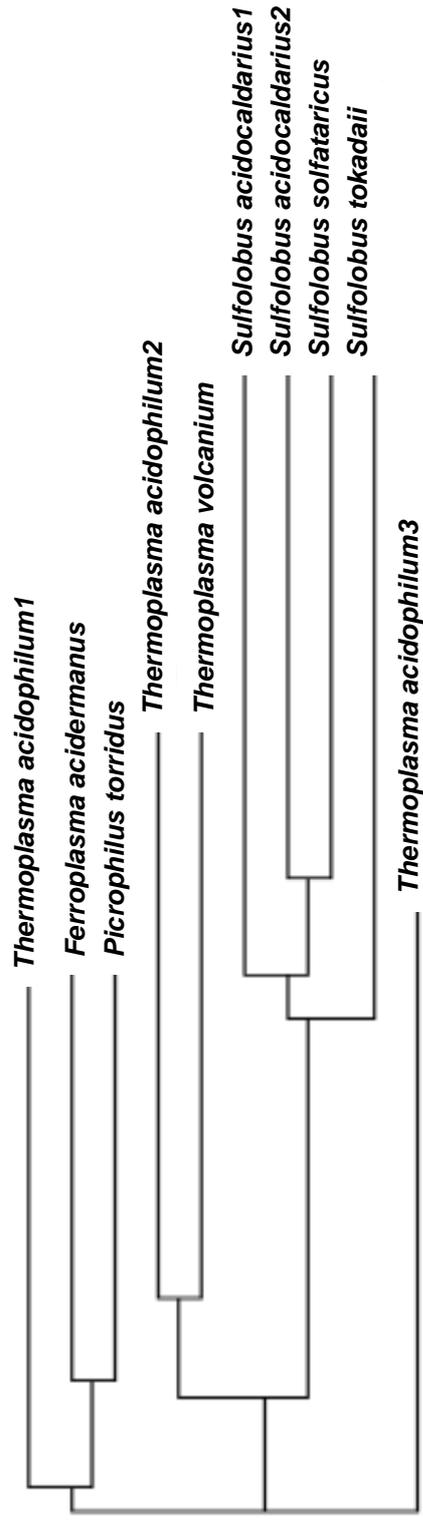


Figure 3.45: Phylogenetic tree of different *Archaea* strains.

CHAPTER IV

DISCUSSION

Proteases were initially characterized as nonspecific hydrolytic enzymes that are responsible for the degradation reaction of proteins. However, it is becoming increasingly recognized that proteolysis represents separate mechanism for achieving precise cellular control of biological processes in all living organisms, through highly specific hydrolysis of peptide bonds (Lopez-Otin and Overall, 2002). Besides, important roles of proteases they are also very valuable for industry. Rao *et al.*, reported that proteases represent one of the three largest groups of industrial enzymes and account for about 60% of the total worldwide sale of enzymes (Rao *et al.*, 1998). Owing to the expanded roles for proteolytic enzymes, there has been increasing interest in the identification and functional characterization of the many proteases that are present in various organisms, from prokaryotes to eukaryotes.

Acid proteases (aspartic proteases) are one of the important class of the proteases because of their critical biological functions in bioprocesses and industry. They have important roles in digestion of foods, activation of protein based biomolecules, invasion process of pathogen microorganisms, germination and senescence processes in plants, virus multiplication during invasion of the host organisms (Rao *et al.*,1998). They are also used in industrial applications especially cheese manufacturing (Preetha and Boopathy, 1997; Fernandez-Lahore *et al.*, 1998). Acid proteases are potential target for the pharmaceutical industry and also used in different bating steps of leather processing (Rao *et al.*, 1998). The value of industrial enzymes is determined according to the stabilization of enzyme during the harsh industrial process. *Archae* usually live under extreme conditions

and their enzymes are resistant to extreme conditions such as high salinity and temperature. Thus enzymes which are obtained from archaea can be used readily under various industrial conditions and they are also important for evolutionary studies.

Previous study in our laboratory several thermophilic microorganisms (bacteria and archaea) were isolated from acidic hot-springs in various hydrothermal sites in Turkey, and significant proportion of which exhibited intra and/or extracellular protease activities at various levels (Erdem, 1999, Kocabiyık and Erdem, 2002). Also, a variety of proteolytic activities from a thermophilic archaeobacterium *Thermoplasma volcanium* were studied (Kocabiyık and Özdemir, 2006; Kocabiyık and Özel, 2006). A pepstatin insensitive thermostable acid protease activity was detected in the culture supernatant of *Tp. volcanium* (Kocabiyık and Özel, in press, 2006). Considering the advantages of thermal stability of archaeal enzymes, especially extracellular one, for industrial applications, in this study we have selected *Tp. volcanium* as the source organism, for production of a recombinant acid protease. In enzyme production, extracellular enzymes are preferred because elaborate and costly methods for cell disruption are not necessary. In addition, extracellular enzymes are easier to purify from the culture medium compared to cytosolic enzymes. In our previous studies, the growth conditions of the source microorganism, *Thermoplasma volcanium*, were optimized and optimum growth temperature and pH were determined as 55⁰C and 2.7 (Kocabiyık and Özel, 2006). Also, the putative thermopsin gene was cloned in *E.coli* (Özel,2003). This study is aimed at expression of the cloned *Thermoplasma volcanium* gene in *E.coli* using a His tag fusion strategy.

There are different expression systems for recombinant proteins, two widely used ones being baculovirus-insect cells and *E.coli* expression systems, *E.coli* expression systems are commonly used because *E.coli* is a rapidly growing microorganism and it is compatible with many reagents and vectors. However, fusion protein strategy has been preferred to these systems for the last two decades. Fusion proteins have some advantages such as availability of several

detection and purification materials. Fusion partners also facilitate the purification of the expressed recombinant protein by affinity chromatography. For example, Murine leukemia virus protease and human foamy virus (HFV) proteases were expressed in fusion with maltose binding protein in *E.coli* (Feher *et al.*, 2004; Boros *et al.*, 2005), *Onchocerca volvulus* (filarial parasite) OV7A and OV4A acid proteases were expressed in fusion with outer membrane protein A (ompA) in *Salmonella typhimurium* and isolated by affinity chromatography (Jolodar and Miller, 2002).

Glutathione-S transferase (GST) and Histidine tags are the other examples of fusion partners. GST tag protein binds to glutathione containing matrix and can be easily purified. Then GST tag is removed by thrombin treatment (Kapust and Waugh, 1999).

Recombinant fusion proteins containing polyhistidine tags on either the N- or C-terminus are widely produced (Hochuli *et al.*, 1987; Hochuli *et al.*, 1988). Histidine exhibits highly selective coordination with certain transition metals and have great utility in immobilized-metal affinity chromatography (IMAC). Under conditions of physiological pH, histidine binds sharing electron density of the imidazole nitrogen with the electron-deficient orbitals of transition metals. Although three histidines may bind transition metals under certain conditions, six histidines reliably bind transition metals in the presence of strong denaturants such as guanidinium (Hochuli *et al.*, 1987). Such protein tags are commonly referred to as '6XHis'. This is known as Ni-NTA technology. Nitriloacetic acid (NTA) is a tetradentate chelating adsorbent. NTA occupies four of the six ligand binding sites in the coordination sphere of the nickel ion, leaving two sites free to interact with the 6XHis tag. For instance, a soluble, unglycosylated, 6XHis tagged human pro-BACE-1 (β -secretase) was expressed in insect cells using baculovirus infection (Bruinzeel *et al.*, 2002). A5 family of the aspartic proteases contains only one member which is called thermopsin. The first thermopsin gene was isolated from *Sulfolobus acidocaldarius* (Lin and Tang, 1990). Thermopsin is thought as a new class of aspartic proteases because the conserved active site

sequence (Asp32 and Asp215) is missing in thermopsin and DAN and EPNP inhibitors do not totally abolish its activity (Fusek *et al.*, 1990). While the *Sulfolobus acidocaldarius* thermopsin was 35kDa protein, *Thermoplasma volcanium* thermopsin protein estimated molecular weight is 109.8kDa. Lin *et al.*, reported that thermopsin enzyme may be rapidly degraded by the bacterial cells because it was not obtained high amount of thermopsin gene in expression vector pET. Thus 35kDa thermopsin protein expressed as a heterologous fusion protein as pepsinogen-thermopsin (PP-TH), procathepsin-D (PCaD), rhizopus pepsinogen (RP-TH) to prevent the degradation, under the control of T7 promotor in *E.coli* strains. Moreover in baculovirus/insect cell expression systems preliminary results were promising in that thermopsin can be synthesized as an active enzyme (Lin *et al.*, 1992).

In our laboratory, putative acid protease gene (thermopsin) sequence of *Thermoplasma volcanium* has been amplified by PCR and 3080bp amplicon has been ligated into the pDrive vector and then transferred into the *E.coli* TG1 competent cells. Expression could not be seen because the gene had been placed reverse orientation relative LacZ promoter and gene did not have its own transcription control elements including the promoter. Subcloning into an expression vector can be the solution for expression and QIAGENexpressionist Kit was chosen for expression.

In this study we followed two different strategies for expression. We modified the PCR amplification primer set for *Thermoplasma volcanium* thermopsin gene based on the instructions of the Kit Manual. TP1'/TP2 primer set was, so designed that, the start codon was missing in the amplified thermopsin gene. Although the start codon of thermopsin is not methionine we still removed the first codon to prevent truncated protein expression. Thermopsin gene was amplified with TP1'/TP2 primer set as a 3070bp PCR product. We ligated it to pDrive vector and cloned in *E.coli* TG1 cells. Then we characterized the positive clones and among which pDriveBSTP'-5 clone possessed recombinant pDrive vectors with PCR product placed the same direction with the LacZ promoter region. The cloned

3070bp PCR products from pDriveBSTP'-5 was sub-cloned to pQE vector using different *E.coli* host strains. We planned to express thermopsin gene from *Thermoplasma volcanium* as fusion protein with 6XHis tag. pQE vectors have different open reading frames and one of them should be the correct one for thermopsin gene expression. Firstly, we sub-cloned 3070bp PCR product to pQE30-31 and 32 vectors then transferred them into *E.coli* M15[pREP4]. Secondly true recombinants were detected with colony blot hybridization and immunological detection method. We used Anti-His HRP conjugate which is specific for 6XHis tag. Conjugates contain Horse raddish peroxidase enzyme and when treated with chromogenic DAB substrate enzymatic color reaction makes possible to select expression colonies easily. This method is a quiet specific method. We obtained strong signals with the membrane blots of pQE31 expression vector but we did not obtain expected restriction digestion profile from any of recombinant plasmids. pREP4 plasmid is available in *E.coli* M15[pREP4] strains which enables the regulation of expression by producing high level of Lac repressor protein is 3740bp. It may cause complication in restriction digestion and profiling. Furthermore M15 strain has RecA⁺ phenotype so there can be the recombination between host DNA and PCR product and may be the reason for unexpected profile of recombinant plasmids. In order to overcome these problems, we transformed *E.coli* TG1 strains and obtained detectible signals on the membrane blots of the recombinant colonies carrying recombinant pQE31 vectors; three of these plasmids pQE31-1, pQE31-5 and pQE31-7 carried the 3070bp cloned thermopsin gene fragment. We are planning to extend our characterization studies to sequence analysis by Southern Blot Hybridization method and sequencing of the cloned genes in the recombinant plasmids.

In the mRNAs the first codon usually encodes methionine but thermopsin gene first codon encodes for Leucine amino acid. As a second strategy for expression, the previously amplified thermopsin gene (3080bp) was sub-cloned to pQE30 and 31 vectors which were transferred into TG1 competent cells. We obtained some signals from the colonies which carried recombinant pQE31 plasmids. However, non of the putative recombinant plasmids yielded positive results after restriction

digestion. The failure might be either, due to incorrect reading frame of the thermopsin gene in pQE30 and pQE31 vectors, or simply interference of the start codon with 6XHis tag. pQE 32 vector has not been tried yet so may be it would provide the true open reading frame for 3080bp PCR product. Therefore pQE32 vector might be used to test this possibility.

Another step for characterization of the expressed protein will be purification of the protein. Ni-NTA matrix will be used for this purpose since it is specific for 6XHis tagged proteins. Histidine residues directly bind to the matrix which makes protein purification possible by affinity chromatography.

The alignment scores according to KEGG database indicated that *Thermoplasma volcanium* thermopsin gene was the most similar one to *Thermoplasma acidophilum* thermopsin gene. Multiple sequence alignment was done by ClustalW 1.8 Version. The results of the alignment of the amino acid sequences of putative and characterized acid proteases from various *Archaea* revealed that *Tp. volcanium* acid protease was the most similar to that of *Tp. acidophilum1*, showing homology of 44 %, respectively, and it showed lowest similarity with the acid protease from *S. tokodaii*.

Gene neighborhoods data has shown that thermopsin gene localization of *Thermoplasma acidophilum* and *Thermoplasma volcanium* are very similar, and both thermopsin genes were between NADH oxidase and citrate synthase genes.

CHAPTER V

CONCLUSION

1. PCR amplified 3080 bp fragment containing putative thermopsin gene, in *Tp. volcanium* genome was sub-cloned into pUC18 vector from pDrive vector in *E. coli*. Cloning was confirmed by restriction analysis of isolated plasmids from recombinant colonies.
2. PCR amplified 3070bp fragment containing putative thermopsin gene, in *Thermoplasma volcanium* genome was cloned to pDrive vector in *E.coli*. Cloning was confirmed by restriction analysis of isolated plasmids from recombinant colonies.
3. 3070bp fragment was sub-cloned to pQE31 vector in *E.coli* for expression. Sub cloning was confirmed by restriction analysis of isolated plasmids from putative recombinant colonies. Expression was confirmed by colony blot hybridization with pQE31-1, pQE31-4 and pQE31-6 plasmids.
4. 3080bp fragment was tried to sub-cloned into pQE30 and pQE31 vectors for expression and significant expression signals were obtained from colony blot hybridization. But restriction analysis of plasmids from recombinant colonies has show that expected size of the bands after restriction digestion did not harmonious with the experiment data. In further studies we are going to try pQE32 vector for expression.
5. Functional analysis regarding the postive expression clone pQE31-1, pQE31-4 and pQE31-6 is still in progress in our laboratory.

REFERENCES

Al-Janabi J., Hartsuck J. A., and Tang J., Kinetics and Mechanism of Pepsinogen Activation. *J. Biol. Chem* 247, 4628-4632 (1972)

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J. *Nucleic Acids Res.* 25, 3389–3402 (1997)

Arima, K., Yu, J. and Iwasaki, S., Milk clotting enzyme from *Mucor pusillus* var *lindt*. *Methods in Enzymology* 19, 446–459 (1970)

Babé L. M. And Craik C. S., Viral proteases: Evolution of diverse structural motifs to optimize function. *Cell* 91, 427-430 (1997)

Backman D. and Danielson U. H., Kinetic and mechanistic analysis of the association and dissociation of inhibitors interacting with secreted aspartic acid proteases 1 and 2 from *Candida albicans*. *Biochimica et Biophysica Acta* 1646, 184-195 (2003)

Barbiero L., Benussi L., Ghidoni R., Alberici A., Russo C., Schettini G., Pagano S. F., Parati E. A., Mazzoli F., Signorini S., Feudatari E. and Binetti G., BACE-2 is overexpressed in Down's syndrome. *Experimental Neurology* 182, 335-345 (2003)

Bardy S. L. and Jarell K. F., Cleavage of preflagellins by an aspartic acid signal peptidase is essential for flagellation in the archaeon *Methanococcus voltae*. *Molecular Microbiology* 50(4), 1339-1347 (2003)

Barkholt V., amino acid sequence of endothiapepsin. *Eur. J. Biochem.* 167, 327-338 (1987)

Baxevanis A. D. and Ouellete B. F. F., Bioinformatics; A practical guide to the analysis of genes and proteins, 2nd edition, A John Wiley & Sons, Inc., Publication, New York, USA (2001)

Beers E. P., Jones A. M. and Dickerman A. W., The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. *Phytochemistry* 65, 43-58 (2004)

Berdowska I., Cysteine proteases as disease markers. *Clinica Chimica Acta* 342, 41-69 (2004)

Berka R. M., Ward M., Wilson L. J., Hayenga K. J., Kodama K. H., Carlomagno L. P. and Thompson S. A., Molecular cloning and deletion of the gene encoding aspergillopepsin A from *Aspergillus awamori*. *Gene* 86, 153-162 (1990)

Blow D. M., The structure of cymotripsin. The Enzymes Volume III Hydrolysis: Peptide Bonds 3rd edition Edited by Boyer D. P., Academic Press New York USA 185-205 (1971)

Boross P., Tözser J. and Bagossi P., Improved purification protocol for wild-type and mutant human foamy virus proteases. *Protein Expression and Purification* in press (2005)

Bruinzeel W., Yon J., Giovannelli S. and Masure S., Recombinant insect cell expression and purification of human β -Secretase (BACE-1) for X-ray crystallography. *Protein expression and Purification* 26, 139-148 (2002)

Bujard H., Gentz R., Lanzer M., Stüber D., Müller M., Ibrahim I., Häuptle M.T., and Dobberstein B. A T5 promoter based transcription-translation system for the analysis of proteins in vivo and in vitro. *Methods Enzymol.* 155, 416-433. (1987)

Calderone R. A. and Fonzi W. A., Virulence factors of *Candida albicans*. *Trends in Microbiology* 9, 327-335 (2001)

Campos L. A. and Sancho J., The active site of pepsin is formed in the intermediate conformation dominant at mildly acidic pH. *FEBS Letters* 538, 89-95 (2003)

Cascella M., Micheletti C., Rothlisberger U. and Carioni P., Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *J. Am. Chem. Soc.* 127, 3734-3742 (2005)

Chitpinyol S. and Crabbe M. J. C., Chymosin and aspartic proteinases. *Food Chemistry* 61, 395-418 (1997)

Chou K.-C., Modeling tertiary structure of human cathepsin-E. *Biochemical and Biophysical Research Communications* 331, 56-60 (2005)

Citron M., β -Secretase inhibition for the treatment of Alzheimer's disease-promise and challenge. *Trends in Pharmaceutical Sciences* 25, 92-97 (2004)

Coopman-Lioudet E., Beaujouin M., Derocq D., Gorcia M., Lassis-Glandu M., Mathau-Louret V., Prébors C., Racherfort H. And Françoise V., Cathepsin D: newly discovered functions of a long-standing aspartic protease in cancer and apoptosis *Cancer Letters* in press (2005)

Dahlgren A., Kvarnstro I., Vrang L., Hamelink E., Hallberg A., Rosenquist A., and Samuelsson B., New Inhibitors of the Malaria Aspartyl Proteases Plasmeprin I and II. *Bioorganic & Medicinal Chemistry* 11, 3423–3437 (2003)

Dame J. B., Yowell C. A., Omara-Opyene L., Carlton J. M., Cooper R. A. and Li T. *Mol. Biochem. Parasitol.* 130, 1±12 (2003).

Dash C., Ahmad A., Nath D. and Rao M., Novel bifunctional inhibitor of xylanase and aspartic protease: implications for inhibition of fungal growth. *Antimicrobial agents and chemotherapy* 45, 2008-2017 (2001)

Delgado-Jarana J., Rincón A. M. and Benitez T., Aspartyl protease from *Trichoderma harzianum* CECT 2413: cloning and characterization. *Microbiology* 148, 1305-1315 (2002)

Demoz M., Castiono R., Follo C., Hasilik A., Sloane B. F. and Isidoro C., High yield synthesis and characterizaion of phosphorylated recombinant human procathepsin D expressed in mammalian cells. *Protein Expression and Purification* 45, 157-167 (2006)

Dunn B. M., Goodenow M. M., Gustchina A. and Wlodawer A., Retroviral proteases *Genome Biology* 3(4), 3006.1-3006.7 (2002)

Erdem B.,1999. Protease activities of thermophilic microorganisms isolated from hot springs in Turkey: detection of alkaline serine protease gene sequences by PCR. M.Sc. Thesis, Department of Biological Science, Middle East Technical University, Ankara

Feher A., Bross P., Sperka T., Oroszlan S. and Tözser J., Expression of the murine leukemia virus protease in fusion with maltose-binding protein in *Escherichia coli*. *Protein Expression and Purification* 35, 62-68 (2004)

Fernandez-Lahore H. M., Auday R. M., Fraile E. R., Bonino M. B., Pirpignani L., Machalinski C. And Cascone O., Purification and characterization of an acid proteinase from mesophilic *Mucor sp.* solid state cultures. *J. Peptide Res.* 53, 599-605 (1999)

Gonzalez-Lopez C. I., Szabo R., Blanchin-Roland S. and Gaillardin C., Genetic control of extracellular protease synthesis in the yeast *Yarrowia lipolytica*. *Genetics* 160, 417-427 (2002)

Farabaugh, P.J., Sequence of the Iac I gene. *Nature* 274, 765 (1978)

Fruton S., Pepsin. *The Enzymes Volume III Hydrolyss: Petide Bonds 3rd edition* Edited by Boyer D. P., Academic Press New York USA 120-152 (1971)

Fusek M., Lin X. and Tang J., Enzymic Properties of Thermopsin. *The Journal of Biological Chemistry* 265, 1496-1501 (1990)

Fusek M, Vetvicka V., Mitogenic function of human procathepsin D-role of the activation peptide. *Biochem J.* 303, 775-780 (1994)

Haki G.D. and Rakshit S.K., Developments in industrially important enzymes: a review. *Bioresource Technology* 89, 17-34 (2003)

Hashimoto H., Iwaasa T. and Yokotsuka T., Thermostable acid protease produced by *Penicillium duponti* K1014, a true thermophilic fungus newly isolated from compost. *Applied Microbiology*, 986-992 (1972)

Hess G. P., Chymosin- Chemical Properties and Catalysis. *The Enzymes Volume III Hydrolyss: Petide Bonds 3rd edition* Edited by Boyer D. P., Academic Press New York USA 213-244 (1971)

Hill J. and Phylip L. H., Bacterial aspartic proteinases. *FEBS Letters* 409, 357-360 (1997)

Hoorn R. and Jones J. DG., The plant proteolytic machinery and its role in defence. *Current Opinion in Plant Biology* 7, 400-407 (2004)

Hochuli E, Dobeli H and Schacher A., New metal chelate adsorbent selective for proteins and peptides containing neighboring histidine residues. *Journal of Chromatography* 411, 177-184 (1987)

Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R. and Stuber, D., Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate absorbent.. *Bio/Technology* 6, 1321-1325 (1988)

Horiuchi H., Yanai K., Okazaki T., Takagi M. and Yano K., Isolation and sequencing of a genomic clone encoding aspartic proteinase of *Rhizopus niveus*. *Journal of Bacteriology* 170, 272-278 (1988)

Inoue H., Hayashi T., Huang X-P., Lu J-F., Athauda S. B. P., Kong K-H., Yamagata H., Udaka S. and Takahashi K., Heterologous expression and site-directed mutagenesis studies on the activation mechanism and the roles of the basic residues in the prosegment of aspergillopepsinogen I. *Eur. J. Biochem.* 237, 719-725 (1996)

Jiang L., Liu Z., Liang Z. And Gao Y., An artificial proteinase system. *Bioorganic&Medicinal Chemistry* article in press (2005)

Jolodar A. and Miller D. J., Construction and expression of aspartic protease from *Onchocerca volvulus* as ompA fusion protein in a mutant strain of *Salmonella typhimurium*. *Genetics and Molecular Biology*, 25, 1, 97-101 (2002)

Jones B. L., Endoproteases of barley and malt. *Journal of Cereal Science* 42, 139-156 (2005)

Kapust, R. B. and Waugh, D. S., *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8, 1668-1674 (1999)

Kawashima T., Yamamoto Y., Aramoki H., Nunoshiba T., Determination of the complete genomic DNA sequence of *Thermoplasma volcanium*. *Proc. Jpn. Acad.* 75, 213-218 (1999)

Kocabiyik S. and Erdem, B., Intracellular alkaline proteases produced by thermoacidophiles: detection of protease heterogeneity by gelatin zymography and polymerase chain reaction (PCR). *Biosource Tech.* 84, 29-33 (2002)

Kocabiyik S. and Özel H., An intracellular-pepstatin insensitive acid protease produced by *Thermoplasma volcanium*. *Bioresource Technology* in press (2006)

Kocabiyik S. and Özdemir İ., Purification and characterization of intracellular chymotrypsin-like serine protease from *Thermoplasma volcanium*. in press (2006)

Kuhn D. M., Mukherjee P. K., Clark T. A., Pujol C., Chandra J., Hajjeh R. A., Warnock D. W., Soll D. R. and Ghannoum M. A., *Candida parapsilosis* characterization in an outbreak setting. *Emerging Infectious Disease* 10, 1074-1081 (2004)

Kumamoto C. and Vines M. D., Contributions of hyphae and hyphae-co-regulated genes to *Candida albicans* virulence. *Cellular Microbiology* 7(11), 1546-1554 (2005)

Kumar S., Sharma N. S., Saharan M. R. and Singh R., Extracellular acid protease from *Rhizopus oryzae*: purification and characterization. *Process Biochemistry* 40, 1701-1705 (2005)

Leal A. T., Pohl P. C., Ferreira C. A. S., Nascimento-Silva M. C. L., Sorgine M. H. F., Logullo C., Oliveira P. L., Farias S. E., Vaz Jr I. S. and Masuda A., Purification and antigenicity of two recombinant forms of *Boophilus microplus* yolk pro-cathepsin expressed in inclusion bodies. *Protein Expression and Purification* 45 107–114 (2006)

LaPointe C. F. and Taylor R. K., The type 4 prepilin peptidase comprise a novel family of aspartic proteases. *The Journal of Biological Chemistry* 275, 1502-1510 (2000)

Ledgerwood E. C., Brennan S. O., Cawley N. X., Loh y. P. and George P. M., Yeast aspartic protease 3 (Yap3) prefers substrates with basic residues in the P₂ P1 and P₂' positions. *FEBS Letters* 383, 67-71 (1996)

Liaudet-Coopman E., Beaujouin M., Derocq D., Garcia M., Glondu-Lassis M., Laurent-Matha V., Prébois C., Rocheford H. And Vignon F., Cathepsin D: newly discovered functions of a long-standing aspartic protease in cancer and apoptosis. *Cancer letters* Article in press 1-13 (2005)

Lin X. and Tang J., Purification, characterization, and gene cloning of Thermopsin, a thermostable acid protease from *Sulfolobus acidocaldarius*. *The Journal of Biological Chemistry* 265, 1490-1495 (1990)

Lin X., Liu M. and Tang J., Heterologous expression of thermopsin, a heat stable acid proteinase. *Enzyme Microb. Technol.* 14, 696-701 (1992)

López-Otín C. and Overall C. M., Protease degradomics: a new challenge for proteomics. *Molecular cell biology* 3, 509-519 (2002)

Maegawa H., Kimura T., Arie Y., Matsui Y., Kasai S., Hayashi Y. and Kiso Y., Identification of peptidomimetic HTLV-I protease inhibitors containing hydroxymethylcarbonyl (HMC) isostere as the transition-state mimic. *Bioorganic& Medicinal Chemistry Letters* 14, 5925-5929 (2004)

Madzak C., Gaillardin C. and Beckerich J., Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *Journal of Biotechnology* 109, 63-81 (2004)

Manteau S., Abouna S., Lambert B. and Legendre L., Differential regulation by ambient pH of putative virulence factor secretion by the phytopathogenic fungus *Botrytis cinera*. *FEMS Microbiology Ecology* 43, 359-366 (2003)

Monod M., Capoccia S., Léchenne B., Zaugg C., Holdom M. and Jousson O., Secreted proteases from pathogenic fungi. *Int. J. :Med. Microbiol* 292, 405-419 (2002)

Nehls U., Bock A., Einig W. and Hampp R., Excretion of two proteases by the ectomycorrhizal fungus *Amanita muscaria*. *Plant, Cell and Environment* 24, 741–747 (2001)

Oda K., Takahashi T., Tokuda Y., Shibano Y and Takahashi S., Cloning, nucleotide sequence, and expression of an isovaleryl pepstatin-insensitive Carboxyl proteinase gene from *Pseudomonas sp.* 101. *The Journal of Biological Chemistry* 269, 26518-26524 (1994)

Ohtaka H. and Freire E., Adaptive inhibitors of the HIV-1 protease. *Progress in Biophysics and Molecular Biology* 88, 193–208 (2005)

Olsen V., Cawley N. X., Brandt J., Egel-Mitani M. and Loh Y. P., Identification and characterization of *Saccharomyces cerevisiae* yapsin 3, a new member of the yapsin family of aspartic proteases encoded by the YPS3 gene. *Biochem. J.* 339, 407–411 (1999)

Ostermann N., Gerhartz B., Worpenberg S., Trappe J. and Eder J., Crystal structure of an activation intermediate of Cathepsin E. *J. Mol. Biol.* 342, 889–899 (2004)

Özel H., Cloning of an extracellular thermostable acid protease from *Thermoplasma volcanium* in *E. coli*. Master thesis METU Ankara Türkiye (2003)

Pekkarinen A., Mannonen L., Jones B. L., and Niku-Paavola M.-L., Production of proteases by *Fusarium* species grown on barley grains and in media containing cereal proteins. *Journal of Cereal Science* 31, 253–261 (2000)

Palma J. M., Sandalio L. M., Corpas F. J., Romero-Puertas M. C., McCharty I. And Rio L. A., Plant proteases, protein degradation, and oxidative stress: role of peroxisomes. *Plant Physiol. Biochem.* 40, 521-530 (2002)

Pessela B. C. C., Torres R., Fuentes M., Mateo C., Fernandez-Lafuente R. and Guisan J. M., Immobilization of rennet from *Mucor miehei* via its sugar chain. Its use in milk coagulation. *Biomacromolecules* 5, 2029-2033 (2004)

Pichová I., Pačličková L., Dostal J., Dolejší E., Hrušková-Heidingsfeldová O., Weber J., Rumi T. and Souček M., Secreted aspartic proteases of *Candida albicans*, *Candida tropicalis*, *Candida parapsilosis* and *Candida lusitanae* inhibition with peptidomimetic inhibitors. *Eur. J. Biochem.* 268, 2669-2677 (2001)

Polgar L. and Csoma C., Dissociation of ionizing groups in the binding cleft inversely controls the endo- and exopeptidase activities of cathepsin B. *The Journal of Biological Chemistry* 262, 14448-14453 (1987)

Poza M., Prieto-Alcedo M., Sieiro C. And Villa T. G., Cloning and expression of *clt* genes encoding milk-clotting proteases from *Myxococcus xanthus* 422. *Applied and Environmental Microbiology*, 70, 6337-6341 (2004)

Prasad B. V. L. S. and Suguna K., Role of water molecules in the structure and function of aspartic proteinases. *Acta. Cryst.*, 250-259 (2002)

Preetha S. and Boopathy R., Purification and characterization of milk clotting protease from *Rhizomucor miehei*. *World Journal of Microbiology & Biotechnology* 13, 573-578 (1997)

Rao B. M., Tanksale A. M., Ghatge M. S. and Dehpande V. V., Molecular and biotechnology aspects of microbial proteases. *Microbiology and Molecular Biology Reviews*, 597-635 (1998)

Reichard U., Monod M., and Ruchel R., Molecular cloning and sequencing of the gene encoding an extracellular aspartic proteinase from *Aspergillus fumigatus*. *FEMS Microbiol. Lett* 130, 69-74 (1995)

Richter C., Tanaka T. and Yada Y., Mechanism of activation of the gastric aspartic proteinase: pepsinogen, progastrin and prochymosin. *Biochem. J.* 335, 481-490 (1998)

Robb, F.T., and A.R. Place, *Archaea, A Laboratory Manual*, Eds. Cold Spring Harbor, Laboratory press, NY, USA, 1995

Sambrook J., Fritsch E.F., and Maniatis T., *Molecular cloning: a laboratory manual* 3rd edition Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2001

Sardinas, J. L., Microbial Rennets. *Advances in Appl. Microbiol.* 15, 39-81 (1972)

Schaaf A., Reski R. and Decker e. L., A novel aspartic proteinase is targeted to the secretory pathway and the vacuole in the moss *Physcomitrella patens*. *Eur. J. Cell Biol.* 83, 145-152 (2004)

Schaller A., A cut above the rest: the regulatory function of plant proteases. *Planta* 220, 183-197 (2004)

Shuker S. B., Mariani V. L., Herger B. E. and Dennison K. J., Understanding HTLV-I protease. *Chemistry&Biology* 10, 373-380 (2003)

Suarez M. B., Sanz L., Chamorro M. I., Rey M., Gonzalez F. J., Llobell A. and Monte E., Proteomic analysis of secreted proteins from *Trichoderma harzianum* identification of a fungal cell wall-induced aspartic protease. *Fungal Genetics and Biology* 42, 924-934 (2005)

Sutherland, K.J., Henneke C. M., Towner P, Hough D.W.,and Danson M.J., Citrate synthase from the thermophilic archaeobacterium *Thermoplasma acidophilum* Cloning and sequencing of the gene *European Journal of Biochemistry*, 194, 839-844 (1990)

Tanaka T. and Yada R., Redesign of catalytic center of an enzyme: aspartic to serine proteinase. *Biochemical and Biophysical communications* 323, 947-953 (2004)

Thanikaivelan P., Rao J. L., Nair B. U. and Ramasami T., Progress and recent trends in biotechnological methods for leather processing. *TRENDS in Biotechnology* 22, 181-187 (2004)

Wang Y. T., Yang C. Y., Chen Y., Lin Y. and Shaw J., Characterization of senescence-associated proteases in postharvest broccoli florets. *Plant Physiology and Biochemistry* 42, 663-670 (2004)

Weigh M, Reichard U. and GroB U., *Aspergillus fumigatus*- virulence and opportunism?. *Mycoses* 44, 351-355 (2001)

Wu L. C. and Hang Y. D., Acid protease production from *Neosartorya fischeri*. *Lebensm.-Wiss. U.-Technol.* 33, 44-47 (2000)

Veerapadian B., Cooper J. B., Sali A., Blundell T. L., Rosati R. L., Dominy B. W., Damon D. B. and Hoover D. J., Direct observation by X-ray analysis of the tetrahedral 'intermediate' of aspartic proteinases. *Protein Science* 1, 322-328 (1992)

Viterbo A., Harel M. and Chet I., Isolation of two aspartyl proteases from *Trichoderma asperellum* expressed during colonization of cucumber roots. *FEMS Microbiology* 238, 151-158 (2004)

APPENDIX A

BUFFERS AND SOLUTIONS

TrisHCl Buffer with Sucrose

- 50 mM Tris, 25 % Sucrose, pH 8.0

TE Buffer

- 10 mM Tris HCl
- 1 mM EDTA, pH 8.0

TAE Buffer

- 0.04 M Tris-acetate
- 0.001 M EDTA, pH 8.0

Buffer B

- 3 M NaCH₃COO, pH 5.5

Gel Loading Buffer

- 0.25 % Bromophenol blue
- 40 % Sucrose

LB Medium

- 10 gr tryptone
- 5 gr yeast extract
- 10 gr NaCl

SDS solution

10% (w/v) sodium dodecyl sulfate

Denaturing solution

0.5M NaOH, 1.5M NaCl

Neutralization solution

1.5M NaCl, 0.5M Tris-Cl, pH 7.4 (25⁰C)

20XSSC

500ml: 87.65g NaCl, 50.25g trisodium citrate.2H₂O

TBS buffer

20mM Tris-Cl, pH 7.5; 150mM NaCl

TBS-TweenTriton buffer

20mM Tris-Cl, pH 7.5; 500mM NaCl;

0.05% Tween 20; 0.2% Triton X-100

Blocking buffer

3% BSA in TBS buffer

DAB working solution

4.5ml in 500μl H₂O₂ solution

Anti-His HRP conjugates solution

125μl stock solution 1/1000 dilution

APPENDIX B

pDrive Cloning Vector

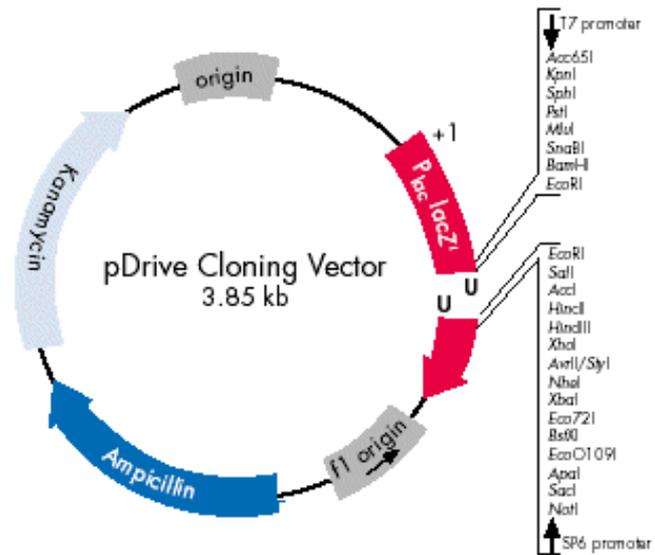


Figure B.1: pDrive cloning vector

pUC18 Cloning Vector

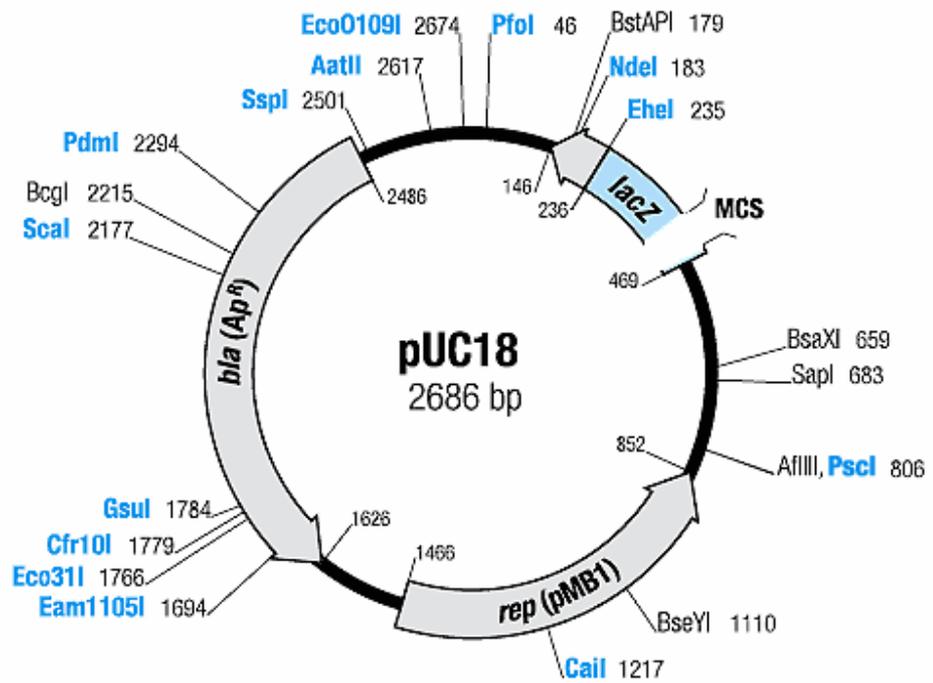


Figure B.2: pUC18 vector

pQE 30-31-32 Expression Vectors

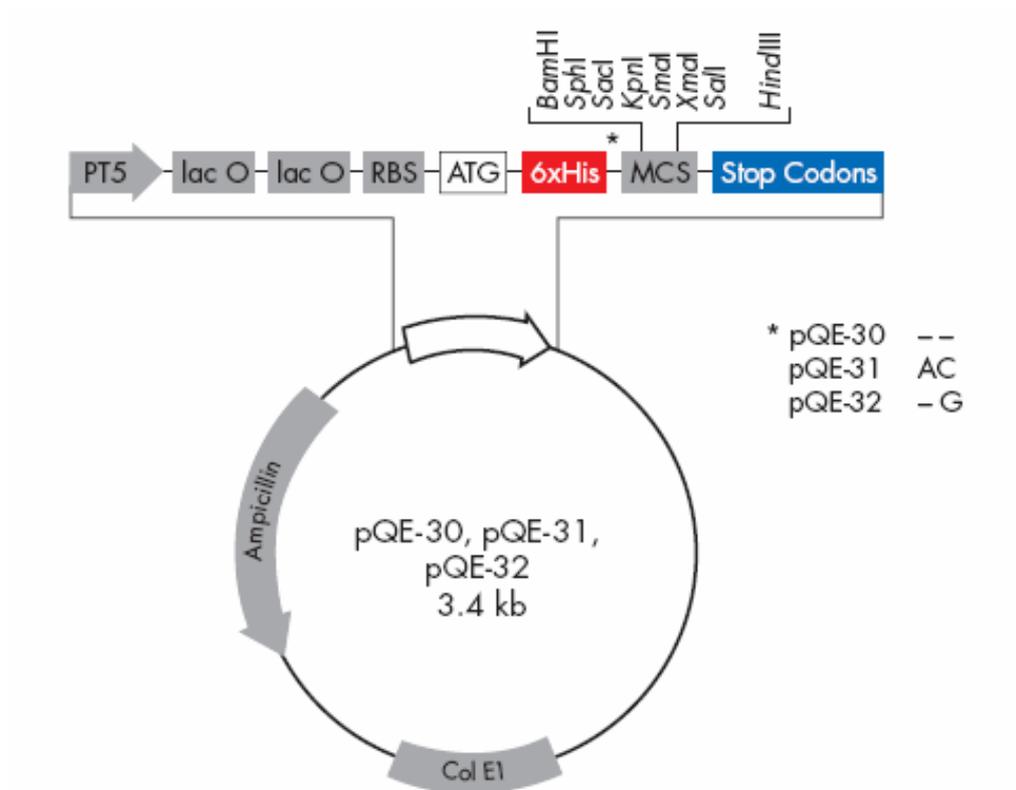
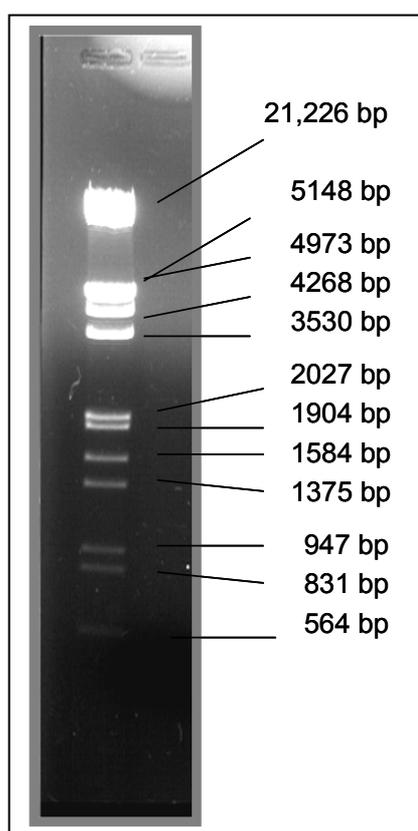


Figure B.3: pQE 30-31-32 vector set

APPENDIX C

MOLECULAR SIZE MARKER



EcoRI and *HindIII* cut Lambda DNA molecular size marker, MBI Fermentas AB, Vilnius, Lithuania

APPENDIX D

Nucleotide sequence of thermopsin of *Tp. volcanium*. START and STOP positions are indicated in RED.

```

248101      ggaatttt ggtaatatct ctgattaccg ttacgccgtt tcaaacacag
248161 gattcgcgatg attaccgccca ttcagctatg cctgcacgcg ctaacattgg atatgttaat
248221 agtactatac gaacaccggt gaaacaaat tatagcatac tgactgcaag caaatctatg
248281 cactactttt accctccaaa cagtggagatc gggaaaatgt accattacga tggaaatcatt
248341 acacctacca agattgcttc accttcgccca atgggtctat cagatatcgg tttatacaac
248401 aactctggcc acttgctctgc atacaatggt accfcaccaa gcgtaatggg agaaaactact
248461 ctgaactcat ctaaigagta ctaccttgac gatgatatgc cgaatagcti tagtatacag
248461 cttaacgctg tcctgaacaa tgtaacactt catggaacgc cgaaaactcgt attttggaca
248521 cagaacgttg tggtttattc tgcaggacg catgaattaa catttattga caacgttgg
248581 aaiftttcaa gcaggagcag ttttgcacti ccaftaagta ctttagaggg gaaaggaaaa
248641 atttgttcaa acaaatatta ttactatggt ggcccaactc tgaatataic ttttccgttt
248701 accctgaaac tataccttaa ctattcattt gcaaacggga acgacttagt atatttcaac
248761 tactctatcc tgaaccctag ttattcgaag tcaggatcat atgatatggt taccttcaat
248821 tctaattctt cagagagcgg ggccctactac tttgtcagcg gttacgtaac tttccatca
248881 gggctcctta atgatgcaga actggctcatt ggccggcccag gaggcggaaag taatgcaaac
248941 itctactcaa tcaatggaac tatgtcccta taciigtcca atgaaactgc tcatgcttat
249001 gaacctgtaa aatctgcgta tgattatggt tttgataccg gggaaacttc ctatggcgta
249061 aatggtggct ggtcaaatca tactgcaaag ctcaactggg gaccaagcct tctatatggc
249121 atgtggaatg tctcagcttc agagaggata attaaaatag aagcgccttc aggatccctc
249181 gtatttattt ctcccggatc tgaattcaat aatacttcgg ctgtatgggc gcctttaaat
249241 ggaattaaca caacgtttta tgcagcttca ggttcataat ccgctgctgt gttgaaaaat
249301 iactataagc agactttcta taatgatatt aatcccttaa ccaatgcaic attactgccg
249361 gacaataatg ctggagtata tacgcctata atgatcttta acaacagcca aatagcagca
249421 tatggctgctt ttgggcttgg gactgcaagc tctccatagc ccatataactt aaatggaagc
249481 gctagcatta gccactgttt tgggtcatta aacgattati tctttcctgt cttcgcaggc
249541 gtatctatcg taaatgttac aaagcatctc atcatttatt ataattcccc gatgtccatt
249601 aatataaccg gacatccaga atatcaactc aaagccgcta acctgcctct taacaactac
249661 ctaccctcc aggtatacgg ttcttcaaac ataictatac tgaacgctag cctaataatca
249721 ggctggttcc caictactat ggccggtttc atcgtagggg accttacagt ttggaattct
249781 accaatgttc ttgtagctc aaatacgttc gttccgtttt cgtatttccc tggtagctg
249841 cagtctgcat attttcttc acttcttatt tatggagctc cggatactct agagaatgat
249901 accatatggg gaaacacatt ccctggcccg atcatcgtg gagtaaacac gacaactggt
249961 ctctttatgg acgctcccgg aaacttggtc tacaacaact ttttaccgga acaagacatt
250021 gcgtattctc ccaacataaa catatacacc ggcaagaacg ccacctatca tgattatgg
250081 agcctttcga acaaggaacc cgtgaattat acgatttacg ttaatggaa accattaata
250141 ggcagcataa taggaggtgg attcgaaggt ggcaattact ggggcagccc ttatattacg
250201 ggcccatgga atgaacaagg atacattgaa tcaggatagc attactatcc gtigaattac
250261 ttcagctata tttttttct gcacgcaicc ggattgccaa aaggcctaag ttggcaggca
250321 tcaatcaacg gtgtttatta tgtgctgaac acgcccgtata tgaacgtatc tcttacaac
250381 gatggcaata gctcattctc agtttatttg ccgccaggat atatctacag cggcgaaaac
250441 tataattatg cttacaaaaa taatacatac tacaacttta cagtaaatc gggcgattg
250501 tttccgтта cgttcaatga agagggcttt ccaggtaatt caacttggga agtacaagtc
250561 gcaggaaaca tatacgcatc catttcgaga tcaattacag ttctgtctcc aaatggtacg
250621 tatacctatg aaatafacgg tccgtatggg tattccccac ataacagcta tggctatctt
250681 aatgtatcag gttctctgtc gatcatttct gtgaaaattta caactagtat gtacatttta
250741 acagtatctg aacacgggct tcaaaaaggt acagattggc tcatagtggg ggatggtaaa
250801 gtcfactacg gaaatacctc taccattaaa atctctctg caccaggcag gtaftcgatc
250861 tacgcagata caatacctgg ttattccaaa actctatacaa caaatactat tataataaat
250921 ggaaaaaact tggatgtgtc ggtcacatat atccagaaga gggacatcat tttctattac
250981 tatattatag caicaacagc tgcagggctc attgcaggag cggccatcgg tttttattt
251041 agaagaaaac atlgaagatc ttctttccta gtttttttag aaaaacttta ttcattattt
251101 aatacagctt ggggaaagat agattttatg gc

```

APPENDIX E

Multiple Sequence Alignment of *Archaeal* Thermopsin Genes

NP_393645	-----MKVLVIFLLILSLIGSTLSAFFVQMSGNTHQNESQIS-----ASSQAANI	46
NP_110756	-----MVISLITVTPFOTQDSDYRHSAMPASPNIGYVNSTIR-----TPVETN--	44
ZP_00609469	-----MKKLLAIVVAIFVALAFVAVSSQGTTPAAPAVSHQSTQONSQRSLLAI	51
AA42746	-----MKKAFVFIAMFLISAFAIINDHGGDKNQISAFVSEK-----DSGNRVVYL	47
NP_394857	MFHGFMMTDIKKSIIVLVTLIFITTLFFAVSAAPVHATTPAVQASSYVNAQQSQMYSSV	60
CAC11875	-----MFITILIAVIAVIGFLGINFAFNSASTPSIQKDTGLKASS-----KPVVETQTSV	50
AA480848	-----MSRIIVLAVVALALLNPLFLVHSLPTG-----	28
NP_343288	-----MLKHIVLVLLLLLTP-----LVAFSPTG-----	25
AA481460	-----MLKVRIVAVLSFLLLTLP-----VIGFAVTAPVG-----	29
NP_378476	-----MKSIIILLAIILFSLPWNLTGIGFIPR-----	28
*		
NP_393645	VSRIGSINVSSRYEFLPNFNNAVSEYSEDGHITLTYDSAPAPMGVADVGLFNSSGSIPTYS	106
NP_110756	-YSILTASKSMHYFYPNSEIGKMYHYDGIITPTKIASPSMGLSDIGLYNNSGHLISAVN	103
ZP_00609469	EKKLEEKGIPLKYASIPAFYNNITRDNG-VIAPSYQSAAPMGIGFYGTQNESKLVGVN	110
AA42746	EKIAEKRGVPLRYISPPAAMPDTNYADG-VIEPSYNNAPAFFGIGFYGTENINGRLTGFN	106
NP_394857	MQKLESRGISPKYAYLPDFNAQRIVHKGNTVSPYATTPAPMGIGDIGLKYLNQQLTLPYE	120
CAC11875	LKPVNN-----YLFPAKDRHFLYQNG-TVQPLYTQSPAPMGIGFFGLNISGLVGSN	103
AA480848	-----VSAHDGPIYTNVAVLGYANITS-----	49
NP_343288	-----VVAYNGPICTNEVLGYANISS-----	46
AA481460	-----ISSYSPVTVTTTILGYANISS-----	50
NP_378476	-----YNASSQYAMGVSSIGSSSLSISTN-----	52
*		
NP_393645	VNTTGMGAVDLNSTQVFLYDDQPVYSFAVQLNAVAHDVDFLGNSSYSFWAQNVVVYSTR	166
NP_110756	VTSSVMGEITLNSSNEYVLLDDMPNSFSIQLNAVLNNTLHGTPKYVFWTONVVVYSVR	163
ZP_00609469	LTTPSVMASIGIQNMSQFYLLNDGPTSETFQLNSVLTNVTLFGNSTYTFWTONVAFYSER	170
AA42746	LTTPGIMASIKIYNMSDFYLLNDGPESETFQLNAVLSNVTLFGKSGYSFWTONVAFYSAR	166
NP_394857	INFTSTEASVTIDNLSDFYLLNDGPHSVTIQLNSVLTNVTILGNSSYSFWTONVVVYSAR	180
CAC11875	YYASSFAATINITNLSVFNLADDAPSMTFQLNTVTANTLFGNSSYTFWTONVASYSVR	163
AA480848	-----LQAYNSSFN-----VPYGASLQLNVLEATSTNGNT-YYFWLQNVAFQVTN	94
NP_343288	-----LLAYNTSASQLGVPPYPYGASLQLNVMLEVN-TSGGE-YYFWLQNVADFITN	94
AA481460	-----LLAYNSTFF-----QVYGASLQLNAILEVD-TPSNT-YYFWVQNVAGFITN	94
NP_378476	-----EVLGYVEVFSLS-----SSPFSIQFNAVIKATNIYGEI-LDFWVQNIIEFNKQ	100
*		
NP_393645	THQLSFIDNVVNFSS---PAFNMTAN--ALHGNGKLVPGV-FYYDIGPTFNISFPFSLK	219
NP_110756	THELTFIDNVVNFSSR---SSFALPLS---TLGKGRKIVSNK-YYYYVGPTLNISFPFSLK	217
ZP_00609469	TQTIQFLTNIWNFSS---PAVTISSN-VFHSDDGILCAPT-FYYAIGPTIHVTTFPFLN	224
AA42746	THQLQFLVNIWNFSS---PAINFNTANSISYHSSGFVCAPT-FYYAVGPTINVTAFPTLN	221
NP_394857	THQITFIDNVVNFSS---PAFNMTN--ALHGNGILVPPV-FYYDIGPTFNVTVPFTVD	233
CAC11875	THEISFVNDNIWNFSS---PTAVMSSN-AILNSTGLLYPYSGVHIALGPTFLPFPFTL	218
AA480848	ESLMCFDNDNVNYYTTATAEIS-NVTGNGGIGFTYNIILFGHPTFYGYSTEPMPYRFLALY	153
NP_343288	ESKVVFGDNIWNSTPPFAGIN-NIVGKEI-YSTDFSSHSSYYAVGTYYIKYHFFFSFY	152
AA481460	NNTLFFNDNIWNATGMDSNIT-EVIGDGNISTCDSQAPQTFYGAASSQYIYVHFFLSFY	153
NP_378476	E--LNFTEIWNFTYSASINPNLISGKGEVYSTDINGHIVNYYVSYSTPYQYSVLEPLTYI	158
*		
NP_393645	LYLNSSLVGGNDAIYFNYSIE---NGKITKSGSYDAVTFNSSG-----KASSPARFLISG	271
NP_110756	LYLNYSFANGNDLVYFNYSIIL---NPSYSKSGSYDMVTFNSNS-----SESG-AYYVFSV	268
ZP_00609469	LYLNSTVVDNRSAVFFNYSIT---SDNHVQSGSYDAIFNSTYQGPSTFTAKAPEYLAGS	281
AA42746	LYMNSGIWNDSAVFFNYSVK---TASMNVSGTYDEVLFNSTR-----AVINPEYLAGS	272
NP_394857	LYLNSTVINGNSAVYFNYSVF---TSGKVYSGSYDVFIFNSTG--HPGYRAPQPEYQIDG	288
CAC11875	LYLNTSEINGNNAVYFNYSIP---TLG---VSGTYDRVTFNSTYGMPPAYKAPPSYFRVSG	273
AA480848	LLINESLSSNGVIVNFYVVLQNGSLVPPHPVYDVSFIPVDPQSAVYIIVNDSLTPSAV	213
NP_343288	LIINESYDTQGVYVSGYVILQNGNISPPNPIFYDVFIPIQNLSFASIIIANQTTPSAN	212
AA481460	MFINVTPTVRGPLVTFGYVILQNGKITSPKVQVYDVAIIPVQGAISAQIVVQNSYTYVYN	213
NP_378476	LFISISYSS-----PYIYVSIYSGNLSDPTSNYDTITINVENLQNAACIYVSPSYTG---	210
*		
NP_393645	YS-----PSPAG-LLYDAELAITGPGGGSNWNVYINGMTGLYLYNSTSK-----	315
NP_110756	YV-----TSPSG-LLNDAELVIGPGGGSNANFYISINGTMSLYLFNETAH-----	312
ZP_00609469	TQ-----ITPTGYIPYDFEIMVGGPGGGSTTSIYNVNATMNLRYMGSSG-----	325
AA42746	TQ-----LTPTGYIPYDFEIIIGGPGGGSTTSLYNINATMNLKYLKYS-----	316
NP_394857	YG-----VDATGYLINDAEVMIIGGPGGGSTTSIYSINATMHLVYLNSTTG-----	333
CAC11875	TQ-----LTPG-LLYDAEIMIGGPGGGSTTTVTDINGMMDLKYIPAANPMHPTPPPPP	327
AA480848	VLNFTLFYQEAALGNLMDTELWVGGYENGEFTTFTQMSYLGFLYLSG-NSFVP-----	265
NP_343288	FG-----IVTYLGNLYLDAELWVGGFGNGESTTFLNMSYLLALLYKKS-GEVVP-----	259
AA481460	QG-----LYSYVGLKQVDELVWGGLSNGEHTTFKEMSSLLAIYYLRN-GQWLS-----	260
NP_378476	-----VGLPYSVELVWGGSGNGQSAFTNMESSDLAIFYQVKEPELYSP-----	252
*		

```

NP_393645 -----AFNIVRSAYDAGVDTGETSYGVVDVGVYGNM----ADL 348
NP_110756 -----AYEPVKSAVDYGFDTGETSYGVNVGVSNHT----AKL 345
ZP_00609469 -----YVNVPSAYDVVGSETGETSEGVAVSWDNDT----AYL 357
AA42746 -----YKNVPSAYDVVGSETGETSDGIAISWHGAT----EHL 348
NP_394857 -----TYQTVPSAFDIAADTGETSEGVAVSWSHGPDGYAKL 370
CAC11875 PFFFFFFPGSQHASGKPPAAMTVQSYVNVPSAFDFGDTGETSVGTAIWNSMD----QVLL 384
AA480848 -----FSTVYTYGNDTAESTGDLYVSIASNG----DAYV 295
NP_343288 -----FSQVYNYGSDTAESTNNLQVLIGKNG----DAYV 289
AA481460 -----FNEIYNYGFDTAESANNLTVVVDKQC----FAHV 290
NP_378476 -----FPIIYNYGFDQESADLSAYLASDG----LVEV 282
      : : : * * : : :

NP_393645 QSGPSLIYGLWNVSSG---IETISGTAEPQSYIFVSNKGQFNNSTASWAPYSG-AHFRFK 404
NP_110756 TGGPSLLYGMWNVSA---ERIIKIEAPSGSFVFIISPGSEFNNTSAVWAPLNG-INTTFY 401
ZP_00609469 TPGPSFIYGMWNIYGSNAMTHVSGTVAPSNAFMFVSTGNTFNASNAQWAPLISGKYSET 417
AA42746 TTGPSFVYGMWNISSNDRMEKFSGRIEFSPNAFAFVSPGCFYNESLASWLPISNGSYNYE 408
NP_394857 TAGPSFIYGMWNVSSVNNMEHFSGSISSPNAFIFASP-DEMNLSTAAMVPTSASGHFSFT 429
CAC11875 NTGPSLLYGMWNVSSVTTMEQFTGSMPSNAFLFVSPGDVINYSQAGYVPLTMSGSYDFW 444
AA480848 TVGTPDYGLLTDSEFHPYIPGFVFNWTSKVPFLVNGT----- 332
NP_343288 TIGRQNPGLLTKFNPSYPSFLYLNISSKIPFLLNKS----- 326
AA481460 ITGNLYKGELTSDFNPPKPYFTFINISSKVPFLINGN----- 327
NP_378476 SVGTPNPPTLLTNYSAPLLPSTVEVTLANSTYNIIDGYNFTYNSIKEYGDNFGLPYYISFT 342
      *

NP_393645 LPKGSYSLAVLRNLYAPEYINDAG--TSPSLNGTSKTGD----- 441
NP_110756 AASGSYSAAVLKNYKQTFYNDINPLTNASLLPKYAG----- 439
ZP_00609469 LPSGQYTAQALLSNYMNA---YFMKQGNVSLNHDYNG----- 452
AA42746 LPAGHYSGISMLSYAFA---YFKPGRITLKNRFYGC----- 443
NP_394857 LPEGTYAYEALMSYAFASG---VLGSDTVISLLKDPMTG----- 466
CAC11875 LPAGQYAGLALLSDHRAQAQ-MLTSSNTFVLPLDMAAG----- 481
AA480848 -----
NP_343288 -----
AA481460 -----
NP_378476 SPNPITVTIYVPKTNNGVIEPYEINVVVEPYTNEEYNTSTTTFTLSNNYFQLIVHYKFLHK 402
      : : :

NP_393645 -----
NP_110756 -----
ZP_00609469 -----
AA42746 -----
NP_394857 -----
CAC11875 -----
AA480848 -----
NP_343288 -----
AA481460 -----
NP_378476 KFYVYINIPMWGI VNGTSM LFKSGYVYVGTIIQLPLVNFTYINSLERYAFFPNVTEIIVI 462
      : : :

NP_393645 -----IYAPVVIENYSDALYYRSNGSQISPFVVISFGQTVSVD 480
NP_110756 VYTPIMIFNNSQIAAYGAFGLGTASSPVAIYLNGSASIS 478
ZP_00609469 IYTPLYAFDQAQIANISLYGNGTAGNPYIMDNVQIMPIS 491
AA42746 IYTPLYAFNNQQLKNISYSGNGTLGNPYIITGKQIFPV 482
NP_394857 IYTPLYAFNNQQLANISVSGTGSSANPVVVENOQYMPIN 505
CAC11875 VYTPPLMAMNNQIANISSSGSGTASNPYMLFNNEYGYIS 520
AA480848 VTNSFVGYVWHVPTITFYRNYTVNSSSYAVLDYPSQS-- 369
NP_343288 LSHAFSGYVTTQIKLGFKNYSINSSFAVLNGNYPS-- 363
AA481460 QLYNFTGYINGPISVVFYSNYSLSNNSFALLNYSSNQKT 366
NP_378476 QNISIHVTEVLQYVIGITSQVPLYGYINNVKQIYSNWFNESCLIKIPSQIYVYVNTETRE 522
      : : :

NP_393645 PLFGVINDYLFVFTGLAIINTSYHIDVENASSFEITYGS---QYDSILASHGLGNTNWM 537
NP_110756 PLFGALNDYFFPVFAGVSI VNVTKHLIIYNSPMSINYTG---HPEYQLKAANLPLNNYL 535
ZP_00609469 SLFEGFNDFAFPVFAVLLHNTNASVVMKNMPSLFIQYTSQNPAYTSYVSDYDLPGYNFL 551
AA42746 PLFEEFNDFVFPVFPVGLIMNWNSTVLYNMPSMYIYD--NPAYENIIDFYDLPGYNYL 540
NP_394857 PLFDEFNDFVFPVFPVGMIAANTNASVNLISMPFYMQYAG---YQVQILKNLKLPTYNYL 562
CAC11875 PLFGEINDYFPPQFAGVMLINTSAYTMASQMPTEFPVLYQG---IFAGYAVFEGLVPPYNYL 577
AA480848 --LTIIPSNTEFKNVIIIPQYTYYYKVIWNSSRPLLLNLSGKETLITNGTLWLPQGTSVSLV 427
NP_343288 --LIEPNVSWFKVLNIIPNYTYYYLVKVNNSQIPVIANVNGKQITLNSTDWFAGQTQISIL 421
AA481460 VRLVINSSTWFDLQFTPNYTFYRLNVISSIPARALINGVNTTLKSG--WYAGNTRIIIVL 425
NP_378476 ILLNPVTILVNSPENYAKWQYQYVNVNNTYPLNGVINNENTTLKSN-WFNASEKIIIP 581
      : : :

NP_393645 GMILYGTSNVSIITGGLNITGWFFETMKGFVVANLMVWNSSNDLIADNVFYSLHASSDDIQ 597
NP_110756 TLQVYGSNISIILNASLISGWFPSTHAGFIVGDLTVWNSTNVLVASNTFVFPFSYFPGETQ 595
ZP_00609469 NYELYNASNVTIWKSNDISGYFASALTGFPVGNLVIWNSTNDLVGSNLFNVMS----- 605
AA42746 NYEFYNDNSNITLWSTP-VSGYFPELISGFPVANVVVVNNTSFLIGSNIFDVMS----- 593
NP_394857 PFEFYNTSNLSLWKDQFISGWFAFYLSGFPDANVIMWNSTHDLIGSNTFYMSMS----- 616
CAC11875 GYWIFNSSHFLYGE-VTGWFTSFDSEPLANVVFVWNTSDSLIASNTFMSMS----- 630
AA480848 NNTYVSSGERFVVVNTYPNLPITVENPTTITVTLKQFKLTINSPIPLLSINGKSRIF 487
NP_343288 N----- 422
AA481460 T----- 426
NP_378476 AQIYVVSNSQREVLLNPTNIVVSSPINYVAKWQYQYVTVSFPVWATIEGESVIFSTGWY 641

```

NP_393645	PPYYSSVLLIYGMNST-----FANNTVWGNVFPDQYVSEN-----	631
NP_110756	SAYFSSLLIYGAPDT-----LENDTIWGNTPFGPIIAG-----	628
ZP_00609469	-----GMLIYKS-----PNVTVWGNVFNAPQTYNSTFEQV	636
AAAT42746	-----GLLTFDS-----HNTTIWGNVFFNDVVDYNNTFENA	624
NP_394857	-----SLLVFGG-----SNNVWVWGNVFLQSPITKTKEFEYI	647
CAC11875	-----SVLIYGG-----MNNTVWGNVFLNSPLLANTQLEQA	661
AAV80848	NGTTWVLQGSVELV-----	502
NP_343288	-----	
AAV81460	-----	
NP_378476	NASEKIVISQNPQYVGKYERWYIYNKTPIVIIVNSPLVVKIYASVQFYVNFMSVILGEVN	701
NP_393645	---GNATVGLFVDSSGNLIYNNAFQYDIAFSPDFNIYTDNSNATYHDNWNISSVEPINYT	688
NP_110756	---VNTTTGLFMDASGNLVYNNFFTEQDIAVSPNINIYTKGNATYHDYWSLSNKEPVNYT	685
ZP_00609469	TDIWGAPLGLAEYSSGDTIYNNFFDITLTAYSPNTSIYSGNPAVYINMWNIT-KQPAYIT	695
AAAT42746	TNIWGAPLGLAVFSSNETVYNNYFDVVITAYSPDISIYTGEHTIYKDLWNIT-KQPASVV	683
NP_394857	N-TNGAPFGVSMYSSGNLVYNNFVKVYITAYSPDYSIYTGLPALYIDENNIS-LQPAFIA	705
CAC11875	WNLWGSPLGLAVFSSNNTIYNNFFATELTAISPSYSIYSGSPAVYINMWNVS-KMPATAV	720
AAV80848	-----NNTYVSSGERYVSSISPSTPIEVSGPMNISVELVKEYKV	543
NP_343288	-----YTYVNGSNERYIISLILPSSSFNVSLPLNITLSTIKQYRV	462
AAV81460	-----TNYVLSNNTRYVITRVEPSTSIVNVNRPINLTVKAQLQYLV	466
NP_378476	GSLELLKSNWYNMYTIIKLNSTYYKYISSERILYIPSIITSINVTAFYVNSFKSITEYFL	761
NP_393645	INVNGYSLNGSIVGGGVEGGNYWGSFFITPPWNESGYIDSG-----YD--YHPLV----	736
NP_110756	IYVNGIPLIGSIIIGGGFEGGNYWGSPPYITGPWNEQGYIESG-----YD--YYPLN----	733
ZP_00609469	HYVNGFALRGSIIIGTHVQGGNYWNNFNGTIPYTNNGLIAYG-----GD--YAPL----	742
AAAT42746	HYFNGFRLYGSIIGTSYQGGNYWVNFNGTIPYNDYGLIYYG-----GD--YEPL----	730
NP_394857	HSPNGFTLSGSIIGTFYQGGNYWVNFNGTIPYNDNGLIYYG-----GD--YVPL----	752
CAC11875	RMFDGLQLTGSIVGGDYQGGNFWVNFNGNIPYNDSGMIEVG-----GD--YEPLNVIDL	772
AAV80848	EIVSSQPVMAIKGE-----	558
NP_343288	LVDSNLPVYVING-----	474
AAV81460	NISGIL-----	472
NP_378476	YFNSTYQLEKINGSEVLLRSGWYNYSETIIPFQYKMINSYSRIALLNPE-----	812
NP_393645	-----YFLYELSFAIHGLPAGTWSVELVADG	761
NP_110756	-----YFSYILFLHASGLPKGLSWQASING	758
ZP_00609469	-----YYNIFFYGIFSHRGM	758
AAAT42746	-----YYNFFYYNF	739
NP_394857	-----VHGFPFYLFRYMF	767
CAC11875	QPPNSVMLSGNVPPYMTAPEFQOYLLNLGPNPDNVINVTLYLNMTNLSQLEQFVALVNS	832
AAV80848	-----	
NP_343288	-----	
AAV81460	-----NITVSGSLTYVADWGIQYVLVILSNHP	839
NP_378476	-----	
NP_393645	VYVVGSGPYINITEFNGN--TTYAVFLPPGYISNASGYAYVTDMSVNVVYARHVMTYG	819
NP_110756	VYVVSNTPYMNVSLYNDGN--SSFVYVLPVGYIYSGENYIYAYKNNTYVNTVNTGVLFS	816
ZP_00609469	-----	
AAAT42746	-----	
NP_394857	-----	
CAC11875	PMSPYFHDYLTQPQEFMASYVPSQSQISQVESYYSAMGFRVWSVAYTPLVIVLQGTVMGFE	892
AAV80848	-----	
NP_343288	-----	
AAV81460	-----	
NP_378476	VYALVNGTNTTLITGWYNKFDEIIIEINISYVNNTVRYSIYNATLLSFRVTSPLTVSVNY	899
NP_393645	IVFHATNVYKNETWYLDIAGSLYSSTSQDLTVSLPNGTYSYSVFAPPN-----	867
NP_110756	VTFNEEGLPGNSTWEVQVAGNIYASISRSITVLLPNGTYTYEIVGPG-----	864
ZP_00609469	-----	
AAAT42746	-----	
NP_394857	-----	
CAC11875	RAFGTMLNFEEFTYPGGSGAVFMTNTANPYIPAQFSLIMHVYGLSYSSEALLSTQNRQD	952
AAV80848	-----TEPVNGTAWLPQGSVIHLTANLPVYLKGYFVGTYN	593
NP_343288	-----ERVNGSVWINAGSSIQLSANVPFYEKGIFGTYN	508
AAV81460	-----TWENNGSEITIPNYKPLLYTMVYEGTYN	500
NP_378476	EPEYLVNINNGKSFVAFNGSGITLYVGSPPFYTVKWIPTYN	939
NP_393645	----NYTAVS-GYVTVNGVSEIVNITFPEQLYRVVIRETGLPAGSSVYVVIDGKAYYSNG	922
NP_110756	----YSPHNSVGYLVNVSQSSSIISVKFTTSMYILTVSEHGLPKGTDWFIIVDQKVVYVYNT	920
ZP_00609469	-----	
AAAT42746	-----	
NP_394857	-----	
CAC11875	LKTEVSGISTLTDQNLNGSSNILTVPVHLAQFYVNTGLEQLGFTGKGVKIGILGVGESANM	1012
AAV80848	-----ASPGEQVTINGPIYEKLVVQLNYTFLEELVPIVITAVVIVVAITRRR	640
NP_343288	-----VTPGSIITVNGPIVETLILSINTELMGIWAVIVIAVVATAAILVLRRRR	556
AAV81460	-----LLPGDTITVTQPISERLIIRLVNVMILIVSIGIITLLTYLVIRKPSN	548
NP_378476	-----VFNKYTIIVHSPITERAVYITNFIVNIVSIGFLLLALISILTNIRIFKKE	988

NP_393645 TEISILLPNG-----RYSVQFFSIPGYKASPOQTAILVSDQPVOY 962
 NP_110756 STIKILLPPG-----RYSIVADTIPGYSKTLSTNTIIINGKNLDV 960
 ZP_00609469 -----
 AAT42746 -----
 NP_394857 -----
 CAC11875 SAISMFWDDQYGIHHPQVVFVNLTTPNGMNPYFEGWEADLDWEWSGAMAPNATIYDVMQPFN 1072
 AAY80848 -----
 NP_343288 -----
 AAY81460 -----
 NP_378476 -----

NP_393645 TEAYTSN--ANYDYVLGAGIIIGAMITAFVYIDRRRR----- 998
 NP_110756 SVTYIQNRDIIFYYVIIASTAAGLIAGAAIGFYIRKK----- 998
 ZP_00609469 -----
 AAT42746 -----
 NP_394857 -----
 CAC11875 LTGIGDNAINLELYYMLNVIHPQIISGSWAELQFHHDGTGFAQIYNLIGLQAAAEGTTIFL 1132
 AAY80848 -----
 NP_343288 -----
 AAY81460 -----
 NP_378476 -----

NP_393645 -----
 NP_110756 -----
 ZP_00609469 -----
 AAT42746 -----
 NP_394857 -----
 CAC11875 GSADSHSIYYLTVMASEYIVSVGGIYPVLNQTGYITGQFGWYQPEYSWYGGPVGSGGGNS 1192
 AAY80848 -----
 NP_343288 -----
 AAY81460 -----
 NP_378476 -----

NP_393645 -----
 NP_110756 -----
 ZP_00609469 -----
 AAT42746 -----
 NP_394857 -----
 CAC11875 FFFARPTYQSSLELIVVPSTFTNRGQPDIAKMPVAFYRDFFGVAGGTSFATPISAGIF 1252
 AAY80848 -----
 NP_343288 -----
 AAY81460 -----
 NP_378476 -----

NP_393645 -----
 NP_110756 -----
 ZP_00609469 -----
 AAT42746 -----
 NP_394857 -----
 CAC11875 ADIESGFLTPNGEVGYFGWIQPALVELGYGYEFLPAYMVDYVQPYPGMVGSGYLQGGW 1312
 AAY80848 -----
 NP_343288 -----
 AAY81460 -----
 NP_378476 -----

NP_393645 -----
 NP_110756 -----
 ZP_00609469 -----
 AAT42746 -----
 NP_394857 -----
 CAC11875 NDFVIGISFNAFNITVDLADYVLDLLIAYGL 1344
 AAY80848 -----
 NP_343288 -----
 AAY81460 -----
 NP_378476 -----