MOLECULAR CLONING OF ALKALINE PROTEASE GENE FROM MARINE FUNGUS ENGYODONTIUM ALBUM

Thesis submitted to the
Cochin University of Science and Technology
under the Faculty of Science
In Partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
In
Biotechnology

by

Jasmin C
Reg. No. 2321

Microbial Technology Laboratory
Department of Biotechnology
Cochin University of Science and Technology
Cochin – 682 022
Kerala, India

January 2007

DEPARTMENT OF BIOTECHNOLOGY
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN-682 022, KERALA, INDIA
Ph: +91 484-2576267 Fax: +91 484-2577595
Email: mchandra@cusat.ac.in

Prof. (Dr.) M. Chandrasekaran 05.01.07

CERTIFICATE

This is to certify that the research work presented in the thesis entitled "**Molecular cloning of alkaline protease gene from marine fungus *Engyodontium album***" is based on the original research work carried out by Ms. Jasmin C under my guidance and supervision at the Department of Biotechnology, Cochin University of Science and Technology, in partial fulfillment of the requirements for the degree of Doctor of Philosophy, and that no part thereof has been presented for the award of any other degree.

M. CHANDRASEKARAN

DECLARATION

I hereby declare that the work presented in this thesis entitled "**Molecular cloning of alkaline protease gene from marine fungus *Engyodontium album***" is based on the original research work carried out by me at the Department of Biotechnology, Cochin University of Science and Technology, Cochin under the guidance of Dr. M Chandrasekaran, Professor of Biotechnology and the thesis or no part thereof has been presented for the award of any degree, diploma, associateship or other similar titles or recognition.

Cochin 22
05.01.07

JASMIN C

Acknowledgement

For a research student, it is a great excitement to work in molecular biology, that is experiencing an information explosion. I am indebted to my guide Dr. M. Chandrasekaran, Professor, Department of Biotechnology, Cochin University of Science and Technology for giving me an opportunity to work in this area. I express my gratitude for his sincere guidance and freedom of experimentation that he provided to me during the course of my study. Dr. Chandrasekaran's dedication to his students and unending desire to see his students grow, both academically and professionally, has been a source of inspiration and support throughout my studies.

I would like to thank the Department of Biotechnology, Government of India, for the financial support to conduct this research through the research grant BT/PR2203/AAQ/03/109/2000 dated 16-08-2002. I am also thankful to Council of Scientific and Industrial Research (CSIR), for providing fellowship during the initial period of my research and Cochin University of Science and Technology for financial support during the last period.

I wish to thank Head, Department of Biotechnology and Dean, Faculty of Science for facilitating my registration and providing administrative support.

Without the encouragement and participation of Dr. K.K. Elyas, Lecturer and Dr. Sarita G. Bhat, Senior Lecturer, Department of Biotechnology, CUSAT this work would never have been accomplished. My personal and very deep gratitude goes to both of them for sharing their ideas and making recommendations to improve the work. I thank Dr. K.K. Elyas for being with me during the arduous works on column packing for cDNA fractionation and library screening. I should specifically thank Dr. Sarita G. Bhat for the request letter she send to Centre for Cellular and Molecular Biology (CCMB), Hyderabad, which gave me an opportunity to work at one of the best Molecular biology labs in India.

An important requirement of my research was the PCR amplification of partial gene, southern analysis using radiolabeling and gene sequencing. I could attain this only because of the broad minded scientific enthusiasm of Dr. Kshitish Majumdar, Scientist, CCMB, Hyderabad. I take this opportunity to pay my heart-felt gratitude to him for allowing me to work at CCMB. I gratefully remember the support and affection rendered by all his students especially my friend Dr. Rajesh Ramachandran who spent many hours for helping me. In addition to being a dear friend, Rajesh is one of the most generous, considerate and brightest people I have ever met. Thank you for all the assistance you have given me.

I always tried to learn from the methodical and patient approach to research by Rajeevettan, my any time approachable senior. The patronizing advice rendered by him through out the tenure of the work was invaluable and it will not be an exaggeration if I fail to express my gratitude to him for want of right words. Thank you very much for all your help, particularly for teaching me the basics of homology modelling and further analysis based on that.

I thankfully remember the recommendations given by Dr. Takeshi Akao, National Research Institute of Brewing, Japan when I was confused with my RNA extraction experiments. I am especially grateful to Dr. P.E. Kolattukudy, Ohio State University, USA

for providing me cDNA for using as heterologous probe in my library screening experiments. In this connection I am also thankful to Dr. Kapil Deo Singh, Professor and Head, Department of Biotechnology, N.D. University of Agriculture & Technology, Uttar Pradesh for taking me through the basics of RNA isolation and phage library handling.

My interests in molecular biology research started during MSc dissertation when working with Dr. Nirmal Babu, Senior Scientist, Indian Institute of Spice Research, Calicut, who initiated me in to this field, for which I will be indebted to him for all my life.

Many of my friends have been involved in the successful completion of this research work. I would first like to thank Bernard Rajeev for being with me when I had problems and questions.

Teamwork is the core of the success of any research lab and I express my huge thanks to all the members of MTL family Soorej, Archana Kishore, Bijna, Madhu, Lailaja, Jissa and Beena, for their kind affection, extensive cooperation and timely assistance. I also thank Siju and Manjusha from our sister lab, Microbial Genetics Laboratory for their cooperation.

I am thankful to all the research scholars of the Neuroscience laboratory, Department of Biotechnology for their cooperation and friendship.

I am also deeply obliged to Dr. K. Seetha, Syngene, Bangalore for teaching me the basics of primer designing and sequence analysis using different bioinformatics tools.

I thankfully remember the suggestions and advice given by Dr. Anup Kumar Kesavan, Johns Hopkins University, USA.

The most difficult part in writing any thesis is the reference section, which was made very easy by the 'Endnote' software, which was thankfully provided by my friend Dr. P.V. Preejith, Research Scientist, Bioengineering Division, Research Technoplaza, Singapore. I am thankful to him for all the support and all other necessary miscellaneous help.

I wish to thank my senior and project colleague Dr. Sreeja Chellappan, who spent many hours with me discussing our research and other things.

I take this opportunity to extend my admiration and gratitude to my seniors at department of Biotechnology, Dr. Sabu, Dr. Renuka, Dr. Swapna, Dr. Ani Das, Dr. Dasan, Dr. Keerthi, Dr. Biju and Dr. Jackson for their support, constructive criticism and encouragement.

I am thankful to my teachers at Department of Biotechnology Dr. C.S. Paulose, Dr. Padma Nambeeshan and Dr. Shyla Raj for the lessons they taught me during my masters course, which were very much useful for me during my research period.

I really appreciate the efforts taken by the administrative staff of Department of Biotechnology and University for processing my papers as early as they could.

I gratefully acknowledge my good friends and masters course classmates Hareesh, Seetha, Deepu, Maya, Sreeja and Geetha, for their moral support and encouragement through out the research period. Special thanks to Deepu for his assistance at CCMB.

I thank my intimate juniors Riyas, Santhosh Varghese and Bindhu for their friendship and also for finishing the products of my cooking experiments several times. Special thanks to Riyas for his expertise in designing the cover page of the thesis and editing the pictures.

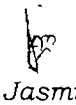
It would be criminal to sign off without acknowledging my friends Vrinda, Roshni, Sandhya, Jyotsana, Aishwarya, Sofia, Leena Varier and Pravitha who made my hostel days very much memorable and enjoyable.

The most difficult but the most important part of writing this acknowledgement is to express the love and gratitude I feel towards two people I cherished most, my husband and my daughter. I wish to offer very special thanks to my husband and best friend, Anas who has been with me since the beginning of my research and has provided the emotional base and companionship without which I would be far less of a person and a researcher. I thank my daughter Reyna for her innocent smiles, which made me happy even during the most distressed times.

Finally and importantly, I wish to extend my appreciation to my parents and sister for their patience and encouragement, especially to my father for his support and patience through out these long, challenging years.

In conclusion I would like to give a word of gratitude to all those who have worked and generated valuable findings and data in this subject and my best wishes to all those who will continue to do so in future.

Above all I thank lord almighty for the benevolence and blessings he showered on me.



Jasmin

LIST OF ABBREVIATIONS

Abbreviations for standard units (SI) of measurement and chemical formulae are not included in the list given below

A	Absorbance
aa	Amino acid
ATP	Adenosine triphosphate
bp	Base pairs
BSA	Bovine serum albumin
cDNA	complementary DNA
cpm	Counts per minute
°C	Degree Celsius
Da	Dalton
dATP	deoxyadenosine 5'-triphosphate
dUTP	deoxyuridine 5'-triphosphate
DEPC	diethyl pyrocarbonate
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxy-ribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
Fig	Figure
g	gram
gDNA	Genomic DNA
hrs	Hours
Ig G	Immunoglobulin G
kb	Kilo base pair
kDa	Kilo Dalton

l	litre
mm	Milli meter
M	Molar
µg	Microgram
µl	Micro litre
ml	Milli litre
mM	Milli molar
mRNA	Messenger RNA
MW	Molecular weight
nmol	Nano mol
N	Normality
ng	Nano gram
nm	Nano meter
No	Number
nt	Nucleotide
OD	Optical density
ORF	Open reading frame
^{32}P	Radioactive phosphorous
PCR	Polymerase chain reaction
pfu	Plaque forming unit
pg	Pico gram
PMSF	Phenyl methyl sulfonyl fluoride
POPOP	1,4-bis (5-phenyloxazolyl) benzene
PPO	2,5, Diphenyloxazole
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribosomal RNA
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction

rpm	Revolutions per minute
rcf	Relative centrifugal force
SDS	Sodium dodecyl sulphate
Sp	Species
Tris	Tris (hydroxy methyl) aminomethane
U	Unit
UTR	Untranslated region
UV	Ultraviolet
V/V	Volume/volume
W/V	Weight/volume
%	Percentage

CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1

INTRODUCTION

1.1 Preface

Enzymes have attracted attention from researchers all over the world because of the wide range of physiological, analytical and industrial applications, especially, from microorganisms, because of their broad biochemical diversity, feasibility of mass culture and ease of genetic manipulation. Despite the fact that more than 3000 different enzymes have been identified and many of them have found their way into biotechnological and industrial applications, the present enzyme toolbox is not sufficient to meet most of the industrial demands. In view of these limitations, researchers have diverted their attention for isolation and characterization of enzymes from different environments. Whenever required, due attention is also paid towards development of recombinant enzymes with desired characteristics and for specific applications.

Microbial proteases constitute approximately 40% of the total enzyme sales in various industrial market sectors, such as detergent, food, pharmaceuticals, leather, diagnostics, waste management and silver recovery besides application as molecular biology reagents. Bacteria are the most abundant group of alkaline protease producers with the genus *Bacillus* being the most prominent source (Anwar & Saleemuddin 1998, Gupta et al. 2002) among the various groups of microorganisms (Table – 1.1).

Table 1.1: Commercial microbial alkaline proteases, sources, application and their industrial suppliers (Gupta et al. 2002)

Product name	Microbial Source	Application	Supplier
Alcalase	*B. licheniformis*	Detergent, Silk	
Savinase	*Bacillus.* sp	Degumming	
Esperase	*B. lentus*	Detergent, textile	
Biofeed pro	*B. licheniformis*	Detergent, food,	Novo Nordisk,
Durazym	*Bacillus* sp	Silk degumming	Denmark
Novozyme 471 MP	*Ns*	Feed	
Novozyme 243	*B. licheniformis*	Detergent	
Nue	*Bacillus.* sp	Photographic gelatin hydrolysis, Denture cleaners, Leather	
Purafact	*B. lentus*	Detergent	
Primatan	Bacterial source	Leather	Genencor, USA
Subtilisin	*B. alcalophilus*	Detergent	
Maxacal	*Bacillus* sp.	Detergent	Gist-Brocades, The
Maxatase	*Bacillus* sp.	Detergent	Netherlands
Opticlean	*B. alcalophilus*	Detergent	
Optimase	*B. licheniformis* Protein engineered variant of *Bacillus* sp	Detergent	Solvay Enzymes, Germany
Maxapem	*B. subtilis*	Alcohol, baking, brewing, feed, food, leather,	
HT-proteolytic Protease	*B. licheniformis*	Photographic waste Food, waste	
Proleather	*Bacillus* sp.	Leather	
Collagenase	*Clostridium* sp.	Technical	Amano
Amano protease S	*Bacillus* sp.	Food	Pharmaceticals, Japan

Contd.

Enzeco alkaline protease	*B. licheniformis*	Industrial	
Enzeco alkaline protease -L FG	*B. licheniformis*	Food	Enzyme Development, USA
Enzeco high alkaline protease	*Bacillus* sp.	Industrial	
Bioprase concentrate	*B. subtilis*	Cosmetic, pharmaceuticals	
Ps. protease	*Pseudomonas aeruginosa*	Research	
Ps. elastase	*Pseudomonas aeruginosa*	Research	Nagase Biochemicals, Japan
Cryst. protease	*B. subtilis* (K2)	Research	
Bioprase	*B. subtilis*	Detergent, cleaning	
Bioprase SP-10	*B. subtilis*	Food	
Godo-Bap	*B. licheniformis*	Detergent, food	Godo Shusei, Japan
Corolase 7089	*B. subtilis*	Food	Rohm, Germany
Wuxi	*Bacillus* sp.	Detergent	Wuxi Synder Bioproducts, China
Protosol	*Bacillus* sp.	Detergent	Advance Biochemicals, India

(*Ns.* Not specified)

Among fungi, species of *Aspergillus* (Rajamani & Hilda 1987, Chakrabarti et al. 2000) is the most exploited group, and the genus *Conidiobolus* (Bhosale et al. 1995), *Rhizopus* (Banerjee & Bhattacharya 1993), *Beauveria* (Joshi et al. 1995), *Metarhizium* spp (St. Leger et al. 1992) are also known to produce alkaline proteases. The most popular fungal alkaline serine protease "proteinase K" is produced by *Tritirarchium* sp. Fungal proteases have an added advantage over bacterial proteases for they are obtained in large quantities from the biological source. Mycelia free culture broth can be easily obtained by conventional filtration of the fungi, whereas cost-intensive filtration technology is required for downstream processing of the bacterial enzyme.

Increasing demand of proteases with specific properties has lead biotechnologists to explore newer sources of the enzyme. Taking into consideration the depth of microbial diversity, the scope for recognizing potential microorganisms capable of producing novel protease with desirable properties suitable for commercial exploitation is very promising. Of late, marine microorganisms have emerged as an abundant source for the isolation of potential industrial enzymes (Suresh & Chandrasekaran 1999, Gupta et al. 2002). Due to their unique natural habitat these marine microorganisms show distinct physiological characteristics, metabolic patterns and nutrient utilization when compared to their terrestrial counterparts. The optimum activity of marine bacterial enzymes usually occurs at high salinity, making these enzymes utilizable in many harsh industrial processes, where the concentrated salt solutions used would otherwise inhibit many enzymatic transformations. In addition, most marine microbial enzymes are considerably thermo tolerant, remaining stable at room temperature over long periods. Therefore, the enzymes from marine microorganisms should be expected to possess unique properties.

Although few investigators have reported production of exoenzymes by marine microorganisms, there is a dearth of knowledge on marine microbial genetics as well as on the molecular biology of marine microbial enzymes. To be particular there is absolute lack of information at molecular level on marine fungi. Whereas, there is enormous scope for effective utilization of marine fungi for diverse industrial applications if adequate information is made available through molecular studies. In this context the present study assumes greater importance since it probes into the isolation of intact gene encoding alkaline serine protease from a marine fungus, sequencing the nucleotides and an in depth analysis of the sequence employing bioinformatics tools towards drawing an insight on the structure and function of the enzyme protein. To the best of our knowledge this thesis is the first of its kind with reference to marine fungus.

1.2 Scope of the present study

During the course of another investigation in our laboratory a fungus identified as *Engyodontium album* was isolated from marine sediments from Kochi, South coast of India (Suresh & Chandrasekaran 1999) as a potential source of alkaline protease enzyme (Sreeja Chellappan et al. 2005). This enzyme is produced as a major protein by the fungus, among other enzymes it secreted extracellularly. The enzyme has been characterized for its physico-chemical and biochemical properties and the molecular weight was estimated as 28,000 Da. Based on the results obtained, the enzyme is classified within the subtilisin family of serine protease (Sreeja Chellappan et al. 2005). The various studies conducted with the alkaline serine protease during that study generates curiosity and keen interest in undertaking the present study towards isolation of the intact gene encoding alkaline serine protease from *Engyodontium album* (*Eap*), sequencing the entire gene and its analysis using various bioinformatics tools and techniques.

Knowledge of full nucleotide sequence of the protease gene will facilitate the deduction of the primary structure of the encoded enzyme and also in identification of various functional regions. Knowledge on the gene sequence will enable phylogenetic analysis of the enzyme, prediction of the secondary structure of the protein and subsequent study of structure–function relationships of the enzyme. Further the information on nucleotide sequence would also provide an excellent means for the manipulation and control of genes towards harnessing the enzyme for various applications.

1.3 Objectives

Specific objectives of the present study includes the following:

1. Construction of cDNA library from *Engyodontium album*
2. Construction of genomic DNA library of *E. album*
3. Preparation of Probes
4. Screening of DNA libraries
5. Isolation of intact gene encoding alkaline serine protease
6. Nucleotide sequencing
7. Sequence analysis using Bioinformatics tools

Chapter 2

REVIEW OF LITERATURE

2.1 Economic importance of Protease

A wide range of enzymes are excreted by fungi and play an important role in the breakdown of organic materials; many of them are now produced commercially (Table 2.1) (Bennett 1998, Hamlyn 1998). This capacity for enzyme secretion enables fungi to be a principal group of microorganisms responsible for nutrient recycling. The main polymeric components of organic matter found in nature are derived from plant cell walls, which include cellulose, hemicellulose, lignin and pectin. Other polymers available are starch, protein and lipid. Fungi secrete a wide range of enzymes capable of degrading each of these polymers even when the polymers are naturally found in association with each other. Fungal genome can encode a very large range of different degradative enzymes and many of them are harnessed in biotechnology.

Table 2.1: Fungal enzymes of commercial importance

Enzyme	**Major Source**
Asparaginase	*Aspergillus* spp. and *Penicillium* spp.
Amylase	*Aspergillus niger* and *Aspergillus oryzae*
Catalase	*A. niger* and *Penicillium* spp.
Cellulase	*A.niger*, *Trichoderma reesei*, *T. viride* and *P. finiculosum*
Dextranase	*Penicillium* spp.
ß-Glucanase	*A. niger*, *Penicillium emersonii*, *T. reesei* and *T. viride*
Glucoamylase	*A. niger* and *A. oryzae*
Invertase	*A. niger* and *A. oryzae*
Hemicellulase	*A. niger*, *A. oryzae*, *T. reesei*, *T. viride*, *P. emersonii*
Laccase	*Pyricularia oryzae* and *Coriolus versicolor*
Lipase	Several species including *A.niger* and *A. oryzae*
Pectinase	*A. niger*, *Rhizopus oryzae* and *Humicola insolens*
Protease	*A. niger*, *A. oryzae*, *A.melleus*, *Rhizopus delemar* and *Engyodontium album*
Rennin	*Mucor miehei*, *Mucor pusillus* and *Endothia parasitica*
Tannase	*A.niger* and *A. oryzae*
Xylanase	*A. niger* and *T. reesei*

The current estimated value of the worldwide sales of industrial enzymes is $1 billion (Godfrey & West 1996). Of the industrial enzymes, 75% are hydrolytic (Rao et al. 1998). 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). Proteases are enzymes that catalyze the hydrolysis of proteins, occupying an important position with respect to their applications in both physiological and commercial fields (Rao et al. 1998). They are involved in a large variety of functions, extending from ordinary growth to unique development. They are responsible for the complex processes involved in the normal physiology of the cell as well as in abnormal pathophysiological conditions. Their involvement in the lifecycle of disease causing organisms has led them to become a potential target for developing therapeutic agents against fatal diseases such as cancer and AIDS. Proteases have a long history of application in the food and detergent industries. Their application in the leather industry for dehairing and bating of hides to substitute currently used toxic chemicals is a relatively new development and has conferred added biotechnological importance (Rao et al. 1998). The vast diversity of proteases in contrast to the specificity of their action has attracted worldwide attention to exploit their physiological and biotechnological application (Poldermans 1990, Fox et al. 1991).

Proteinases are widely produced amongst fungi and serve a number of different roles within fungal systems, including nutrient cycling, activation of zymogens, post translational processing, protein turnover, cell autolysis (Chung & Goldberg 1981, North 1982, Santamaria & Reyers 1988, Desphande 1992, Gottesman & Maurizi 1992) and sporulation (Gomez et al. 1995, Hofmeister et al. 1995, Stragier 1996). Serine proteinases are generally involved in nutrient acquisition (Segers et al. 1999). The action of serine proteinases is especially important during periods of nitrogen starvation, as proteins are a source of both nitrogen and carbon. Most fungal serine proteinases isolated so far belong to the Proteinase K subfamily of subtilases (Siezen & Leunissen 1997). This family of enzymes tends to share a high degree of similarity (>55% identity) and the structure of

these enzymes can generally be inferred from the crystal structure of proteinase K (Siezen et al. 1991) isolated from *Tritirarchium album* (Kolvenbach et al. 1990).

Proteases of fungi have widely been used for various industrial applications. The greatest industrial use of protease is for laundry detergents where they help to remove protein based stains (such as blood and egg) from clothing (Rao et al. 1998). The second largest use of protease is for cheese making. Enzymes from calf stomach and microbial sources are used to clot milk, one of the first steps in cheese making. Proteases are also used for bating (softening) leather, modifying food ingredients (e.g. soy protein whipping agents), meat tenderizers, and flavor development (Cohen 1977, Godfrey & West 1996, Anwar & Saleemuddin 1998, Hamlyn 1998, Rao et al. 1998, Ganesh Kumar & Takagi 1999, Wedde et al. 1999, Gupta et al. 2002, Sreeja Chellappan 2005).

2.2 Classification of proteases

According to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, proteases are classified in subgroup 4 of group 3 (hydrolases) (Inoue et al. 1991). However, proteases do not comply easily with the general system of enzyme nomenclature due to their huge diversity of action and structure. Proteases are classified on the basis of three major criterions such as type of reaction catalyzed, chemical nature of the catalytic site, and evolutionary relationship with reference to structure.

Proteases are grossly subdivided into two major groups, i.e., exopeptidases and endopeptidases, depending on their site of action (Figure 2.1). Exopeptidases cleave the peptide bond proximal to the amino or carboxyl termini of the substrate, whereas endopeptidases cleave peptide bonds distant from the termini of the substrate.

Based on the functional group present at the active site, proteases are further classified into four prominent groups, i.e., serine proteases, aspartic proteases, cysteine proteases, and metalloproteases (Hartley 1960). There are few miscellaneous proteases, which do not precisely fit into the standard classification, e.g., ATP-dependent proteases which require ATP for activity (Menon & Goldberg 1987).

Figure 2.1: Classification of protease enzymes (Rao et al. 1998)

Protease	Mode of action[a]	EC no.
Exopeptidases		
Aminopeptidases	●↓○-○-○-○---	3.4.11
Dipeptidyl peptidase	●-●↓○-○-○---	3.4.14
Tripeptidyl peptidase	●-●-●↓○-○---	3.4.14
Carboxypeptidase	---○-○-○-○-○↓●	3.4.16–3.4.18
Serine type protease		3.4.16
Metalloprotease		3.4.17
Cysteine type protease		3.4.18
Peptidyl dipeptidase	---○-○-○-○↓●-●	3.4.15
Dipeptidases	●↓●	3.4.13
Omega peptidases	*-●↓○-○---	3.4.19
	---○-○-○↓●-*	3.4.19
Endopeptidases	----○-○-○↓○-○-○---	3.4.21–3.4.34
Serine protease		3.4.21
Cysteine protease		3.4.22
Aspartic protease		3.4.23
Metalloprotease		3.4.24
Endopeptidases of unknown catalytic mechanism		3.4.99

[a] Open circles represent the amino acid residues in the polypeptide chain. Solid circles indicate the terminal amino acids, and stars signify the blocked termini. Arrows show the sites of action of the enzyme.

Based on their amino acid sequences, proteases are classified into different families (Argos 1987) and further subdivided into clans to accommodate sets of peptidases that have diverged from common ancestor (Rawlings & Barrett 1993). Families are grouped by their catalytic type, the first character representing the catalytic type: A, aspartic; C, cysteine; G, glutamic acid; M, metallo; S, serine; T, threonine; and U, unknown. A clan that contains families of more than one type is described as being of type P.

Proteolytic enzymes that exploit serine in their catalytic activity are ubiquitous, being found in viruses, bacteria and eukaryotes (Rawlings & Barrett 1993, 1994). They include a wide range of peptidase activity, including exopeptidase, endopeptidase, oligopeptidase and omega-peptidase activity. Over 20 families (denoted S1 - S27) of serine protease have been identified, these being grouped into 6 clans (SA, SB, SC, SE, SF and SG) on the basis of structural similarity and other functional evidence (Rawlings & Barrett 1994). Structures are known for four of the clans (SA, SB, SC and SE): these appear to be totally unrelated, suggesting at least four evolutionary origins of serine peptidases and possibly many more (Rawlings & Barrett 1994).

The serine endoprotease are divided into two super families that independently evolved similar catalytic mechanisms. The trypsin super family includes the trypsins and chymotrypsins that are ubiquitous in animals. The subtilase super family is similarly ubiquitous in bacteria and fungi. The classification of (Siezen & Leunissen 1997) delineates three families of subtilases in fungi. The Proteinase K family (fungal or class II) named after an enzyme found in the ascomycete *Tritirarchium album*, the subtilisin family (also called bacterial or class I) which until recently were only thought to be found in bacteria, and kexins (preprotein convertases) that are also found in animals (Hu & St. Leger 2004).

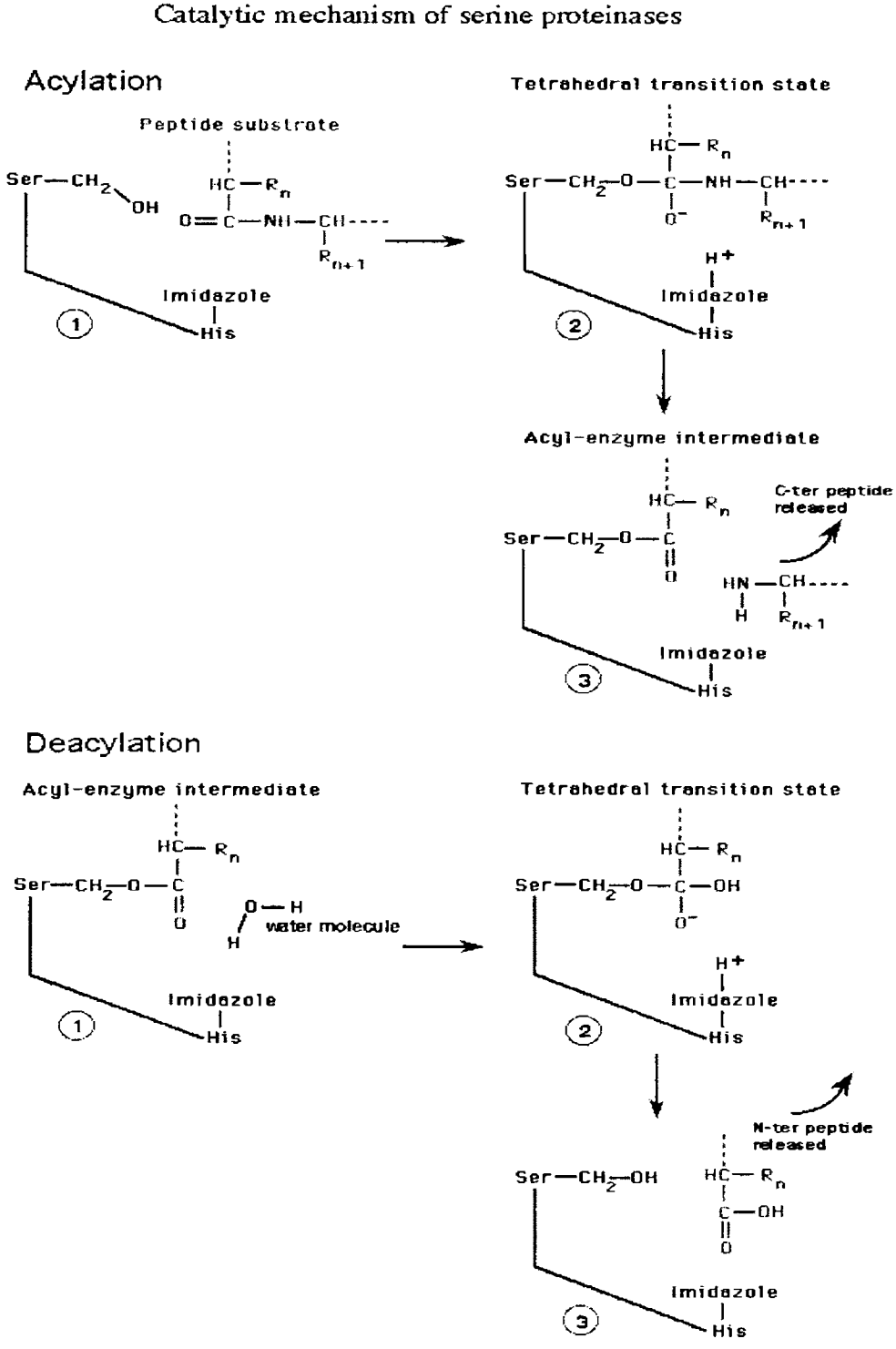
The subtilisin family is the second largest serine protease family characterised to date. Over 200 subtilises are presently known, more than 170 of which with their complete amino acid sequence (Siezen & Leunissen 1997). It is widespread, being found in eubacteria, archaebacteria, eukaryotes and viruses (Rawlings & Barrett 1994). The vast majority of the family is endopeptidases, although there is an exopeptidase and tripeptidyl peptidase (Rawlings & Barrett 1993, 1994). Structures have been determined for several members of the subtilisin family: they exploit the same catalytic triad as the chymotrypsins, although the residues occur in a different order (HDS in chymotrypsin and DHS in subtilisin), but the structures show no other similarity (Rawlings & Barrett 1993, 1994). Some subtilisins are mosaic proteins, and others contain N- and C-terminal extensions that show no sequence similarity to any other known protein (Rawlings & Barrett 1994).

The proprotein-processing endopeptidases kexin, furin and related enzymes form a distinct subfamily known as the kexin subfamily (S8B). These preferentially cleave C-terminally at paired basic amino acids. Members of this subfamily can be identified by subtly different motifs around the active site (Rawlings & Barrett 1993, 1994). Sedolisins (serine-carboxyl peptidases) are proteolytic enzymes whose fold resembles that of subtilisin; however, they are considerably larger, with the mature catalytic domains containing approximately 375 amino acids. The defining features of these enzymes are a unique catalytic triad, Ser-Glu-Asp, as well as the presence of an aspartic acid residue in the oxyanion hole (Wlodawer et al. 2003).

2.3 Mechanism of action of serine protease

The serine proteinases exhibit different substrate specificities, which are related to amino acid substitutions in the various enzyme subsites interacting with the substrate residues (Schechter & Berger 1967). Some enzymes have an extended interaction site with the substrate whereas others have a specificity restricted to the P1 substrate residue. Three residues which form the catalytic triad is essential in the catalytic process i.e His, Asp and Ser. The first step in the catalysis is the formation of an acyl enzyme intermediate between the substrate and the essential Serine. Formation of this covalent intermediate proceeds through a negatively charged tetrahedral transition state intermediate and then the peptide bond is cleaved. During the second step or deacylation, the acyl-enzyme intermediate is hydrolyzed by a water molecule to release the peptide and to restore the Ser-hydroxyl of the enzyme. The deacylation, which also involves the formation of a tetrahedral transition state intermediate proceeds through the reverse reaction pathway of acylation. A water molecule is the attacking nucleophile instead of the Ser residue. The His residue provides a general base and accept the -OH group of the reactive Ser (From http://delphi.phys.univ-tours.fr/Prolysis/).

Figure 2.2: Mechanism of action of serine protease enzyme



2.4 Molecular biology of extracellular proteases of fungi

Extracellular fungal proteases are usually secreted as zymogen/ precursors such as the precursors of the mammalian serine proteinases (eg. trypsinogen, chymotrypsinogen, proelastase ect). This is to mask their proteolytic activity during their intracellular biosynthesis and transportation. This enzyme precursor comprises of a signal peptide, an N-terminal propeptide and a mature protease polypeptide. The signal peptide assists the translocation of the zymogen into the Endoplasmic reticulum, which later gets cleaved away from the enzyme within the endoplasmic reticulum lumen by a signal peptidase. The N-terminal propeptide generally functions as an intramolecular chaperone and is cleaved away from the mature domain after the completion of enzyme folding (Eder et al. 1993, Yabuta et al. 2001). The inactive enzyme precursor becomes activated by this proteolytic cleavage. Concerning the processing of the pro region, there are two types of cases reported. In the first case a processing enzyme is involved in the processing of pro region. The second type of processing is catalysed by autoproteolytic activity of the protease itself (Tatsumi et al. 1989, Tatsumi et al. 1991).

Identification of zymogens at the protein level requires a transient presence, at least, in stable form. The gene sequence however provides unequivocal evidence for the existence of precursor forms (Neurath 1975). The mature regions of the serine proteases in common possess a catalytic triad consisting of Ser, Asp and His (Siezen & Leunissen 1997).

2.5 Protease gene cloning in Fungi

Gene encoding protease enzyme in fungi has been the subject of study for several years owing to the importance of the enzyme in various fields of applications including medical, agricultural and industrial. In medical field the studies were oriented towards investigating the role of these enzymes in exhibiting virulence and then the control of disease by targeting the enzyme. Whereas in industrial field, the aim of such studies were

pointed towards: (i) overproduction of the enzyme either by gene dosage effect or by modifying the gene regulation events, (ii) protein engineering to locate the active site residue and/or to alter the enzyme properties to suit its commercial applications, (iii) regulation/suppression of proteases which deleteriously reduce the yield of heterologous protein production and (iv) development of new expression hosts for recombinant protein production. In agricultre, the studies are targeted towards (i) studying the crucial role of proteases in the infection and colonization processes of phytopathogenic fungi and (ii) unravelling the importance of proteolytic activity in biocontrol processes and then use genetic engineering to increase the pathogenecity and survival of the introduced host. A summary of the various investigations reported in the literature on cloning and sequencing of protease gene in various species of fungi and their significance is presented in Table 2.2.

Gene manipulation studies were conducted for the purpose of overproduction of the enzyme either by gene dosage effect or by modifying the gene regulation events and regulation/suppression of proteases, which deleteriously reduce the yield of homologous/heterologous protein production. In several cases, proteases are actively involved in the post-secretional processing of extracellular hydrolytic enzymes. Though some of these activities are proved to be harmful to homologous and/ or heterologous protein production, a few of them are reported to be useful in the correct processing of the recombinant fusion proteins.

Table 2.2: A summary of various investigations on cloning and sequencing of protease gene in various species of fungi and their significance

Organism	Protease enzyme	Significance	Reference
Absidia zychae	Serine carboxypeptidase *scpZ* gene	Industrial strain (koji culture)	Lee et al. (1995)
Acremonium chrysogenum	Alkaline protease	Industrial – protease production	Isogai et al. (1991)
Agaricus bisporus	Serine proteinase	Mushroom senescence	Kingsnorth et al. (2001)
Aphanomyces astaci	Two serine proteinase	Fish pathogen	Bangyeekhun et al. (2001)
Arthrobotrys oligospora	Cuticle-degrading serine protease	Biocontrol-nematode trapping	Ahman et al. (1996)
Arthrobotrys oligospora	Subtilisin	Biocontrol-nematode trapping	A°hman et al. (2002)
Aspergillus awamori	Aspergillopepsin B (*pepB*)	Protease deficient industrial strain	Moralejo et al. (2002)
Aspergillus awamori	Aspergillopepsin A	Protease deficient industrial strain	Berka et al. (1990)
Aspergillus flavus.	Metalloproteinase (*mep20*)	Human pathogen	Ramesh et al. (1995)
Aspergillus flavus.	Elastinolytic serine protease	Human pathogen	Ramesh et al. (1994)
Aspergillus fumigatus	Extracellular Alkaline serine protease	Human pathogen	Jaton-Ogay et al. (1992)
Aspergillus fumigatus	Elastinolytic serine protease	Human pathogen	Kolattukudy et al. (1993)
Aspergillus fumigatus	Elastinolytic metalloproteinase	Human pathogen	Sirakova et al. (1994)
Aspergillus fumigatus	Metalloproteinase (*mep20*)	Human pathogen	Ramesh et al. (1995)
Aspergillus fumigatus	Alkaline protease	Bird pathogen	Katz et al. (1998)
Aspergillus fumigatus.	Extracellular aspartic proteinase, *alp 2.*	Human pathogen	Reichard et al. (1995)
Aspergillus nidulans	*KpcA* gene, *kex-2* like endoprotease	Proprotein/ recombinant fusion protein processing	Kwon et al. (2001)
Aspergillus nidulans	Extracellular protease	Protease deficient industrial strain	Vankuyk et al. (2000)
Aspergillus nidulans	Alkaline protease	Industrial	Katz et al. (1994)
Aspergillus nidulans	Intracellular carboxypeptidase, *cpy A*	Vacuolar biogenesis and intracellular transport of Vacuolar proteins	Ohsumil et al. (2001)
Aspergillus niger	Acid proteinase A	Industrial	Inoue et al. (1991)
Aspergillus niger	*PepC* gene, serine protease	Industrial	Frederick et al. (1993)
Aspergillus niger	The Kexin-Like Maturase	Recombinant fusion protein processing	Jalving et al. (2000)
Aspergillus niger	Aspartic protease *pepE*	Industrial	Jarai et al. (1994b)
Aspergillus niger	Subtilisin-like protease *pepD*	Industrial	Jarai et al. (1994a)

Contd.

Aspergillus niger	Three acid proteases	Protease deficient industrial strain	van den Hombergh et al. (1997)
Aspergillus niger	*PepF*	Industrial	van den Hombergh et al. (1994)
Aspergillus niger	Proprotein convertase	Proprotein/ recombinant fusion protein processing	Heerikhuisen et al. (2000)
Aspergillus niger var. macrosporus	Proctase B	Industrial strain	Lu et al. (1995)
Aspergillus oryzae	Acid protease-encoding gene (*pepA*)	Industrial over production of protease	Gomi et al. (1993)
Aspergillus oryzae	Alkaline protease	Industrial strain	Tatsumi et al. (1989)
Aspergillus oryzae	Alkaline protease	Industrial	Murakami et al. (1991)
Aspergillus oryzae	Alkaline protease, *alpA*	Industrial over production of protease	Cheevadhanarak et al. (1991)
Aspergillus oryzae	Aspergillopepsin O	Industrial strain	Berka et al. (1993)
Beauveria bassiana	Cuticle degrading protease, bassiasin I	Biocontrol-entamopathogen	Kim et al. (1999)
Beauveria bassiana	Cuticle degrading protease, *Prl 1*	Biocontrol-entamopathogen	Joshi et al. (1995)
Beauveria bassiana.	Cuticle degrading enzyme CDEP-1	Biocontrol-entamopathogen	Fang et al. (2002)
Beauveria brongniartii	Cuticle degrading serineprotease, *Prl 1.*	Biocontrol-entamopathogen	Sheng et al. (2006)
Botrytis cinerea	Aspartic proteinase	Plant pathogen	Have et al. (2004)
Candida albicans	Endoprotease B, *PRB1*	Human pathogen	Orozco et al. (2002)
Candida boidinii	Proteinase A gene. (*PEP4*)	Heterologous protein expression host	Komeda et al. (2002)
Candida boidinii	Proteinase B gene. (*PRB* 1)	Heterologous protein expression host	Komeda et al. (2002)
Coccidioides immitis	Aspartyl protease	Human pathogen	Johnson et al. (2000)
Cochliobolus carbonum	Three extracellular proteases	Plant pathogen	Murphy & Walton (1996)
Cryphonectria parasitica	Acid proteinases *eapB* and *eapC* genes	Plant pathogen	Jara et al. (1996)
Cryphonectria parasitica	Endothiapepsin (aspartic proteinase)	Cheese industry-milk clotting	Jara et al. (1995)
Dermatophilus congolensis	Serine protease, *nasp*	Animal pathogen	Garcia-Sanchez et al. (2004)
Fusarium oxysporum.	Subtilase, MPM1	Plant pathogen	Pietro et al. (2001)
Glomerella cingulata	A secreted aspartic proteinase.	Plant pathogen	Clark et al. (1997)
Hansenula polymorpha	Proteinase A, (PEP4)	Recombinant expression host	Bae et al. (2005)

Contd.

Lecanicillium psalliotae	Extracellular serine protease	Biocontrol-nematode trapping	Yang et al. (2005)
Leptosphaeria maculans	Subtilisin-like serine protease	Plant pathogen	Wilson & Barbara (2005)
Magnaporthe grisea	Vacular serine protease	Plant pathogen	Fukiya et al. (2002)
Metarhizium anisopliae	Cuticle degrading protease	Biocontrol-entamopathogen	St. Leger et al. (1992)
Metarhizium anisopliae	Cuticle-degrading protease	Biocontrol-entamopathogen	Smithson et al. (1995)
Metarhizium anisopliae	Subtilisin-like protease (*Pr1B*)	Biocontrol-entamopathogen	Joshi et al. (1997)
Metarhizium anisopliae	Mecca, zinc carboxypeptidase	Biocontrol-entamopathogen	Joshi & Leger (1999)
Metarhizium anisopliae	Subtilisins (Pr1A,B,C,D,E,F,J,H,I,J,K)	Biocontrol-entamopathogen	Bagga et al. (2004)
Metarhizium anisopliae	Toxic protease	Biocontrol-bioinsecticide	St. Leger et al. (1996)
Metarhizium anisopliae	Fungal chymotrypsin,	Evidence for lateral gene transfer	Screen & Leger (2000)
Metarhizium strains	Subtilisin-like protease isoforms (*Pr1*-A, B, C)	Biocontrol-entamopathogen	Bidochka & Melzer (2000)
Microsporum canis	Three subtilisin like proteases.	Animal pathogen	Descamps et al. (2002)
Monacrosporium megalosporum	Serine Protease Gene (*Spr1*)	Biocontrol-nematode trapping	Kanda et al. (Article in Press)
Monacrosporium microscaphoides	Extracellular protease	Biocontrol-nematode trapping	Wang et al. (2006)
Mucor miehei	Aspartate protease	Cheese industry	Gray et al. (1986)
Mucor miehei	Aspartyl protease	Dairy industry	Gray et al. (1986)
Mucor pusillus.	Mucor rennin, an aspartate protease	Cheese industry	Tonouchi et al. (1986)
Ophiostoma piliferum	Subtilase gene, albin1	Cosmetic defect	Hoffman & Breuil (2004b)
Paecilomyces lilacinus	Basic serine protease	Biocontrol-nematode trapping	Bonants et al. (1995)
Paracoccidioides brasiliensis	Serine protease, *kex* 2	Human pathogen	Venancio et al. (2002)
Penicillium chrysogenum	Alkaline serine protease	Human pathogen	Chou et al. (2002)
Penicillium chrysogenum	Pen ch 18, the vacuolar serine protease	Human pathogen	Shen et al. (2001)
Penicillium chrysogenum	Protease, *Epg222.*	Industry (Meat curing)	Benito et al. (2006)
Penicillium citrinum	Alkaline serine protease, *PenC 1*	Human pathogen	Su et al. (1999)
Penicillium marneffei	Extracellular Proteinases	Human pathogen	Moon et al. (2006)
Penicillium oxalicum	Pen o 18, the vacuolar serine protease	Human pathogen	Shen et al. (2001)
Penicillium roqueforti	Aspartyl protease, *aspA*	Dairy industry	Gente et al. (1997)
Phaffia rhodozyma	Aspartic protease	Industrial	Bang et al. (1999)
Pichia pastoris	Carboxypeptidase Y PRC1 gene	Recombinant expression host	Ohi et al. (1998)
Pleurotus ostreatus	Metalloprotease	Mushroom fruiting	Joha et al. (2004)

Contd.

Pleurotus ostreatus	Fungal subtilases, *posI* and *pcsI* gene	Regulation of lignolytic enzymes	Faraco et al. (2005)
Pneumocystis carinii	Subtilisin-like serine proteases	Human pathogen	Lugli et al. (1997)
Pneumocystis carinii	KEX1	Human pathogen	Lee et al. (2000)
Pneumocystis jiroveci.	Serine endoprotease Kex1	Human pathogen	Kutty & Kovacs (2003)
Pochonia chlamydosporia.	Protease VCP1	Biocontrol-nematode trapping	Morton et al. (2003)
Podospora anserina	Subtilisin-like serine protease PSPA	Vegetative incompatibility	Paoletti et al. (2001)
Pyrenopeziza brassicae	Psp2, serine protease	Plant pathogen	Keniry et al. (2002)
Rhizopus microsporus	Extracellular aspartic proteinase	Human pathogen	Schoen et al. (2002)
Rhizopus niveus.	Aspartic proteinase	Industrial	Horiuchi et al. (1988)
Saccharomyces cerevisiae	Subtilisin-like serine protease, KEX2	Protein processing	Mizuno et al. (1998)
Sapstaining fungi	Subtilase genes	Cosmetic defect	Hoffman & Breuil (2002)
Sapstaining fungi	Three subtilase genes	Cosmetic defect	Hoffman & Breuil (2004a)
Stagonospora nodorum	SNP1 trypsin protease	Plant pathogen	Bindschedle et al. (2003)
Trichoderma asperellum	Two aspartyl proteases	Biocontrol-mycoparasite	Viterbo et al. (2004)
Trichoderma harzianum	Trypsin like protease, prA1	Biocontrol-nematicidal	Suarez et al. (2004)
Trichoderma harzianum	Proteinase, prb1	Biocontrol-mycoparasite	Geremia et al. (1993)
Trichoderma harzianum	Aspartyl protease, papA	Industrial	Delgado-Jarana et al. (2002)
Trichoderma reesei	Trypsin-like fungal serine proteases	Heterologous protein production Host	Dienes et al. (In Press)
Trichoderma virens	Serine protease, tvsp1	Biocontrol-mycoparasite	Pozo et al. (2004)
Trichophyton rubrum	Secreted subtilisin genes (SUB1-7)	Human pathogen	Jousson et al. (2004)
Tritirarchium album	Proteinase K	Industrial	Gunkel & Gassen (1989)
Tritirarchium album	Proteinase R	Thermostable	Samal et al. (1990)
Tritirarchium album	Proteinase T,	Thermostable	Samal et al. (1989)
Verticillium dahliae	Trypsin protease	Plant pathogen	Dobinson et al. (1997)
Yarrowia lipolytica	Alkaline extracellular protease XPR2	Industrial protein processing	Davidow et al. (1987)
Yarrowia lipolytica	Kex2p like proteases	Proprotein/ recombinant fusion protein processing	Enderlin & Ogrydziak (1994)

Interestingly the most widely studied fungus as a pathogen and useful organism is the *Aspergillus* spp (Kothary et al. 1984, Kolattukudy 1985, Tatsumi et al. 1989, Kolattukudy et al. 1993, Kwon et al. 2001, Tunga et al. 2003). Filamentous fungi especially *Aspergillus niger* and *Aspergillus oryzae* produces a broad spectrum of proteases. In the case of *A. oryzae*, most of the reported studies on protease gene are aimed at over production of the enzyme.

Aspergillus oryzae is used for manufacturing soy sauce, and the alkaline protease (Alp) is considered to play a most important role in producing the delicious taste of soy sauce by hydrolyzing raw materials (Murakami et al. 1991). Alp is reported to be a serine protease with an alkaline pH optimum (Nakagawa 1970). In order to obtain some information about Alp at molecular level, Tatsumi et al. (1989) cloned the cDNA and deduced the primary structure of the mature region. In continuation to this they have reported the isolation and structural analysis of an entire cDNA, coding for prepro-Alp, and recombinant secretion of enzymatically active Alp from *Saccharomyces cerevisiae*. From the nucleotide sequence of the entire Alp-cDNA, they have deduced that Alp consists of a pre-pro-region of 121 aminoacids and a mature region of 282 aminoacids. A consensus sequence of signal peptide consisting of 21 amino acids is found at the N-terminus of the pre-pro-region. The primary structure of the mature region shared extensive homology (29 – 44%) with those of subtilisin family, and the three residues (Asp32, His 64 and Ser 221 in Subtilisin BPN) composing the active site were preserved. The entire cDNA, coding for pre-pro Alp, when introduced in to the yeast *S. cerevisiae*, directed the secretion of enzymatically active Alp into the culture medium, with its N-terminus and specific activity identical to that of native *Aspergillus* Alp.

Murakami et al. (1991) reported the complete nucleotide sequence of the alkaline protease (Alp) of the fungus *Aspergillus oryzae* isolated using synthetic oligonucleotides as hydridization probe. The Alp gene reported was 1374 nucleotides long and contained three introns, one of which was in the pro region and two in the mature coding region.

Sequences related to the TATA box (TATAAAT) and the CAAT box (CCAAAT) was found in the 5'-noncoding region. Furthermore, primer extension analysis showed that three transcriptional start points are present.

Later, Gomi et al. (1993) cloned the gene encoding an acid protease (PEPA) from *Aspergillus oryzae* using a 0.6kb fragment of the gene as a probe. This fragment was obtained by PCR using primers designed from the partial amino acid sequences of peptide fragments of the purified protein. The *pepA* gene encoded a 404 amino acid protein and contained 3 putative introns ranging in length from 50 to 61 nucleotides. The deduced amino acid sequence of the *A. oryzae* PEPA had a high degree of homology (67%) to the *A. awamori* PEPA. Southern hybridization analyses showed that the pepA gene existed as a single copy in the *A. oryzae* chromosome. They concluded the report with an interesting observation that the transformants of *A. oryzae* containing multiple copies of the *pepA* gene showed a 2-6 times increase in acid protease activity compared with the recipient strain.

An aspartyl protease gene from *Trichoderma harzianum* CECT 2413 was isolated and characterized (Delgado-Jarana et al. 2002). Based on conserved regions of other fungal acid proteases, primers were designed to amplify a probe that was used to isolate the *papa* gene from a genomic library of *T. harzianum*. *papA* was an intron less ORF which encoded a polypeptide of 404aa, including a prepropeptide at the N-terminal region. The promoter possessed potential AreA, PacC, and MYC sites for nitrogen, pH and mycoparasitism regulation respectively. Transformants from *T. harzianum* CECT 2413 cultivated in yeast extract supplemented medium over expressed *papA* and had a fourfold increase in protease activity compared to the wild type, while transformants that over expressed the β-1,6- glucanases gene *bgn* and *papA* had an additional 30% increase in β-1,6- glucanase activity compared to *bgn* single transformants.

Berka and colleagues (Berka et al. 1990) reported the cloning and disruption of the major extracellular aspartic protease, aspergillopepsin A from *Aspergillus awamori*. The genomic pepA sequence was isolated using a synthetic oligonucleotide probe. The sequence data of the *pepA* gene revealed that it contained three introns (51, 52, and 59 bp in length). Based on similarities to other aspartic proteinases, the deduced aminoacid sequence of the *pepA* was suggested to have a 20-aa signal peptide followed by a 49-aa propeptide that is rich in basic aa residues.

Katz et al. (1994) had reported a (*prtA*) gene encoding an alkaline protease (*Alp*) from *Aspergillus nidulans*. The gene was isolated from the *A. nidulans* library with a probe, a fragment of the Alp-encoding gene from an *Aspergillus oryzae*. The deduced amino acid sequence of the prtA product showed a high degree of similarity to proteases from *A. oryzae*, *A. fumigatus* and *A. flavus*. Levels of the *prtA* transcript were found to mirror extracellular protease levels (Katz et al. 1994).

Serine protease from the mold, *Tritirarchium album* has been extensively characterized. *T. album* probably secretes three related serine proteases where the production of proteolytic enzymes can be modulated by changing the source of organic nitrogen. Among the three, the best characterized one is proteinase K (Jany et al. 1986) and rest of the two designated as Proteinase T and Proteinase R are also characterized to some extent, while the genes for all of these have been cloned (Gunkel & Gassen 1989, Samal et al. 1989, Samal et al. 1990). Now, proteinase K is widely being used as a molecular biology reagent.

Initially the proteinase K gene from *T. album* was cloned and sequenced (Gunkel & Gassen 1989). The study resulted in the elucidation of the entire coding region and the 5' and 3' flanking regions of the gene. The deduced primary translation product of the gene encoded for a protein of 384 aa (Mw 40231 Da) out of which 105 aa from the N-terminal region was not present in the mature protein. By analogy to the evolutionary

related bacterial subtilisins and other serine proteinases it was inferred that the primary secreted product is a zymogen containing a 15 amino acid signal sequence and a 90 amino acid propeptide, and the propeptide is assumed to be removed in the later steps of the secretion process or upon secretion of the enzyme into the medium. The nucleotide sequence analysis of the gene and its flanking regions revealed that the proteinase K gene is composed of two exons and the single intron (63 bp) is located in the proregion. Furthermore, a putative promoter sequence and a capping site was identified, suggesting that the transcription start site is located 103 bp upstream of the ATG initiation codon. When attempted the expression of proproteinase K cDNA in *E. coli*, the recombinants tested exhibited slight proteolytic activity on skimmed milk plates suggesting that some fusion proteins were correctly secreted into the periplasm and processed to the mature proteinase K.

Later Proteinase T and R were reported from *T. album* (Samal et al. 1989, Samal et al. 1990). The genomic and cDNA clones encoding proteinase T were isolated (Samal et al. 1989) and observed that the coding sequence for this enzyme is interrupted by two introns. The deduced amino acid sequence was about 53% identical to that of proteinase K. Four cysteines were present in the mature proteinase, probably in the form of two disulphide bonds, which might explain the thermal stability of the proteinase. Furthermore, the protease showed considerable sequence similarity with thermitase (Meloun et al. 1985), aqualysin (Kwon et al. 1988), as well as with the alkaline extracellular protease of *Yarrowia lipolytica* (Davidow et al. 1987). They also attempted the recombinant expression of the proteinase T cDNA in *E. coli* and then the authenticity of the product was determined by western blot using a polyclonal antibody raised against the purified native proteinase T and N-terminal analysis of the recombinant product (Samal et al. 1989).

Later they identified a full length clone of proteinase R from the cDNA library (Samal et al. 1990). This clone contained sequences coding for the 108 amino acid prepro

leader as well as for the 279 aminoacid mature proteinase. Proteinase R belonged to the subtilisin group of serine proteases that contains disulphide bonds. Homology between proteinase R and proteinase K was found to be 87% at the nucleotide as well as at the amino acid level. There are 32 amino acid substitutions in proteinase R compared to that of mature proteinase K. Proteinase R and K were isolated from two different strains of *Tritirarchium album*. The amino terminal analysis of the purified proteinase K and R demonstrated the differences in two of the sixteen amino acid residues (Samal et al. 1990). The prepro sequences of proteinase R and K, although remarkably similar, contain a significant number of aminoacid substitutions and deletions. Mature proteinase K was generated by cleavage between aspargine and alanine residues (Gunkel & Gassen 1989), while the mature proteinase R was probably generated by a cleavage between phenylalanine and alanine residues. The Brookhaven Protein Data Base co-ordinate file of proteinase K was used as a template to study the proteinase R substitutions in three-dimensional space. The majority of the substitutions of proteinase R with respect to proteinase K were found to be on the exterior of the protein model.

2.6 Protease from marine microbes

The marine microorganisms are widely accepted as a potential source of alkaline protease, there are only few studies reported on the extraction of alkaline protease from marine microorganisms, and most of them are from marine bacteria. Some marine bacteria belonging to the genera *Vibrio*, *Pseudomonas* and *Bacillus* spp have been reported to produce protease with novel attributes like thermostability, salt tolerance and solvent stability (Makino et al. 1981, Farrel & Crosa 1991, Tsujibo et al. 1993, Tsujibo et al. 1996, Croocker et al. 1999, Salamanca et al. 2002, Estrade-Badillo et al. 2003, Teo et al. 2003, Kumar et al. 2004a, Sreeja Chellappan 2005, Barindra Sana et al. 2006, Venugopal & Saramma 2006). Barindra Sana et al. (2006) reported purification and characterization of a salt, solvent, detergent and bleach tolerant alkaline serine protease having a broad substrate specificity produced by a true marine bacterium *Gamma proteobacterium* isolated from marine environment of Sundarbans. Similarly Venugopal

& Saramma (2006) reported the extraction and characterization of an alkaline protease from *Vibrio fluvialis* isolated from the mangroove forests of Cochin estuary. Chi et al. (2006) screened 327 strains of marine yeast isolates for protease production, 12 strains showed enzyme production in casein agar plates out of which *Aureobasidium pullulan* showed maximum production.

Recently Damare et al. (2006) projected marine fungus as a potential source of alkaline protease. They have isolated 221 isolates of fungus from 5000 m depth in the Central Indian Basin (9-16° S and 73 – 76° E) and out of these isolates 33 % showed good protease production with maximum in *Aspergillus ustus*. The terrestrial counterparts of *Aspergillus* sp are well studied at biochemical and molecular level for the protease enzyme production (Tatsumi et al. 1989, Murakami et al. 1991, Jaton-Ogay et al. 1992, Frederick et al. 1993, Ramesh et al. 1994, Vankuyk et al. 2000).

Protease gene from several bacteria, fungi and viruses have been cloned and sequenced. There are few reports regarding the cloning of protease gene from marine bacteria, but marine fungal protease gene cloning is not reported till date. Tsujibo et al. (1993) reported the cloning and sequencing of an alkaline serine protease encoding gene *apr* II from a marine bacterium *Altermonas* sp. (Teo et al. 2003) isolated a protease gene, encoding a metalloprotease of a pathogenic strain of *Vibrio harveyi* isolated from the coastal waters of East Java, Indonesia. The recombinant enzyme was expressed in *Escherichi coli* TOP 10 cells and subjected for N-terminal sequencing and site-directed mutagenesis.

The molecular cloning of fungal gene is difficult compared to the bacterial gene cloning because of the size and complexity of the fungal genome. Even though the protease gene from marine fungus is not reported, there are reports explaining the sequencing of other genes from marine fungal isolates. Hou et al. (2006) isolated xylanase gene (*xyl*) from a marine fungus *Penicillium chrysogenum* FS010 isolated from

Yellow Sea sediments. The deduced aminoacid sequence of the protein encoded by *xyl* showed high homology with the sequence of glycoside hydrolase family. The gene was subcloned into an expression vector pGEX-4T-1 and the encoded protein was over expressed as fusion protein with glutathione-S-tranferase in *Escherichia coli* BL 21.

Kim and colleagues (Kim et al. 2003) has reported the cloning and expression analysis of the β-lactam genes from a marine fungi, *Kallichroma tethys*. They have identified the β -lactam bisynthesis genes *pcbAB* and *pcbC* from the cosmid genomic library made from the fungus. The phylogenic analysis indicated a close relationship with homologous genes of the cephalosporin-producing pyrenomycete, *Acremonium chrysogenum*. Expression analysis by reverse transcription–PCR suggested that both genes are highly regulated and are expressed in the late growth phase of *K. tethys* cultures.

A thorough survey of available literature on the public domain indicate that so far no work is reported on either alkaline serine protease by marine isolate of *Engyodontium album* other than our investigative group or information on molecular cloning of alkaline serine protease gene from marine fungus to the best of our knowledge.

Chapter 3

MATERIALS AND METHODS

3.1 Microorganisms

3.1.1 Fungus

The fungus used in this study, *Engyodontium album* (BTMFS 10) was obtained from the marine microbial culture collection of Microbial Technology Laboratory, Department of Biotechnology, Cochin University of Science and Technology. The strain was originally isolated from marine sediments of Cochin (Suresh & Chandrasekaran 1999) and was identified at MTCC, IMTECH, Chandigarh (Sreeja Chellappan 2005) and further confirmed by ribosomal RNA gene sequencing (Jasmin et al. 2006c, a, b). The culture was grown on Bennet's agar medium prepared in 50% seawater at 25°C and maintained on the same medium at 4°C and was subcultured once in thirty days. A set of stock culture was also maintained under mineral oil at room temperature.

3.1.2 Bacterial strains:

The following *E.coli* strains were used in the present study

DH5α: F$^-$λ$^-$, *recA1 endA1 hsdR17* (r_k^-, m_k^+) D (*lacZYA-argF*) *U169* (*ϕ80 lacZ* d*M15*) *supE44 thi-1 gyrA96 relA1.*

XL-1-Blue MRF' D (mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F'*proAB lacIqZ̃* DM15 Tn10 (Tetr)]

3.2 Plasmid and phages

The following plasmids and phages were used in the study

pUC18 (Bangalore Genei, India)

pMOS Blue (Amersham Pharmacia Biotech, Germany)

λZAPII (Stratagene, USA)

3.3 Media

The following media were prepared using distilled water and the composition of each per Litre is presented. All the media were sterilized and used.

Bennet's Agar (g/L)

Casein enzymatic hydrolysate	2.0
Beef extract	1.0
Yeast extract	1.0
Dextrose	10.0
pH	7.3 ± 0.2
Agar	20.0

Luria-Bertani medium (LB) (g/L)

NaCl	10.0
Tryptone	10.0
Yeast extract	5.0
pH	7.3 ± 0.2

LB Agar LB containing 1.5% (w/v) agar

NZY Agar (g/L)

NaCl	5.0
$MgSO_4$	2.0
Yeast extract	5.0
Casein hydrolysate	10.0
pH	7.3 ± 0.2
Agar	15.0

NZY Top agar NZY broth containing 0.7% (w/v) agarose

SOB Medium (g/L)

Bacto trypton	20.0
Bacto yeast extract	5.0
NaCl	0.5
KCl (250 mM)	10 ml
$MgCl_2$ (2 M)	5 ml added after sterilization

SOC Medium

SOC Medium is identical to SOB medium, except that it contains 20 mM glucose. After the SOB medium has been autoclaved, allow it to cool to 60°C or less and then add 20 ml of a sterile 1M solution of glucose.

Domingues Media (g/L)

KH_2PO_4	15.0
$(NH_4)_2 SO_4$	5.0
$MgSO_4. 7H_2O$	1.23
$FeSO_4. 7H2O$	0.0027
$MnSO_4.7H_2O$	0.0016
$ZnSO_4.7H_2O$	0.0014
$CoCl_2. 6H_2O$	0.0036
$CaCl_2. 2 H_2O$	0.8
NaCl	10.0
pH	7.3 ± 0.3

GPYS Medium (g/L)

Glucose	1.0
Peptone	0.5
Yeast extract	10.0
pH	7.3 ± 0.3
Aged Seawater	500ml
Distilled water	500 ml

3.4 Solutions and Buffers

All buffers and solutions were prepared in deionized water unless otherwise mentioned.

Ampicillin	100 mg/ml
Tetracyclin	5 mg/ml in 50% ethanol
$CaCl_2$	1 M
IPTG	0.5 M
X-gal	20 mg/ml in Dimethyl formamide
Chloroform: Isoamyl alcohol	24:1
EDTA (pH 8.0)	0.5 M
Ethidium bromide	10 mg/ml
Glucose	1 M
Maltose	20 %
Magnesium chloride	1 M
Magnesium sulphate	1 M
Phenol	Saturated with Tris-HCl (pH 8)
RNase stock solution	10 mg /ml
Proteinase K	20mg/ml
SDS	10%

Contd.

Sodium acetate (pH 5.2)	3 M
Sodium chloride	5 M
Sodium hydroxide	5 N
Sodium phosphate	1 M
Tris-HCl (pH 7.5 & 8.0)	1 M
DEPC treated water	Diethyl pyrocarbonate (DEPC) was added to deionized water to a final concentration of 0.1% and kept at 37°C over night with shaking and autoclaved
Tris-HCl (pH 8.0)	1 M in autoclaved DEPC treated water

20 x SSC

Sodium chloride	3 M
Sodium citrate	0.3 M
pH	7.3 ± 0.3

TAE (50X)

EDTA - 0.5 M (pH 8.0)	100 ml
Glacial Acetic acid	57.1 ml
Tris	242 g

TBE (5 X)

EDTA - 0.5 M (pH 8.0)	20 ml
Boric acid	27.5 g
Tris	54 g

TPE (1 X)

Tris-phosphate	0.09M
EDTA (pH 8.0)	0.002M

TE Buffer

EDTA (pH 8.0)	1 mM
Tris-HCl (pH 8.0)	10 mM

CTAB buffer (DNA Extraction buffer)

Tris-HCl (pH 8.0)	100 mM
NaCl	1.4 M
EDTA (pH 8.0)	20 mM
CTAB	2% (w/v)

SM Buffer

NaCl	5.8 g
$MgSO_4$	2 g
Tris-HCl (pH 7.5)	50mM
Gelatin	0.01 mg

DNA Loading Buffer (Type III)

Bromophenol blue	0.25%
Xylene cyanol FF	0.25%
Glycerol	30%

STE buffer

Tris-HCl (pH 8.0)	10 mM
NaCl	0.1 M
EDTA (pH 8.0)	1 mM

Solution I (GTE solution)

Glucose	50 mM
Tris-HCl (pH 8.0)	25 mM
EDTA (pH 8.0)	10 mM

Solution II (Lysis solution)

NaOH	0.2 N
SDS	1%

Solution III

Potassium acetate (pH 7.5) - 5 M	60 ml
Glacial CH_3-COOH	11.5 ml
Deionized water	28.5 ml
Final pH	5.2

Depurination solution

HCl	0.25 N

Denaturing solution

NaOH	0.5 M
NaCl	1.5 M

Neutralizing solution

Tris-HCl (pH 8.0)	0.5 M
NaCl	1.5 M
EDTA	0.001 M
pH	7.2

Washing solution:

Tris-HCl (pH 7.5)	0.2 M
SSC	2X

TBS

Tris-HCl (pH 7.5)	10 mM
NaCl	150 mM

3.5 Chemicals and Biochemicals

The chemicals and biochemicals used in the routine study were of either molecular biology grade or extra pure analytical grade procured from Qualigens, Himedia, SRL, Merck and Sigma. Developer and fixer for the processing of X-ray films were from Excell photochem Pvt. Ltd, India. Other than the commonly used chemicals and biochemicals, the following were also used.

5-Brom-4-Chlor-3-Indolyl-β-D-Galactopyranosid (X-Gal)	Sigma
TRI reagent	Sigma
Bovine serum albumin (BSA)	SRL
Disodium 3-(4-methoxyspiro (1,2-dioxetane-3, 2'-(5'-chloro) tricyclo [3.3.1.1 3,7] decan}-4-yl) phenyl phosphate (CSPD)	Roche diagnostics
Diethylpyrocarbonate (DEPC)	Sigma
Isopropyl- β-D-thiogalacto pyranoside (IPTG)	Sigma
Agarose	Qbiogene
Xylene cyanol FF	Qbiogene
λDNA *Eco*RI/ *Hind* III double digest marker	Bangalore genei
dNTP set	Bangalore genei
Cetyl trimetyl ammonium bromide (CTAB)	Bangalore genei
Ethidium bromide	Bangalore genei
500bp ladder	Bangalore genei
Ampicillin	Bangalore genei
Tetracycline	Bangalore genei

3.6 Enzymes and kits

3.6.1 Enzymes

T4 DNA Ligase	New England Biolabs
Taq DNA polymerase	Qbiogene & Bangalore genei
Proteinase K	Bangalore genei
Rnase A	Bangalore genei
Alkaline phosphatase	New England Biolabs
DNase I	Promega
Restriction endonucleases (Aat II, Bam HI, Bgl II, Eco R I, Hind III, Kpn I, Nco I, Nde I, Not I, Pst I, Sac II, Sal I, Sau 3 AI, Spe I, Sma I, Xba I, Xho I)	New England Biolabs & Bangalore genei

3.6.2 Kits

PCR Core System I	Promega
RT-PCR System	Promega
DIG High Prime DNA labeling and detection Starter Kit II	Roche Applied Science
ZAP-cDNA® synthesis, Gigapack® III gold cloning kit	Stratagene
PMOS Blue PCR Cloning Kit	Amersham Biosciences
PolyAtract® mRNA isolation system	Promega
Wizard PCR Prep DNA purification kit	Promega

3.7 Culturing of *Engyodontium album* BTMFS10

Initially fungal inoculum was developed using the stock agar slope culture of *E. album*. The fungus was inoculated onto Bennet's agar slants and allowed to grow for 14 days at 25°C. Later five ml sterile saline (0.85% NaCl) containing 0.1% Tween 80 was added to the slants and the conidia were detached from the mycelium by gentle scraping with a sterile loop. The conidial spore suspension thus obtained was transferred to a sterile tube and inoculated into GPYS medium supplemented with 1% casein at a

concentration of 2% (v/v). The culture was incubated at 25°C for 4-7 days under static condition and the mycelia mat obtained was used for DNA isolation. For extracting the total RNA, the fungus was cultured in Domingues Media supplemented with 1% casein following the same condition.

3.8 Isolation of Nucleic acids

3.8.1 Chromosomal DNA Isolation from *Engyodontium album* BTMFS10 (Rogers & Bendich 1994)

The mycelia mat of *E. album* was harvested from GPYS medium, washed with sterile distilled water and blot dry. Approximately 1 g of mycelia were frozen in liquid nitrogen and disrupted into fine powder using a sterile mortar and pestle. This powder was transferred into a 50 ml capacity centrifuge tube containing 16 ml of pre heated (65°C) CTAB buffer and incubated at 65°C for 30 minutes in a water bath followed by incubation at room temperature. Equal volume of chloroform: isoamyl alcohol (24:1) was added to the tubes and mixed thoroughly to form an emulsion. Subsequently it was centrifuged at 10,000rpm for 10 minutes and the aqueous layer was saved into a new centrifuge tube using cut tip. The chloroform: isoamyl alcohol extraction was repeated twice and the aqueous phase was collected carefully. Sodium acetate (pH 5.2) was added to the aqueous phase at a final concentration of 300mM and DNA was precipitated by the addition of 2/3 volumes of ice-cold isopropanol. The DNA strands were spooled out gently with the aid of a glass rod, washed with 70% ethanol, air dried briefly, and then dissolved in Milli Q water (Millipore, USA).

3.8.2 Total RNA Isolation by TRI REAGENT ™

The RNA from fungal biomass was isolated following the method of (Chomcynski & Sacchi 1987). Approximately 100 mg of fungal mycelium harvested from *E. album* culture was frozen in liquid nitrogen, ground to a fine powder in a pre-chilled nuclease free mortar and pestle. The tissue powder was transferred quickly to a nuclease free micro centrifuge tube and one ml of TRI reagent™ was added. The content

of the tube was mixed well and was kept at room temperature for five minutes. The sample was centrifuged at 12,000g for 10 minutes at 4°C. The supernatant was transferred to an RNase free micro centrifuge tube and 200µL of chloroform was added per one ml Trizol^{TM} reagent used. It was mixed well to form an emulsion and incubated at room temperature for 15 minutes and centrifuged at 12,000g for 15 minutes at 4°C. The aqueous phase was collected in a fresh micro centrifuge tube and 500µL isopropanol was added and maintained at room temperature for 10 minutes to precipitate the total RNA. The sample was centrifuged at 12,000g for 10 minutes at 4°C and the pellet was washed with 75% ethanol by centrifugation at 7,500g for 5 minutes at 4°C. After centrifugation, ethanol was decanted, RNA pellet dried at room temperature, dissolved in 50µL of nuclease free water and stored at -70°C.

3.8.3 Isolation of mRNA from total RNA

mRNA was separated from total RNA using Poly A Tract mRNA isolation system III from Promega according to the manufacturer's protocol, as described below.

3.8.3.1 Annealing of the oligo dT probe

Total RNA was dissolved in 500µL of nuclease free water and incubated at 65°C for 10 minutes. To the denatured total RNA, 3µL of oligo dT probe and 13µL of SSC (20 X) were added and incubated at room temperature for 30 minutes to allow the slow cooling of the mix to room temperature.

3.8.3.2 Washing of Streptavidin-Paramagnetic particles (SA-PMPs)

The Streptavidin-Paramagnetic particles (SA-PMP) provided in the kit in aliquots were resuspended by gently flicking the bottom of the tube until they were completely dispersed. To capture the SA-PMPs, the tube was placed in a magnetic stand provided in the kit until the SA-PMPs were collected on the side of the tube. The supernatant was carefully removed and SA-PMPs were washed thrice with 0.5X SSC (300µL/ wash),

each time capturing them using the magnetic stand and carefully removing the supernatant. Washed SA-PMPs were resuspended in 100µL of 0.5X SSC.

3.8.3.3 Capture and washing of annealed oligo (dT)–mRNA hybrids

The entire content of the annealing reaction was added to the tube containing washed SA-PMPs and mixed well by gentle pipetting. The mixture was incubated at room temperature for 10 minutes with intermittent mixing by inversion. The SA-PMPs were captured using magnetic stand and the supernatant was carefully removed without disturbing the SA-PMP pellet. The particles were washed four times with 0.1X SSC (300µL/ wash) by gently flicking the bottom of the tube until all particles were resuspended. After the final wash the aqueous phase was removed by pipetting as much as possible without disturbing the pellet.

3.8.3.4 Elution of mRNA

The final SA-PMP pellet was resuspended in 100µL of nuclease free water to elute mRNA. SA-PMPs were captured magnetically and the aqueous phase with mRNA was collected in a sterile RNase free micro centrifuge tube. The elution was repeated by adding additional 50µL of nuclease free water. Final elution was carried out with the pooled 150µL mRNA solution. Eluted mRNA was precipitated by adding 0.1 volume of 3M sodium acetate (pH 5.2) and one volume of isopropanol, incubated overnight at -20°C, washed in 75% ethanol, air dried and dissolved in 30µL of nuclease free water. An aliquot of the sample was quantified spectrophotometrically and stored at –70°C.

3.8.4 Isolation of plasmid DNA

3.8.4.1 Small-scale preparation of plasmid DNA (Sambrook et al. 1989)

Overnight culture (3ml) was centrifuged at 12,000rpm for a minute and the pellet was resuspended in 100µl of double distilled water. One hundred µl lysis buffer (1% SDS, 10mM EDTA, 0.1N NaOH) was added and maintained in a boiling water-bath for

2 minutes. Subsequently 50µl of 1.0M $MgCl_2$ was added and mixed by tapping. It was kept on ice for 2 minutes and spun at 12,000rpm for 2 minutes. The supernatant was transferred to another tube and 50µl of 3M potassium acetate was added, mixed by tapping, kept on ice for 2 minutes and spun at 12,000rpm for 2 minutes. Without disturbing the pellet the supernatant was transferred into a fresh tube. Six hundred µl of isopropanol was added, kept on ice for 5 minutes and spun at 12,000rpm for 2 minutes. Finally the pellet was washed with 70% ethanol, dried and dissolved in TE buffer (pH 8.0) or distilled water.

3.8.4.2 Large scale preparation of plasmid DNA

Plasmid DNA was isolated by a modified version of alkali lysis method (Sambrook et al. 1989). The cell pellet generated from a 50ml LB broth was washed with 20ml STE buffer and centrifuged at 5000rpm for 5 minutes at room temperature. The pellet was again resuspended in 8ml of ice-cold solution I and approximately 2mg/ml lysozyme was added to break the bacterial cell walls. To the dispersed cells, 16ml of Solution II was added, mixed well by inversion and incubated on ice for 10minutes.This was followed by the addition of 12 ml of Solution III, mixed by gentle inversion and then incubated on ice for 15minutes. Subsequently, bacterial cell debris was precipitated by centrifugation of the sample at 10,000rpm for 10 minutes at 4°C. The supernatant was transferred to a new centrifuge tube to which 0.6 volume of isopropanol was added, followed by incubation at room temperature for 10 minutes. The plasmid DNA was pelleted by centrifugation at 10,000rpm for 10 minutes at room temperature. Finally the plasmid DNA was washed twice with 70% ethanol, air dried and dissolved in 2ml sterile Milli Q water. The DNA sample was stored at -20°C.

3.8.5 Precipitation of DNA

DNA in solution was precipitated by the addition of 0.1 volume of 3M sodium acetate (pH 5.2) and 0.6 volume of isopropanol or 3 volumes of absolute ethanol. High

molecular weight DNA was precipitated at room temperature while DNA with low molecular weight precipitated on incubation at –20°C for 12 hours. The precipitated DNA was separated by centrifugation at 10,000 rpm for 15 minutes at room temperature. The excess salt was removed by washing the DNA pellet with 70% ethanol. The DNA was dried under vacuum and dissolved in appropriate volume of TE buffer (pH 8.0) or Milli Q water.

3.8.6 Nucleic acid quantification

The quantitative analysis of DNA, RNA or oligonucleotides was done spectrophotometrically by measuring the optical density at 260 (A_{260}) and 280nm (A_{280}). The reading at 260nm allows calculation of the concentration of the nucleic acid in the sample. An OD of 1 corresponds to approximately 50μg/ml for double-stranded DNA; 40μg/ml for single stranded DNA or RNA; and 20μg/ml for single stranded oligonucleotides. The ratio between readings at 260nm and 280nm (A_{260}/A_{280}) provides an estimate of the purity of the nucleic acid. Pure preparations of DNA and RNA have ratio of 1.8 and 2.0, respectively. If there is contamination with protein, the ratio (A_{260}/A_{280}) will be significantly less than the values given above.

3.8.7 Agarose gel electrophoresis

3.8.7.1 DNA gel electrophoresis

Agarose gel (0.8-1%) was prepared with electrophoresis grade agarose and known volume of TAE buffer (1X). The contents were mixed and melted in a microwave oven. The molten agarose at 55°C was poured without air bubbles into the gel-casting tray and allowed to solidify at room temperature for about 60 minutes.

DNA sample was mixed with 6X concentration gel loading buffer to a final concentration of 1X and loaded into the wells as 10μL aliquots. The electrophoresis tank was attached to a DC power pack and resolved at 8 V/cm. The gel was removed when the

bromophenol blue in the tracking dye reached 2/3 of the gel and the DNA in the gel was observed on a UV transilluminator after staining with ethidium bromide solution.

3.8.7.2 RNA gel electrophoresis

The molten agarose (1%) at 55°C, mixed with appropriate amount of iodoacetic acid (1.86μg/ml) and ethidium bromide was poured without air bubbles into the gel-casting tray and allowed to solidify at room temperature for about 60 minutes. RNA sample was mixed with 6X concentration gel loading buffer (prepared in DEPC treated water) to a final concentration of 1X and loaded into the wells as 10μL aliquots. The electrophoresis tank was attached to a DC power pack and resolved at 5V/cm. The gel was removed when the bromophenol blue in the tracking dye reached half of the gel and RNA in the gel was observed on a UV transilluminator.

3.8.8 Elution of DNA from agarose gel

Elution of the DNA from agarose gel was usually done with the elution apparatus (Exelutor, FinnPCR). When current passes between the electrodes, the negatively charged DNA get eluted out from the gel piece in to the small volume of (500μl) TAE buffer (1X) which cover the gel piece as a bubble. DNA was precipitated from this TAE buffer (1X), by adding 0.1 volume of 3M sodium acetate (pH 5.2) and one volume of isopropanol, incubated overnight at -20°C. Next day, the sample was centrifuged, the pellet of DNA washed with 70% ethanol and dissolved in Milli Q.

For elution of PCR products from the agarose gel, Wizard PCR Prep DNA purification kit (Promega, USA) was used. In this method, the agarose gel piece was dissolved in one ml of resin provided with the kit. The solution was transferred to the syringe barrel spin-column-vacuum manifold assembly and vacuum was applied to draw the solution through the assembly. After breaking the vacuum, 2ml of 80% isopropanol was added to the spin-column through the syringe barrel and vacuum was continued.

Subsequently the assembly was dismantled and the spin-column with the DNA adhered on its membrane was fitted on to a clean micro centrifuge tube. The tube was spun at 10,000g for 2 minutes to remove any residual alcohol and the attached micro centrifuge tube was discarded. After attaching a fresh tube with the mini column, 50µl of distilled water was added on to the spin column, incubated for 2-5minutes and spun at 10,000g for 20 seconds. Later the spin column was discarded and the eluted DNA was precipitated and dissolved in a minimum quantity of distilled water.

3.8.9 Primer Designing

The following primers were designed and used for PCR based experiments. All the primers used in the study were synthesized either by Bioserve biotechnologies, India.

3.8.9.1 Degenerate primers

Degenerate primers were designed based on the conserved domain deduced from the already reported nucleotide sequence encoding alkaline protease of several fungi. The nucleic acid sequences coding for alkaline serine protease were retrieved from the GenBank (http://www.ncbi.nlm.nih.gov/) and aligned by CLUSTAL W (Thomson et al. 1994) alignment programme of Bio-Edit software (Hall 1999). The bases at the most conserved regions of the alignment were considered for the primer designing.

Degenerate primers designed for amplifying partial gene are given below

ASP1R- 5' GCCATSGARGTRCCMGAGATGG 3'
ASP 1F - 5' AAGTACATYGTCAAGYTCAAGGA 3'

ASP 2R -5' AAGGCACCACCMAGAGACATGTT 3'
ASP 2F -5' TACATTGTYGTCATGAAYGATGG 3'

ASP 3R - 5' GTGGCCATGGAGGTACCGGAGA 3'
ASP 3F - 5' GTTGCTGCCGGTAACGACAAC 3'

3.8.9.2 UTR based primers

A set of primers were designed from the 5' and 3' untranslated regions of the gene (obtained from the partial genomic DNA library) using GeneTool software. This primer pair was designated as Eap UTR primers and used to amplify the coding region of the gene (ORF) from *E. album.*

Eap UTR F: 5' TCATCAACAGCCATCGCAGCAATAC 3'
Eap UTR R: 5' GACTAAATATGGTCGTAAGACCGATATGAATG 3'

Primers were diluted to 100 pmol working stock solution using sterile Milli Q water and stored at -20°C. For PCR reactions, primers were diluted from the working stock to a concentration of 10 pmol using nuclease free water and used.

3.9 Polymerase chain reaction (PCR)

PCR reactions were performed in Peltier thermal cyclers (MJ Research, USA and Eppendorf, Germany) in 200µl capacity thin walled tube in a final volume of 25µl. Taq DNA polymerase was used for amplification reactions. Genomic DNA (100ng) or 1µl of first strand-synthesized product from the RT-reaction (~50ng) was used as template in PCR reactions. 'Hot start' PCR was done for amplification in which the reaction mix was heated to 95°C for five minutes, quick chilled on ice followed by the addition of 1unit of Taq DNA polymerase. A typical PCR reaction comprises of denaturation at 95°C for 25 seconds, primer annealing at 55-65°C for 50 seconds and extension at 68-72°C for 60-120 seconds, depending on the target to be amplified. The reaction cycle was repeated 35 times and the product was analyzed on a 1% agarose gel.

3.9.1 Reverse transcription-polymerase chain reaction (RT-PCR)

Before reverse transcription, the RNA sample was given an RNase free DNase I (Promega, USA) treatment at 37°C for 30 minutes to remove DNA contamination.

Subsequently the DNase I was heat inactivated at 70°C for 30 minutes. RT reaction was done with ~1μg of RNA as template and 0.5μg of oligo dT as primer for first strand synthesis as described by the manufacturer (Promega, USA). The reaction was performed in 1X RT buffer, in the presence of 1mM dNTPs, 5mM $MgCl_2$, 20 units of RNase inhibitor and 20units of AMV reverse transcriptase to a final volume of 20μl. The reaction was performed at 42°C for 1hour. AMV reverse transcriptase was heat inactivated at 99°C for 5 minutes. Samples were stored at 4°C until use. For RT-PCR, 1μl of the RT product (from a total of 20μl) as template, and specific primer pairs were used. Amplified product was checked by agarose gel electrophoresis.

3.9.2 Colony PCR

Each colony was grown in LB broth separately and one millilitre each from five cultures were pooled together, extracted plasmid and used as template for PCR reaction. The 25μL reaction mix contained 1X PCR buffer, 200μM dNTP mix, 20 pmol of ASP1 forward and reverse primers and 1 unit of Taq DNA polymerase. The PCR conditions were set to 94°C, 5 minutes of initial denaturation, 35 cycles of 94°C for 1 minute, 50°C for 1 minute, and extension at 72°C for 2 minutes, followed by final extension at 72°C for 7 minutes. Checking the PCR products on a 1.2% agarose gel identified the positive groups. Plasmids were extracted from individual colonies of the positive groups and amplified using the same set of primer for selecting the right clone with the protease gene.

3.10 DNA Manipulations and cloning

3.10.1 Restriction enzyme digestion of DNA

Restriction digestion of plasmid DNA samples was carried out with one unit of restriction enzyme per μg of DNA. For digesting genomic DNA, upto 4.0 units of restriction enzyme per μg of DNA were used. The buffers and incubation conditions for digestion were the same as given by the manufacturer of respective restriction

endonuclease enzymes. The restriction reaction was stopped by incubating at 65°C for 15 minutes and chloroform: isoamyl alcohol extraction also was performed whenever required.

3.10.2 Dephosphorylation of vector DNA

The plasmid DNA linearized with appropriate restriction enzyme was dephosphorylated using calf intestinal phosphatase (CIP). To 1.2μg of Linearised vector, 2 μl of 10X CIP dephosphorylation buffer (10mM $ZnCl_2$, 10mM $MgCl_2$ and 100mM Tris-HCl pH 8.3), and 0.1 unit of CIP were added and incubated at 37°C for 30 minutes in a final reaction volume of 20μl. After the reaction, CIP was inactivated by heating at 75°C for 10 minutes in the presence of 5.0mM EDTA. Further purification was done by phenol: chloroform (1:1) extraction. Sodium acetate (pH 5.2) was added to a final concentration of 300mM followed by the addition of 3 volumes of absolute ethanol to precipitate the DNA. The mixture was spun at 8,000g for 10minutes at room temperature. The precipitate was washed with 70% ethanol and dissolved in TE buffer (pH 8.0).

3.10.3 Ligation of DNA into vector

Dephosphorylated vector DNA (50ng) was used in all ligation reactions. The concentrations of the target DNA to be ligated varied according to its size. The vector: insert molar ratio was kept 1:3 in a 10μl ligation reaction. PCR products were made blunt (in case of amplification using *Taq* DNA polymerase) using T4 DNA polymerase and phosphorylation of the 5' end of the insert DNA was performed using polynucleotide kinase. PCR products obtained from reactions involving Pfu DNA polymerase were given only the phosphorylation reaction prior to ligation, as the products were blunt. All ligation reactions were performed in 10μl reaction with 1unit of T4 DNA ligase enzyme at 16°C or 22°C.

3.10.4 Preparation of competent cells (Sambrook et al. 1989)

A single colony of *E. coli* host cell (DH5α) was inoculated in 5ml of Luria-Bertani (LB) broth and grown overnight with constant shaking at 200 rpm at 37°C. Five hundred µL of overnight culture was inoculated to 50ml of LB broth and incubated overnight at 37°C in an incubator shaker at 200rpm. The cells were harvested by centrifuging at 4,000rpm for 5 minutes at 4°C, the pellet was resuspended in 20ml of ice cold 0.1M $CaCl_2$ and incubated on ice for 30 minutes. Subsequently the cells were segregated from the solution by centrifugation at 7,000rpm 4°C, 10 minutes and the pellet was resuspended in 2ml of ice cold 0.1M $CaCl_2$ and kept in ice. This was aliquoted with glycerol and stored at -80°C until use.

3.10.5 Transformation

Ten µl of plasmid DNA (10-20ng) was added to the prepared competent cells (90µl) and incubated in ice for 30 minutes. Later the cells were given a heat shock at 42°C for 90 seconds by immersing the tube in a water bath followed by quick chilling in ice for 10 minutes. Afterwards, 600 µl of SOC medium was added to the transferred cells in the tube and incubated at 37°C for 1 hour. The grown cells (50-100µl) were plated on LB agar plates (LB media containing 1.5% bactoagar) containing 50µg/ml of ampicillin, 100µg/ml IPTG and 40µg/ml X-gal employing spread plate technique and incubated at 37°C. The plates were checked for transformants after overnight incubation.

3.11 Probes

Both nucleic acid and antibody probes were used for screening cDNA and genomic libraries. Details are as given below.

Nucleotide Probes

Heterologous probe	The partial cDNA sequence of alkaline serine protease from the fungus *Aspergillus fumigatus*- Generously gifted by Dr. P.E. Kolattukudy, University of Central Florida, USA.
Homologous probe	Partial genomic DNA sequence of alkaline serine protease from the fungus *Engyodontium album*.
Antibody	Alkaline protease polyclonal antibody raised in rabbit against *Engyodontium album* alkaline protease (Sreeja Chellappan 2005).

3.12 DNA labeling

3.12.1 Multiprime labeling of the radioactive probe

The multiprime labeling of DNA was done as described by (Feinberg & Vogelstein 1983). Double stranded DNA (50-200 ng) was denatured in a boiling water bath for 10 minutes followed by chilling on ice. The reaction was performed as per the directions of the multiprime labeling kit manufacturer (Jonaki, BARC, India). Random hexanucleotides (75ng), 5.0µl of 10x reaction buffer (900mM HEPES adjusted to pH 6.6 with NaOH, 10mM $MgCl_2$ and 40mM DTT), 4mM each of unlabeled dNTPs, and 4µl of α^{32}P-dATP (40 µCi; specific activity 3000Ci/mmol, Jonaki, BARC, India) and 2.0 units of Klenow (*E. coli* DNA polymerase I - large fragment) were added to the DNA in a total volume of 50µl. The samples were incubated at 37°C for one hour. Adding EDTA to a final concentration of 10mM terminated the reaction. Sheared and denaturated *E. coli* DNA (10µg) was added as carrier and the DNA precipitated at 0°C overnight. Subsequently, the probe was filtered through a G-50 spin column (Amersham Pharmacia Biotech, USA) to remove the unincorporated radioisotope.

The radio labeled probe (1.0µl) was spotted on a piece of nylon membrane, dried and washed in ice cold 10% trichloroacetic acid for 10 minutes, and rinsed in absolute ethanol at room temperature for 10 minutes. The filter was air dried, immersed in scintillation fluid (0.5% PPO, 0.03% POPOP in toluene) and the percentage of incorporation of radioactivity determined by counting in a Packard scintillation counter. Usually nick translation results in probes with a specific activity of $2.5x10^7$ cpm/µg while the probes prepared by random priming are of ten-fold higher specific activity.

3.12.2 DIG DNA labeling (Non radioactive)

Digoxigenin-labelled DNA probe was generated using DIG-High Prime labeling kit (Roche Applied Sciences, Germany) according to the random primed labeling technique. The reaction was performed as per the directions of the manufacturer. Double stranded DNA (400-600ng) was mixed with sterile double distilled water to a final volume of 16µl in a reaction vial and denatured in a boiling water bath for 10 minutes followed by chilling on ice. Four µl of the DIG-High Prime (the 5X concentrated labeling mixture contain random hexamers, dNTP mix containing alkali-labile Digoxigenin-11-dUTP, Klenow enzyme and an optimized reaction buffer) was added to the DNA and incubated overnight at 37°C. The reaction was stopped by adding 2µl of 0.2M EDTA (pH 8.0) and heating to 65°C for 10 minutes.

The expected yield of the DIG-labeled DNA in the probe was estimated using the chart provided along with the kit. The yield depends on the starting amount of the template and incubation time. The efficiency of the labeled reaction was evaluated by comparing the different dilutions of the labeled probe with that of the DIG-labeled control DNA, by dot blot analysis. For this purpose, both the samples were first diluted to 1ng/µl according to the expected yield to start the dilution series (10, 3, 1, 0.3, 0.1, 0.03, 0.01 pg/µl).

The diluted samples were spotted on a strip of positively charged nylon membrane (1µl each) in two rows, and the DNA fixed on the membrane by baking at 120°C for 30 minutes. The spots were detected by chemiluminescent method, compared with the control and the amount of DIG labeled DNA was calculated.

3.13 Construction of phage cDNA library of *E.album* BTMFS10

Phage cDNA library of *E. album* was constructed following the instruction manual of the Lambda ZAP® II /*Eco R* I vector kit (Stratagene) as described below:

3.13.1 Synthesizing first strand cDNA

First strand cDNA was synthesized from mRNA by reverse transcription using StrataScript RT. The reaction mix was prepared for 50µl volume with 5µg of mRNA, 1 X first strand buffer, 75 units of StrataScript RT, 3µl of the first strand methyl nucleotide mix (10mM stock), 40 units of RNase block, Ribonulcease inhibitor and 2.8 µg oligo dT18 linker primer. mRNA, oligo dT18 primer, first strand buffer, nucleotide mix, RNase inhibitor and nuclease free water were added to a nuclease free micro centrifuge tube, mixed gently and incubated at room temperature for 10 minutes. To this, 75 units of StrataScript RT enzyme was added, mixed gently and spun down. RT reaction was incubated at 42°C for 1 hr and then kept on ice.

3.13.2 Second-strand synthesis reaction

To the first strand synthesis reaction, 1X second strand buffer, 6µl second strand dNTP mix (10mM stock), DEPC water, 3U of RNase H and 11µl of DNA polymerase I (9 U/µl) were added and incubated at 16°C on a water bath for 2.5 hrs. After incubation the reaction tube was, immediately placed on ice.

3.13.3 Blunting the cDNA termini

To the second strand reaction tube, 23µl of blunting dNTP mix (2.5mM stock) and 5 units of the enzyme Pfu DNA polymerase were added and incubated for exact 30 minutes at 72°C. After incubation, the sample was extracted with 200µl Phenol: chloroform (1:1) mixture, continued with equal volume of chloroform: isoamyl alcohol (1:1) mixture. The cDNA was precipitated from the supernatant recovered by adding 20µl sodium acetate (3M) and 400µl of absolute alcohol and incubating overnight at -20°C. The precipitated cDNA was collected by centrifugation at 12000rpm for 60 minutes at 4°C and the pellet was gently washed with 500µl of 70% ethanol and dried briefly. The pellet was resuspended in 9µl of Eco RI adapters (0.4µg/µl) and incubated at 4°C for 30 minutes.

3.13.4 Ligating the Eco RI Adapters

To the tube containing blunt ended cDNA resuspended in adapter solution, 1X ligase buffer, 1µl of ATP (10mM) and 4 units of T4 DNA Ligase were added and incubated at 4°C for 2 days.

3.13.5 Phosphorylating the Eco RI ends

After heat inactivating the ligase (at 70°C for 30 minutes) and cooling it down to room temperature, the following reagents were added to phosphorylate the adapter ends of cDNA: 1X ligase buffer, 2µl of ATP (10mM), 5µl sterile water, and 10 units of T4 Polynucleotide Kinase. The reaction mixture was incubated for 30 minutes at 37°C followed by heat inactivation of the enzyme at 70°C for 30 minutes.

3.13.6 Size fractionation of cDNA

The cDNA was precipitated with absolute alcohol and resuspended in 14µl of STE buffer (1X). The sample mixed with 3.5µl of column loading dye was loaded on the drip column containing Sepharose CL-2B gel filtration medium. STE buffer (1X) was

used as the mobile phase, and the eluted fractions were collected separately in micro centrifuge tubes. An aliquot from each fraction was electrophoresed on 1.2% agarose gel and the fractions containing cDNA of size >400bp were pooled, extracted with equal volume of Phenol: chloroform (1:1) and then with equal volume of chloroform: isoamylalcohol (1:1). From the supernatant collected, the cDNA was precipitated with two volume of absolute alcohol by incubating overnight at -20°C. The precipitated cDNA was collected by centrifugation at 12,000rpm for 60 minutes at 4°C. The pellet was carefully washed with 200µl of 80% ethanol and dried briefly under vacuum. The cDNA was resuspended in 3.5µl of sterile distilled water.

3.13.7 Ligation of cDNA into the vector

Approximately 100ng of cDNA was ligated to 1µg of *Eco*R I predigested λZAP II vector in a total volume of 5 µl and the ligation mixture was incubated at 4°C for 2 days.

3.13.8 *In vitro* packaging of lambda phage

Three µl of the ligation reaction was mixed with 25µl Gigapack® III gold packaging extract (Stratagene, USA) and the mixture was incubated at 22°C for 2hrs. To this, 500µl of SM buffer and 20µl of chloroform were added, mixed gently and spun briefly. The supernatant was transferred to a fresh tube and stored at 4°C until use.

3.13.9 Preparation of host bacteria

The glycerol stock of XL1-Blue MRF' was streaked on to the LB plate and incubated overnight at 37°C. 10ml of LB media supplemented with 10mM $MgSO_4$ and maltose (0.2% w/v) was inoculated with a single colony of XL1-Blue MRF'. It was grown at 37°C, with shaking for 4.0-6.0 hours (it was not grown past OD(A_{600}) = 1.0). The cells were pelleted at 2,000rpm for 10 minutes and gently resuspended in half the original volume with sterile 10mM $MgSO_4$ and further diluted to OD(A_{600}) = 0.5 with sterile 10mM $MgSO_4$.

3.13.10 Titering the packaging reaction

Consecutive dilution of packaged phages was made in SM Buffer. One µl of each dilution was added to 200µl of the diluted host cells and the attachment of the phages to the host cells was achieved by incubation of the mixture at 37°C for 60 minutes. Then 5ml of pre-warmed NZY top agar was added and the mixture was plated on NZY agar plates. Plaques were visible after 8-12 hours of incubation at 37°C. Phage plaques were counted and the plaque-forming unit per ml (Pfu/ml) concentration of the library was determined as:

$$\frac{\text{Number of plaques x dilution factor} \quad \text{x } 1000\mu\text{l /ml}}{\text{Volume plated } (\mu\text{l})}$$

3.13.11 Library amplification

A single colony of XL1-Blue MRF' host cells was picked and inoculated in 20ml LB supplemented with 0.2% maltose and 10mM $MgSO_4$ and grown overnight at 37°C. One ml of the culture was inoculated in 50ml of fresh, pre-warmed medium and allowed to grow for 4-6 hours under vigorous agitation until the OD (A_{600}) reached to around 0.7. The cells were harvested and resuspended to attain an OD of 0.5 (A_{600}) in 10mM $MgSO_4$. Aliquots of the packaged library suspension were mixed and incubated with 200µl of the diluted host cells for 60 minutes at 37°C in sterile tubes. About 5ml of melted top agar (48°C) was poured into the tube, mixed well, and spread evenly onto a freshly poured 110mm plate of NZY bottom agar. The plates were incubated at 37°C for 6-8 hours. The plaques were allowed to grow about 1-2mm. Twenty such plates were prepared for a library having 1.0×10^5 plaques. Thereafter 3ml of SM buffer was added to the plates and the phages were allowed to diffuse into the buffer by overnight incubation at 4°C. The phage suspensions were harvested from the plates, extracted with chloroform (5% v/v), the clear supernatant was added with 0.3% chloroform and kept at 4°C. Later, the library was aliquoted and stored with 7% (v/v) DMSO at -80°C. The titer of the amplified library

was checked using appropriate amount of host cells and serial dilutions of the library. The final titer of the library was found to be 10^9 Pfu/ml.

3.14 Plaque hybridization screening

3.14.1 Screening with heterologous cDNA probe

Probe labeling and hybridization of the membranes were done with DIG High Prime DNA labeling and detection Starter Kit II (Roche, Germany) according to the manufacturer's protocol, briefly described below.

For plaque blotting, positively charged nylon membranes of appropriate size were placed on the surface of the pre-chilled agarose plates containing plaques, removed after one minute and dried briefly by placing, plaques side up, on sterile filter paper. Thereafter the membrane was treated on a piece of filter paper soaked in denaturing solution for 5 minutes and subsequently transferred for 5 minutes, plaques side up, to filter paper soaked in neutralizing solution. The filters were rinsed for 10 minutes on filter paper soaked in 2X SSC solution and then dried briefly on Whatman 3mm paper. To fix the nucleic acids to the filters, the blots were baked at 80°C for 2 hrs. Filters were pre-hybridized in hybridization buffer for 1 hour at 52°C. Digoxigenin-labelled DNA probe was generated by random priming using a DIG High Prime enzyme system. Partial cDNA sequence of *Aspergillus fumigatus* alkaline serine protease was used as the heterologous probe. After denaturation the labeled probe was added to the hybridization buffer. Hybridization was carried out overnight at 52°C.

Post hybridization washes included a low stringency wash in which the blots were washed twice in 2.0X SSC + 0.1% SDS at room temperature for 5 minutes each and a high stringency wash, where the blots were washed twice in 0.5X SSC + 0.1% SDS at 65°C for 15 minutes each. After washing, the blots were rinsed briefly in 1X washing buffer and then incubated in blocking solution (1X) for 30 minutes at room temperature. The blots were then transferred to the antibody solution (anti-Digoxigenin antibody,

diluted 1/10,000 times in blocking solution) and incubated at room temperature for 30 minutes. After rinsing them twice in washing buffer (1X) for 15 minutes each, the blots were equilibrated in 1x detection buffer for 5 minutes. The blots were incubated in the CSPD substrate solution (Chemiluminescent substrate, 0.25mM in 1X detection buffer) for 5 minutes in dark. The blots were sandwiched between two transparency sheets by heat-sealing them after allowing excess liquid to drip out. The blots were incubated at 37°C for 10 minutes and then exposed to X-ray films (FUJI medical X- ray film, super RX and blue sensitive) at room temperature for varying lengths of time in lightproof cassettes with intensifying screen. Positive clones were marked on the master plates stored at 4°C. The positive plaques were picked individually and extracted into 500µl of SM buffer containing 20µl chloroform and incubated overnight at 4°C. This phage stock was plated again.

3.14.2 Screening with antibody probe

Partially purified polyclonal antibodies, raised against protease enzyme in New Zealand white rabbits, available in Lab (Sreeja Chellappan 2005) was used as the antibody probe to screen the phage library. The library was induced for fusion protein expression by incubating the phage plates with IPTG impregnated nitrocellulose membranes. The recombinant fusion proteins were identified using the primary antibody (Ab) directed against the protease enzyme.

E. coli XL1 Blue MRF' bacteria were incubated with an appropriately diluted phage solution. This solution was mixed with top agar and poured onto NZY plates. The plates were incubated at 42°C for 4 hours until a dense bacterial lawn could be seen. Nitrocellulose membranes (Protran membranes, Schleicher & Schuell) were wetted in IPTG (10mM) solution. The membranes were placed onto the plates, which are incubated at 37°C for another 4 hours. The plates were stored at 4°C. The membranes' orientation was marked before they were removed and washed 3-5 times with wash buffer (TBST-Tris buffered saline containing 0.05% Tween 20). Subsequently the membranes were

blocked in 1% BSA in Tris buffered saline (TBS) by incubating the membrane for 1-2 hours at room temperature. Next, the membranes were incubated in protease antibody solution (diluted to 1/100 in the blocking solution) overnight at 4°C. After incubation, they were thoroughly rinsed in wash buffer for 3-5 times and incubated for 2 hours in Anti-rabbit IgG solution (conjugated with alkaline phosphatase -1/50,000 dilution). The membranes were washed 3-5 times in wash buffer and a final wash using Tris buffered saline alone to remove residual Tween 20.

3.15 Construction of partial genomic DNA library

A partial genomic DNA library containing DNA fragments from the 1.5-5kb region of the *Pst* I digested *E. album* chromosomal DNA was constructed in pUC18 vector predigested with *Pst* I. The ligated DNA was transformed to the *E. coli* DH5α and the transformants were selected by blue-white assay. Screening for Positive clones was done by colony hybridization using nucleotide probe (partial gene fragment of Eap), and further confirmed by colony PCR.

3.16 Colony Hybridization Screening

Probe labeling and hybridization of the membranes were done with DIG High Prime DNA labeling and detection Starter Kit II (Roche Applied Sciences, Germany) according to the manufacturer's protocol. The colonies were lifted onto positively charged nylon membranes cut to the size of the petri dishes. To identify the orientation, proper markings in the membrane were made by cutting the membrane at 3 random places. The membranes were serially treated with denaturation solution for 15 minutes, neutralization solution for 15 minutes and SSC (2X) for 10 minutes. The DNA from lysed colonies was fixed to the membrane by baking at 80°C for 30 minutes. After this, membranes were incubated for 1hr at 37°C with Proteinase K solution prepared in 2X SSC (60U/ml). The cell debris from the membranes was removed by pressing with Whatman 3mm filter paper (damp with sterile distilled water). This was repeated 3-4

times until the membranes turned clean. Hybridization of the membranes was carried out at 50°C in Hybridization buffer provided in the kit. *Eap* partial gene sequence was used as the probe. The subsequent washing and detection steps were performed as described under the heading "Plaque hybridization screening with heterologous cDNA probe" in section 3.14.1.

3.17 Southern analysis

3.17.1 DNA digestion and size fractionation

Genomic DNA (10μg) was digested to completion with 4units/μg of appropriate restriction enzyme at the recommended temperature for 16 hours. The digested DNA samples were mixed with an appropriate volume of 6X dye (30% glycerol, 0.25% bromophenol blue and 0.25% xylene cyanol) and size fractionated by electrophoresis in 0.8% agarose gel in TPE at 2.0-5.0 volts/cm till the bromophenol blue reaches one cm above the bottom of the gel. After ethidium bromide staining (0.5μg /ml), the DNA was visualized under UV and photographed. Depurination of the size fractioned DNA was done with 0.25N HCl for 15-20 minutes at room temperature. Southern transfer (Southern, 1975) was done by capillary transfer or by vacuum blotting.

3.17.2 Vacuum blotting

The DNA was transferred to Hybond-N^+ membrane at a constant vacuum of 70mm Hg using a vacuum blotting assembly Transfer, first in denaturing solution (0.5N NaOH, 1.5M NaCl) for 40 minutes, followed by neutralizing solution (1.5M NaCl, 0.5M Tris-HCl pH 7.2, 1.0mM EDTA) for 2 hours. The Hybond-N^+ membrane was used directly for hybridization after air-drying without baking.

3.17.3 Southern hybridization (Southern 1975)

Blots were prehybridized in a solution of 0.5M sodium phosphate buffer (pH 7.5) and 7.0% SDS at 65°C for 2 hours in a hybridization oven (Hoefer, USA). Two hundred

µl of prehybridization mix was used per sq.cm of filter area. Hybridization was carried out in the same buffer with ^{32}P labeled probe (specific activity 10^8-10^9 cpm/µg) with a concentration of 1-5x10^6 cpm/ml at 65°C for 24 hours. The probe used was the *Eap* gene fragment.

3.17.4 Post-hybridization washing and autoradiography

Low stringency wash: The blots were washed twice in 2.0X SSC+0.1% SDS at 60°C for 30 minutes each.

High stringency wash: The blots were washed thrice in 0.1XSSC and 0.1% SDS at 65°C for 30 minutes each.

After washing, the blots were rinsed thrice in SSC (2X) at room temperature for 5 minutes each to remove SDS. Further, the blot was exposed to X-ray films at -70°C for varying lengths of time in lightproof cassettes with intensifying screen.

3.18 DNA sequencing

DNA sequencing was performed using the di-deoxy chain termination method of Sanger (Sanger et al. 1977) modified by (Chen & Seeburg 1985). Cycle sequencing was carried out with M13 universal primers or custom-made oligonucleotide primers. Purified super coiled plasmid DNA or PCR product was used for sequencing after purification and precipitation. 'Big Dye' terminator cycle sequencing kit (Perkin Elmer) provided the reaction mixture containing all the dNTPs and fluorescent labeled four deoxynucleotides along with Taq DNA polymerase in the sequencing buffer.

The final reaction mixture was made by mixing 2.0µl of terminator ready reaction mix (Perkin Elmer sequencing manual), 5.0 pmoles of primer and 300ng of plasmid DNA in a final reaction of 10µl. Cycle sequencing was carried out in a thermal cycler (MJ Research, USA) by first giving a rapid ramp to 96°C and holding at this temperature for

10 seconds. This step was followed by rapid thermal ramp to 50°C and holding for 10 seconds. Last step was rapid thermal ramp to 60°C and holding for 4 minutes. The cycle consisting of the above three steps was repeated 30 times. After 30 cycles, the reaction was rapidly brought to 4°C. The unincorporated dye terminators were removed by ethanol precipitation. The pellet was resuspended in 10µl of loading buffer (deionized 50% formamide), vortexed, spun at 10,000 rpm for 1-2 minutes and loaded on an automated DNA sequencer (Applied Biosystems, model 3700).

3.19 Homology modelling

The sequence of EAP corresponding to the mature secreted protein (residues 109–387) was used in homology modelling. Homology modelling was done using the program Swiss-PDB Viewer, SPDBV (now called DEEP VIEW at http://www.expasy.ch/swissmod/SWISS-MODEL.html). Proteinase K was used as a structural template (1IC6A, at the Protein Data Bank http://www.rcsb.org/pdb/) to model EAP (Betzel et al. 1988) and proteinase K complexed with the inhibitor 2-methoxysuccinyl-Ala-Ala-Pro-Ala-chloromethyl ketone (3PRK) as a template to model a substrate/ inhibitor into the active site of EAP.

3.20 Analysis of the nucleic acid sequences and deduced amino acid Sequence by various computer algorithms

The nucleic acid as well as deduced amino acid sequences were analyzed by various online and off line algorithms for structural prediction, phylogenetic relation with other published alkaline serine protease gene and their predicted protein sequences, and various other parameters (Table 3.1). The sequences were analyzed time to time by online BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al. 1997).

Table 3.1: Different bioinformatics tools used for sequence analysis

Bioinformatics tool	Task	Source
Genetool	o PCR Primer designing o Comprehensive analysis of DNA sequences	http://www.DoubleTwist.com
BioEdit	o Multiple alignment of sequence using Clustal W o Restriction analysis o Sequence Identity matrix o Kyte & Doolittle Mean Hydrophobicity profile	Hall 1999
MEGA 3	o Phylogenetic tree	Kumar et al. 2004b
BLAST	o Protein and nucleotide similarity search	http://www.ncbi.nlm.nih.gov
Codon usage	o Codon usage analysis	http://www.bioinformatics.org/sms2/codonusage.html
Translate	o Nucleotide translation	ExPASy Molecular Biology Server (Expert Protein Analysis System, proteomics server of the Swiss Institute of Bioinformatics (SIB), (http://us.expasy.org/)
SignalP 3.0	o Signal Peptide analysis	
TargetP v1.1	o Protein location	
NetNGlyc 1.01	o Protein N-glycosylation site	
NetOGlyc 3.1	o Protein O-glycosylation site	
Sulfinator	o Protein Tyrosine sulfation site	
ScanProsite	o Protein motif search	
InterPro	o Protein domain search	
PSI PRED	o Protein secondary structure analysis	
Swiss-PDB Viewer	o Homology modelling	
Antheprot	o Protein secondary structure analysis	http://antheprot-pbil.icbp.fr
PROCHECK	o 3D model quality analysis	Laskowski et al. 1993
GG v 1.0	o Calcium binding site prediction	http://www.chemistry.gsu.edu/faculty/Yang/GG.htm
DISULFIND	o Prediction of disulphide linkage	http://www.disulpfind.dsi.unifi.it
VMD	o Molecular graphics	http://www.ks.uiuc.edu/Research/vmd/

Chapter 4

RESULTS

4.1 cDNA library construction and screening

The cDNA library was constructed using the good quality RNA extracted from *Engyodontium album* using Tri reagent (Fig.4.1). Approximately 50 to 70 µg of RNA was obtained during independent isolations from 100 mg of fresh mycelia. From this total RNA, mRNA was obtained using Poly (A) tract mRNA isolation system III and was used for cDNA synthesis. Initial titre of the library was enumerated as $2.4x10^4$ plaque-forming unit per millilitre (Pfu/ml) and this on further amplification has increased to 10^9 Pfu/ml (Fig. 4.2).

Figure 4.1: Total RNA isolated from *E.album* mycelia, resolved on 1% agarose gel

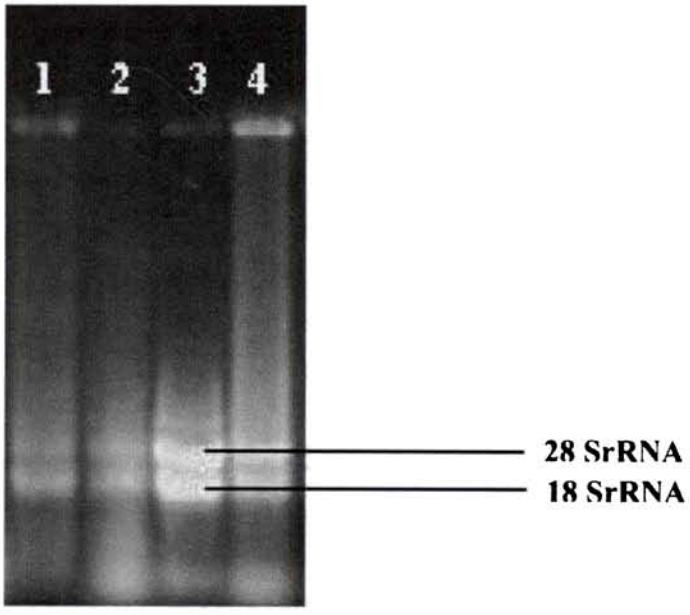
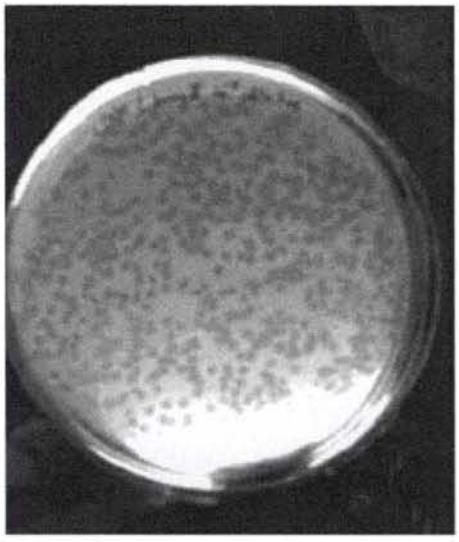


Figure 4.2: cDNA library in λZAP II vector (10^{-5} dilution)



4.2 Screening of the cDNA library

The prepared cDNA library was screened for the presence of alkaline serine protease gene in *E. album* (Eap) using a heterologous probe and an antibody prepared against the purified enzyme as detailed under materials and methods.

4.2.1 Using Heterologous probe

The probe sequence packed in pUC19 was first amplified in *E. coli* DH5α followed by a double digestion with Hind III and *EcoR*I. The released fragment (900bp) was eluted and purified (Fig.4.3). This isolated 900 bp fragment was labeled non-radioactively using DIG High Prime Labeling kit.

In order to assess the efficiency of DNA labeling, a dot blot analysis was carried out. The spot intensity of the sample was compared with that of the control and it was observed that up to 0.1pg DNA concentration, the spots from both sample and control had the same intensity, confirming the efficiency of labeling (Fig. 4.4). Based on this, the expected yield of the labeled DNA was calculated from the standardized chart given along with the kit which shows the yield of DIG High Prime

labeling under optimal conditions (Table 4.1). The concentration of labeled DNA obtained after 20hrs of incubation was calculated as 52.5ng/µl from the table. The recommended concentration of this probe DNA in the hybridization reactions is 25ng/ml hybridization solution.

Figure 4.3: *A. fumigatus* cDNA fragment (900bp) released from pUC19, eluted and purified for labeling. Lane 1- λ DNA Eco R I *Hind* III double digest. Lane 2-eluted fragment purified for the preparation of the heterologous probe

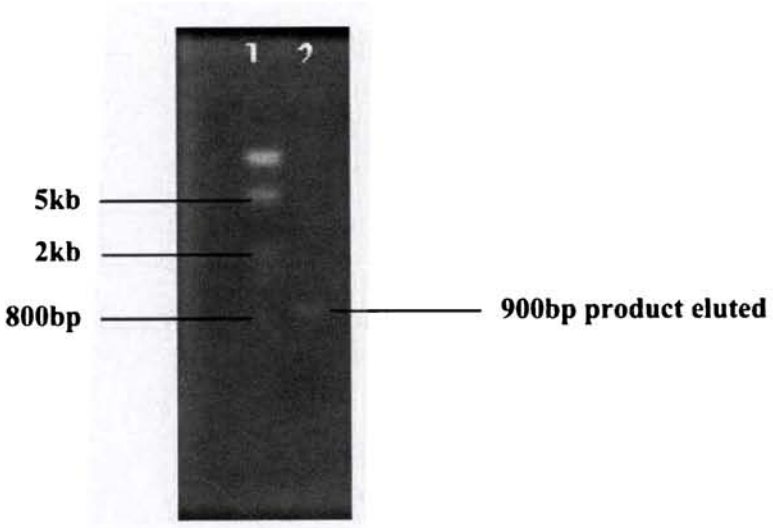


Figure 4.4: Dot blot analysis of the DIG labeled heterologous probe

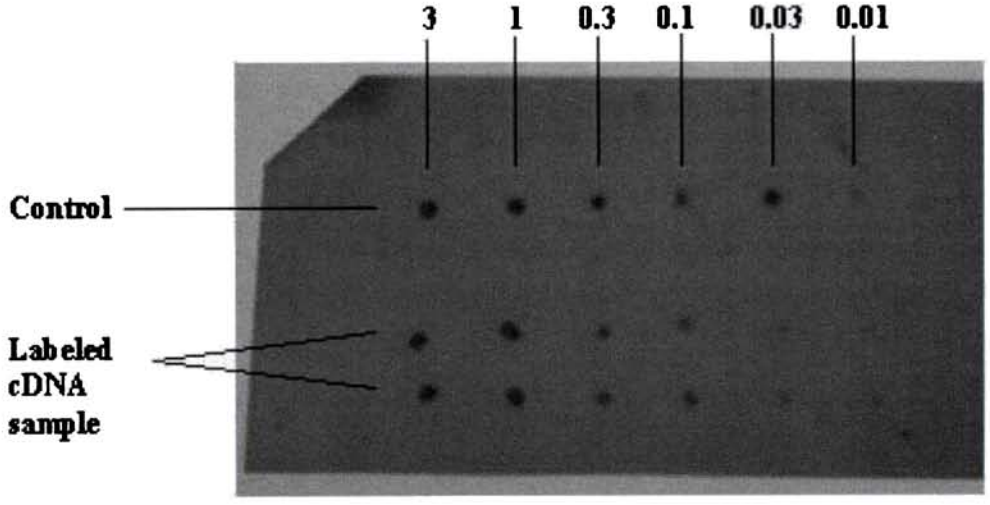


Table 4.1 Yield of DIG High Prime labeling under optimal conditions

Concentration of Template DNA (ng)	Concentration of labeled probe after incubation (ng)	
	1 hr	**20 hr**
10	45	600
30	130	1050
100	270	1500
300	450	2000
1000	850	2300
3000	1350	2650

The screening of cDNA library with the heterologous probe did not give productive results. The heterologous probe seemed to cross-hybridize with the ZAP II vector, which was indicated by the high background, obtained in X-ray sheets after the chemiluminescent detection. Different trials were done to improve the results, like increasing the stringency of washing conditions to the maximum level and raising the hybridization temperature several degrees higher than the prescribed hybridization temperature, but the results were not promising.

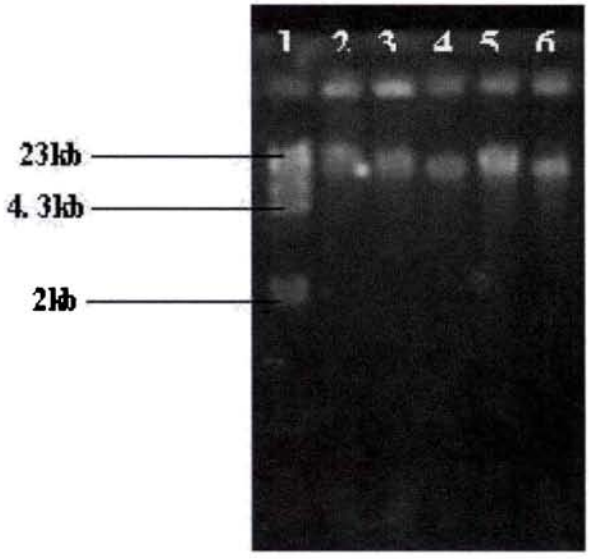
4.2.2 Using antibody probe

The cDNA library was screened with partially purified polyclonal antibodies raised against protease enzyme (Sreeja Chellappan, 2005). On the X-ray films, very low-level signals were obtained even after using low-stringency conditions and prolonged exposure (2-3days).

4.3 Isolation of high molecular weight chromosomal DNA from *Engyodontium album* BTMFS10

High molecular weight chromosomal DNA from *E. album* was isolated using modified CTAB method. Analysis of the isolated DNA by gel electrophoresis confirmed efficacy of the procedure in yielding high molecular weight DNA (Fig.4.5).

Figure 4.5: Genomic DNA isolated from *E. album* by CTAB method resolved on a 0.8% agarose gel. Lane 1- λ DNA *Hind* III digest. Lane 2-6 RNase treated Genomic DNA samples



4.4 Isolation of *Eap* gene fragment from *E. album* BTMFS10

Three degenerate primers were designed based on the conserved regions of the serine protease nucleic acid sequences available in NCBI. These three primers namely ASP1, ASP2 and ASP3 were used for the PCR amplification of serine protease gene of *E. album*, and an 878bp fragment was obtained by using the primer ASP1 (Fig 4.6). Subsequently the PCR product was cloned in the plasmid vector pMOS Blue, transformed in *E. coli* DH5α and plated. Colonies were selected randomly and plasmids were isolated (Fig 4.7). Colony with the asp1 insert was selected based on electrophoretic mobility and the presence of insert was confirmed by PCR with ASP1 primer. This recombinant plasmid was then sequenced and subsequently the sequence was searched against the entries in NCBI. The results indicated a very good homology with fungal serine alkaline protease genes and also with amino acid sequence indicating that it belongs to an internal coding region of the *E. album Eap* gene.

Figure 4.6: *E.album* genomic DNA PCR using ASP1 degenerate primers, which amplified a fragment of the gene. Lane 1- 10kb ladder, Lane 2-4 Amplification product of the primer ASP1

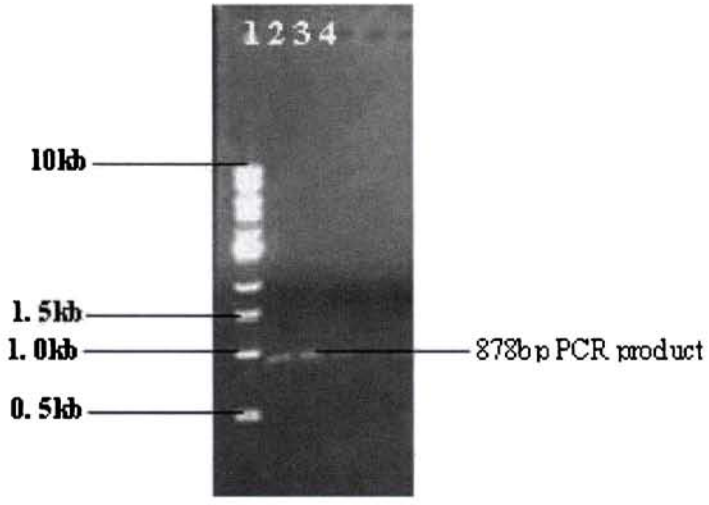
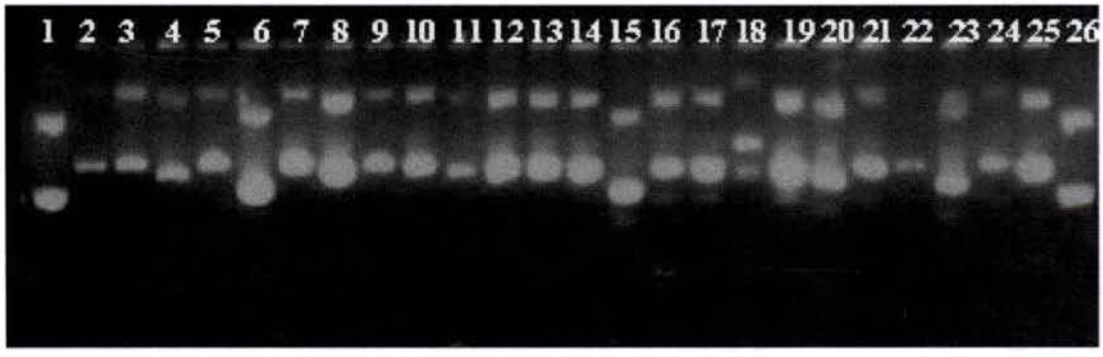


Figure 4.7: Aspl gene fragment cloned in pMOS blue plasmid vector. Lane 1- 10kb ladder Lane 1&26-control plasmid (pMOS blue without insert DNA), Lane 2-5, 7, 9-14 and 16-25 – recombinant plasmids with aspl insert DNA



4.5 Southern hybridization

The plasmid pEap1 containing an internal coding region of the *Eap* gene was used as the probe in the southern analysis of genomic DNA isolated from *E. album*. The 878bp fragment obtained through genomic DNA PCR was purified by gel elution and labeled with ^{32}P by multiple labeling method. The specific activity of the labeled probe was calculated as 10^{8}-10^{9} cpm/μg DNA.

Genomic DNA of *E. album* was digested to completion with nine different restriction enzymes viz *Hind* III, *Eco* RI, *Eco* RV, *Kpn* I, *Nde* I, *Pst* I, *Xba* I, *Bam* HI and *Not* I. The samples were then size fractionated by agarose electrophoresis and visualized after staining with ethidium bromide (Fig.4.8). The DNA in the gel after Depurination was transferred to Hybond- N^+ membrane by vacuum blotting method. The southern hybridization of the *Eap* gene using the prepared probe was carried out under high stringency conditions employing a higher temperature of 65°C for hybridization and up to 0.5X SSC for washing. A unique hybridizing fragment was revealed for each DNA digest (Fig.4.9).

The restriction fragments of *Hind* III, *Kpn* I, Nde I and *Pst* I yielded single signals of an approximate size of 6.5 kb, 4.5 kb, 8 kb and 5.5 kb respectively while the restriction fragments *Eco* RI, *Not* I, *Xba* I, *Bam* HI and *Eco* RV yielded a higher range of banding pattern like 13kb, 23kb, 20kb, 18kb and 17kb respectively. The results of southern blot indicate that there is only a single copy of the *Eap* gene in the *E. album* genome.

Figure 4.8: Restriction digestion profile of *E.album* genomic DNA before southern blotting. Lane 1 - *Hind* III digest, Lane 2 – *Eco* RI digest , Lane 3 – *Eco* RV digest, Lane 4 – *Kpn* I digest, Lane 5 – *Nde* I digest, Lane 6 – *Pst* I digest, Lane 7 – *Xba* I digest, Lane 8 – *Bam* HI digest, Lane 9 – *Not* I digest & Lane 10-λ DNA *Hind* III digest

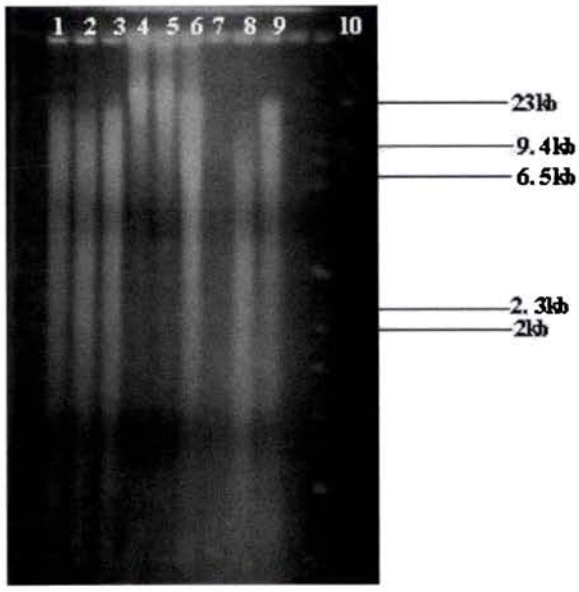
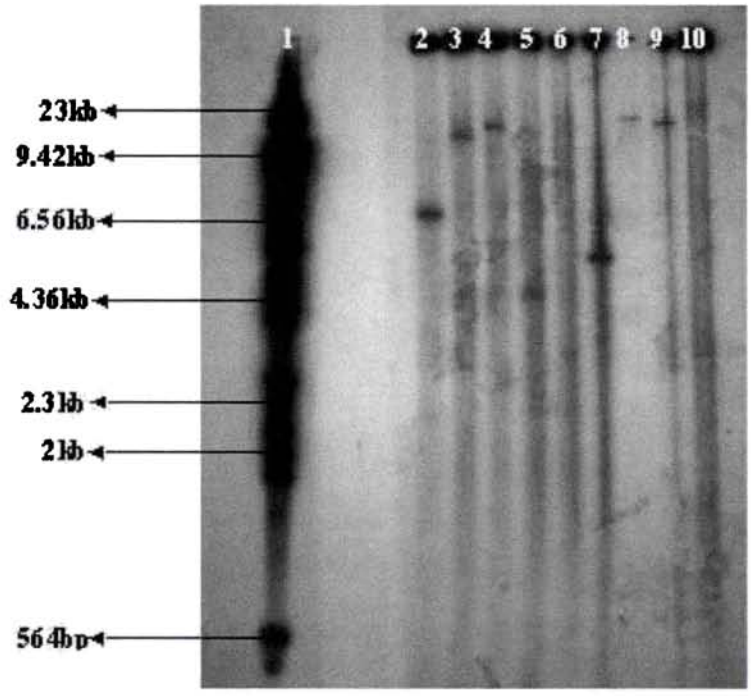


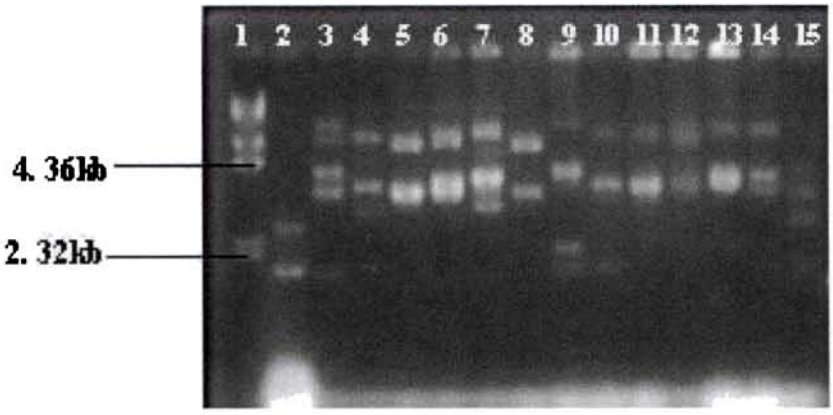
Figure 4.9: Southern Hybridization of *E. album* genomic DNA with partial *Eap* gene fragment as probe. Lane 1-λ□DNA *Hind* III digest, Lane 2 - *Hind* III digest, Lane 3 – *Eco* RI digest, Lane 4 – *Eco* RV digest, Lane 5 – *Kpn* I digest, Lane 6 – Nde I digest, Lane 7 – *Pst* I digest, Lane 8 – *Xba* I digest, Lane 9 – *Bam* HI digest, Lane 10 – *Not* I digest



4.6 Partial genomic DNA library construction

In order to facilitate the cloning and characterization of the full gene sequence of protease from *E. album* BTMFS10, a partial genomic library was constructed based on the data obtained from genomic southern analysis. The high molecular weight chromosomal DNA of *E. album* was digested with *Pst* 1, the fragments ranging from 1.5kb to 5kb size were purified from the gel and ligated to the plasmid vector pUC18 predigested with *Pst* 1. Following transformation and plating of the library, a total of 750 recombinants were obtained. The presence of inserts in clones was confirmed by checking the electrophoretic mobility of isolated plasmids (Fig 4.10).

Figure 4.10: Few selected recombinant plasmids isolated from the partial genomic DNA library of *E.album*. Lane-1-λDNA *Eco* RI digest, Lane-2 control plasmid (pUC18 without insert) Lane-3-15 Recombinant plasmids (pUC18 with *E.album* genomic DNA fragments as inserts)

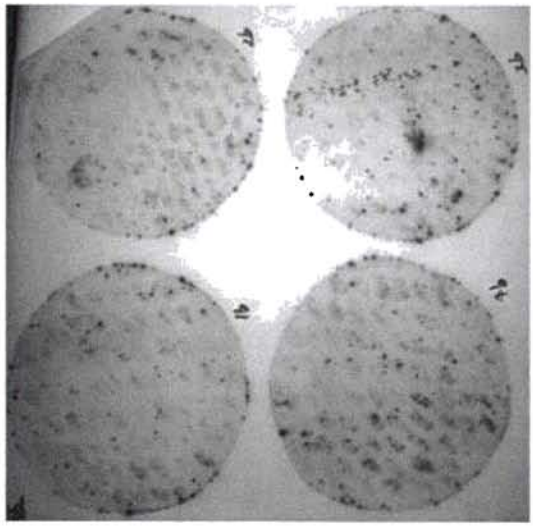


4.7 Screening of the partial genomic DNA library

4.7.1 Using *asp*1 probe

The recombinant colonies plated on LB ampicillin plates were blotted on to positively charged nylon membranes, treated with denaturation and neutralization solutions, washed in 2X SSC and subsequently the DNA from lysed colonies were fixed to the membrane by baking. Additionally, a proteinase K treatment was given to the membranes to remove the cell debris completely to avoid background and non-specific binding problems and afterwards hybridization was carried out. The *Eap* gene fragment was used as a probe to screen the genomic DNA library of *E. album* BTMFS10 by in situ colony hybridization. The probe was labeled with digoxigenin and detection was done with the CSPD in the chemiluminescent method (Fig 4.11). Out of the 750 recombinant colonies, forty clones that hybridized strongly to the probe were selected.

Figure 4.11: Colony hybridization of the partial genomic DNA library of *E. album* to determine the positive clone carrying *Eap* gene



4.7.2 Colony PCR

Considering the large number of positive clones (40) obtained after colony hybridization screening, a further short listing was done by screening them by colony PCR technique using the degenerate primer ASP1. In colony PCR, forty cultures bearing the positive clones were grouped in to 8 groups with 5 cultures in each group. Plasmid extracted from each group was amplified with the primer ASP1 and identified two positive groups. Plasmid was extracted from individual colonies of the positive groups (Fig 4.12) and the right clone with the protease gene was selected after conducting a PCR amplification using the same set of primer. Two recombinants pEap23 and pEap288 were selected finally. The sizes of the insert present in both the plasmids were compared to the standard DNA markers after digesting the plasmids with the restriction enzyme *Pst* 1. The inserts were of the size 1.4kb and 4.5kb respectively (Fig 4.13; Fig 4.14). The purified recombinant plasmid DNA (pEap23 and pEap288) was subjected to DNA sequencing. The plasmid pEap23 was sequenced successfully while sequencing of the plasmid pEap288 was unsuccessful for reasons unknown. The obtained sequence of pEap23 when compared with the GenBank

entries by BLAST search revealed its identity as a fungal serine protease homologue. More over the sequence covered complete coding region of the gene.

Figure 4.12: Plasmid profile of individual clones of the positive groups selected by Colony PCR. Lane 1-λDNA *Eco* RI + *Hind* III double digest, Lane 2- control plasmid (pUC18 without insert) and Lane 3-9 Recombinant plasmids (pUC18 with *E. album* genomic DNA fragments as inserts)

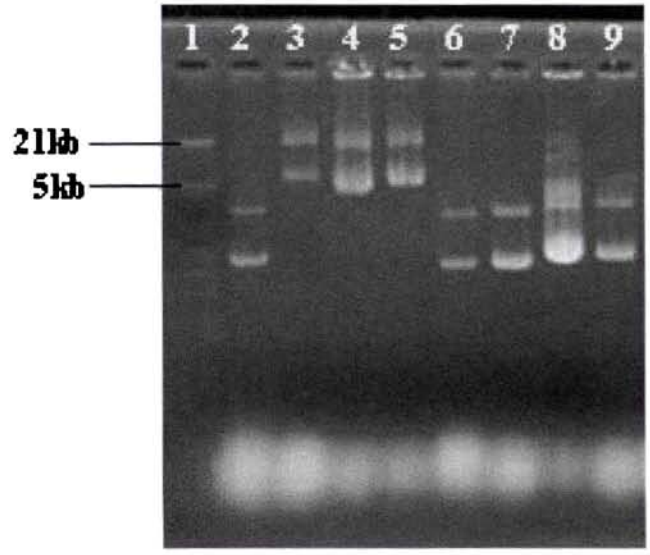


Figure 4.13: Insert release from pEap23. Lane 1. 500bp ladder, Lane 2-7. *Pst* I digest of pEap23 , Lane 8- uncut plasmid, pEap23

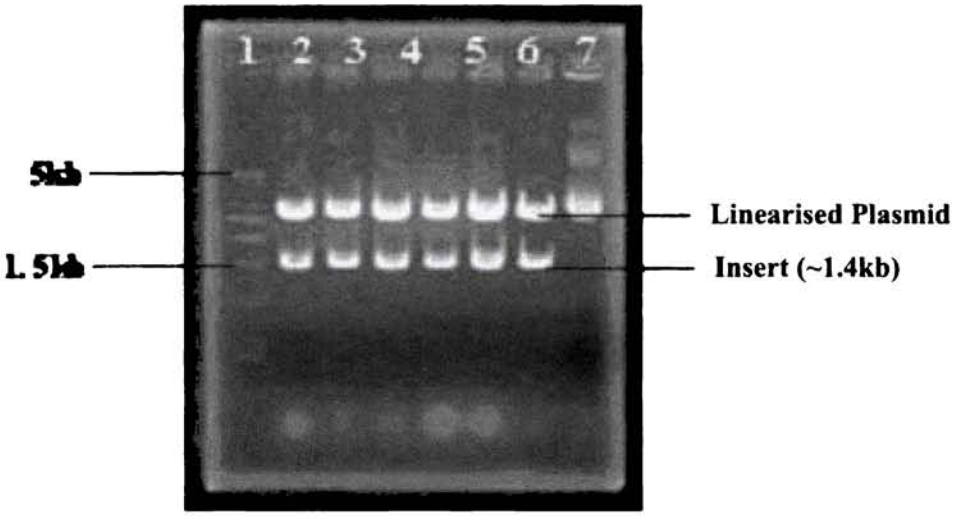
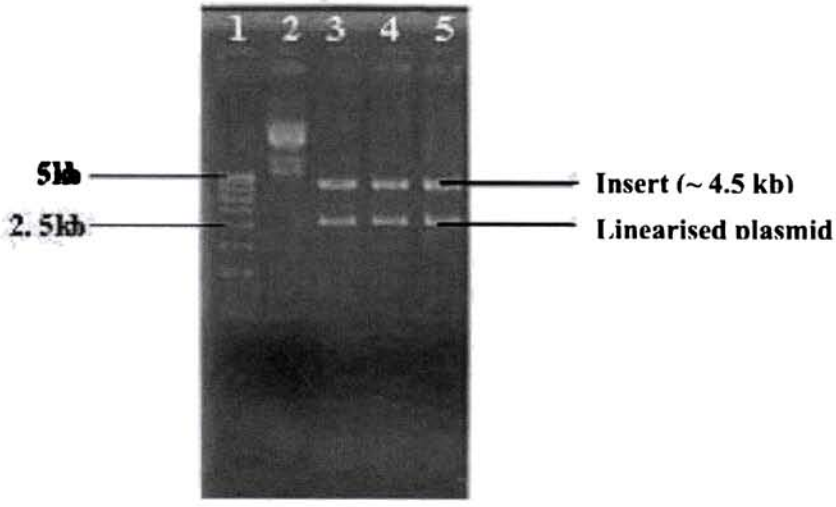


Figure 4.14: Insert release from pEap288. Lane 1- 500bp ladder, Lane 2- uncut plasmid, Lane 3 to 5- pEap288



4.8 Restriction profile of pEap23

The recombinant pEap23 was digested with different restriction enzymes and size fractionated on agarose gel (Fig 4.15). The restriction enzyme digestion profile obtained matched with the theoretical restriction profile obtained using BioEdit software. A schematic diagram of this restriction map is presented in Fig 4.16. For the analysis, the insert DNA was released from the vector backbone by digesting the plasmid, pEap23 with *Pst* I. The vector backbone can be seen in the picture as a 2.7kb fragment. The other enzymes used in the study were *Bgl* II, *Kpn* I, *Nco* I, *Sac* II and *Sal* I all in combination with *Pst* I. The restriction enzymes except for *Pst* I and *Kpn* I were selected based on the absence of their restriction sites in the vector backbone and the presence of restriction site in the insert.

Bgl II was found to be a unique cutter for *Eap,* producing two fragments of 749bp and 675bp. *Sal* I digested the *Eap* gene into two fragments of 149bp and 1275bp. Two fragments of 199bp and 1225bp were produced by *Sac* II digestion, whereas, *Nco* I digested the gene into three fragments, two of them were of same size, 270bp and the third one of 884bp. *Kpn* I cut the Eap gene four times producing five

fragments of 542bp, 390bp, 273bp, 168bp and 51bp. In the gel picture, the lower molecular size fragment (51bp) is not visible.

Figure 4.15: Restriction profile of *Eap* genomic DNA clone (pEap23). Lane 1-500 bp ladder; Lane 2- Udigested plasmid; Lane 3- *Pst* I digest; Lane 4- *Bgl* II/ *Pst* I digest; Lane 5- *Kpn* I/ *Pst* I digest ; Lane 6- *Nco* I / *Pst* I digest; Lane 7- *Sac* II/ *Pst* I digest; Lane 8- *Sal* I/ *Pst* I digest

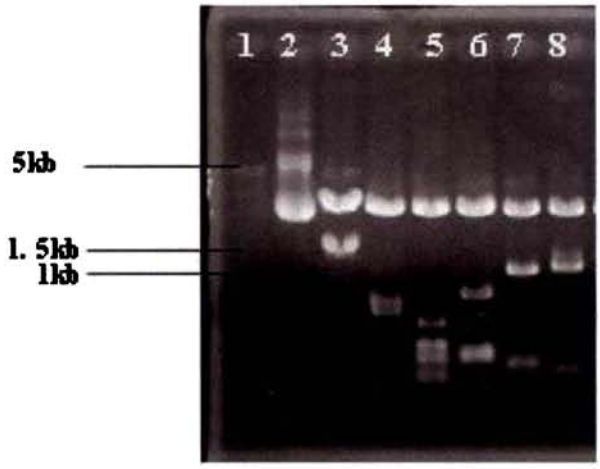
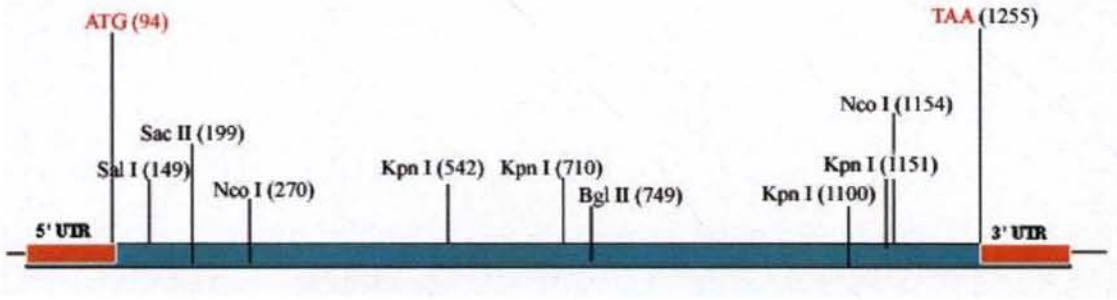


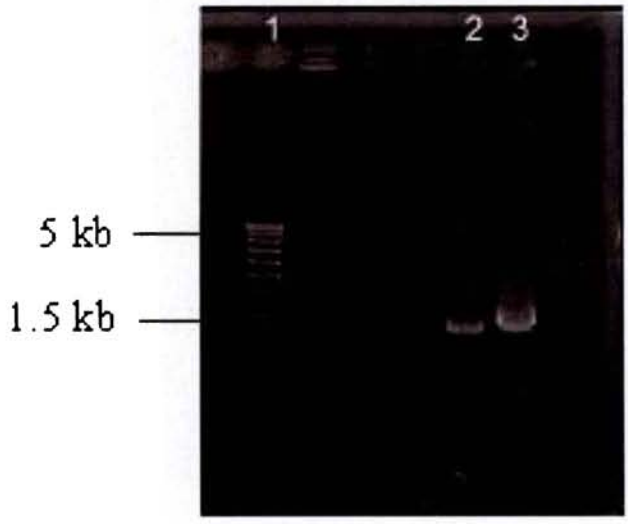
Figure 4.16: Schematic representation of the restriction sites present in the *Eap* genomic DNA



4.9 Amplification of *Eap* cDNA

The gene specific primers Eap UTR Forward and Reverse were designed based on the 5' and 3' non-coding sequences obtained from the genomic clone pEap23. The RT-PCR amplification of *Eap* yielded approximately 1350bp product after 35 cycles of amplification (Fig.4.17). The amplified product was gel eluted, cloned in pMOS blue and sequenced to confirm intron splicing.

Figure 4.17: RT-PCR amplification of coding sequence of *Eap*. Lane 1- 500bp ladder, Lane 2- *Eap* RT-PCR product, Lane 3- *Eap* PCR product



4.10 Sequence analysis by various computer algorithms

4.10.1 Nucleotide sequence of *Eap* gene

The *Eap* locus in the genomic clone pEap23 contains a total of 1,424bp (Fig.4.18). Sequence analysis revealed an open reading frame consisting of 1161bp which is flanked by 93bp in the 5'non-coding region and 102bp in the 3' non-coding region. The ORF starts with an ATG initiator codon at nucleotide position 94 and ends in a TAA stop codon at position 1320. The two lines of evidence support the assignment of translation initiation to the mentioned ATG codon: it is the first ATG codon in the cDNA in frame with the mature EAP protease and its flanking regions are similar to the consensus sequences of eukaryotic ribosomal initiation sites (Kozak 1986). The ORF is interrupted by a putative 65bp intron, starting at position 379, that displays the 5' and 3' consensus sequences for intron splicing and the internal sequence for lariat formation, described for filamentous fungi (Ballace 1991). Comparison of the nucleotide sequence of the genomic clone (pEap23) with that of the cDNA confirmed the presence of the single intron (65bp) in the coding region of EAP.

Coding region of subtilisin like serine proteases from various fungi were aligned by Clustal W alignment method of "BioEdit" (Hall 1999) (Fig.4.19). Identity between the *Eap* sequence and other fungal serine protease sequences at the nucleic acid level was obtained by plotting the identity matrix of the alignment (Table 4.2). *Eap* shows 96-87.5% identity to *T. album* proteinase R and K, 67.4-67.5% with *Fusarium* sp Alp and *Lecanicillum psaliotale* Ver112, 64-63.9% with *Beauveria bassiana bsn*1 and *Cordyceps brongniartii prl*, 62.5-63.9% with *Verticillum chlamydosporium* p1 and *Tolypocladium inflatum* prots and 65.5% with prlA of *Metarhizium anisopliae*.

The G+C content of the *Eap* coding region is 59%. The codon usage for *Eap* is presented in Table 4.3 (http://www.bioinformatics.org/sms2/codon_usage.html). Of the 61 possible codons, 44 are used in the *Eap* gene; however, 90% of the amino acids are coded by only 28 codons. The codon usage in *Eap* shows a very marked bias as is true in nearly all highly expressed gene from filamentous fungi. As a rule, whenever possible, a T or a C is preferred at the third positions. In *Eap*, of all the codons used, 58% end with C, 40% end with T, 25% end with G and only 3% of the codons used end with A (Gurr et al. 1988). A codon bias is also evident in the codon representation of several amino acids. For example, the codon AAC for Asparagine is used 22 times and the codon AAT is not used. For the amino acids proline and serine, two or more codons are absent: CCA and CCG for proline and AGU, UCG and UCA for serine.

Sequence of the pEap23 clone containing EAP gene was submitted to GenBank through BankIt programme, at NCBI site (http://www.ncbi.nlm.nih.gov/ BankIt). The EAP sequence was given the accession number DQ 268654.

Figure 4.18: Nucleotide sequence of *Eap genomic* DNA (GenBank Accession: DQ 268654). Single letter amino acid translations are indicated below the nucleotide sequences. Start and stop codons are underlined and shaded in green. The region is shaded in magenta denotes the intron of the *Eap* gene

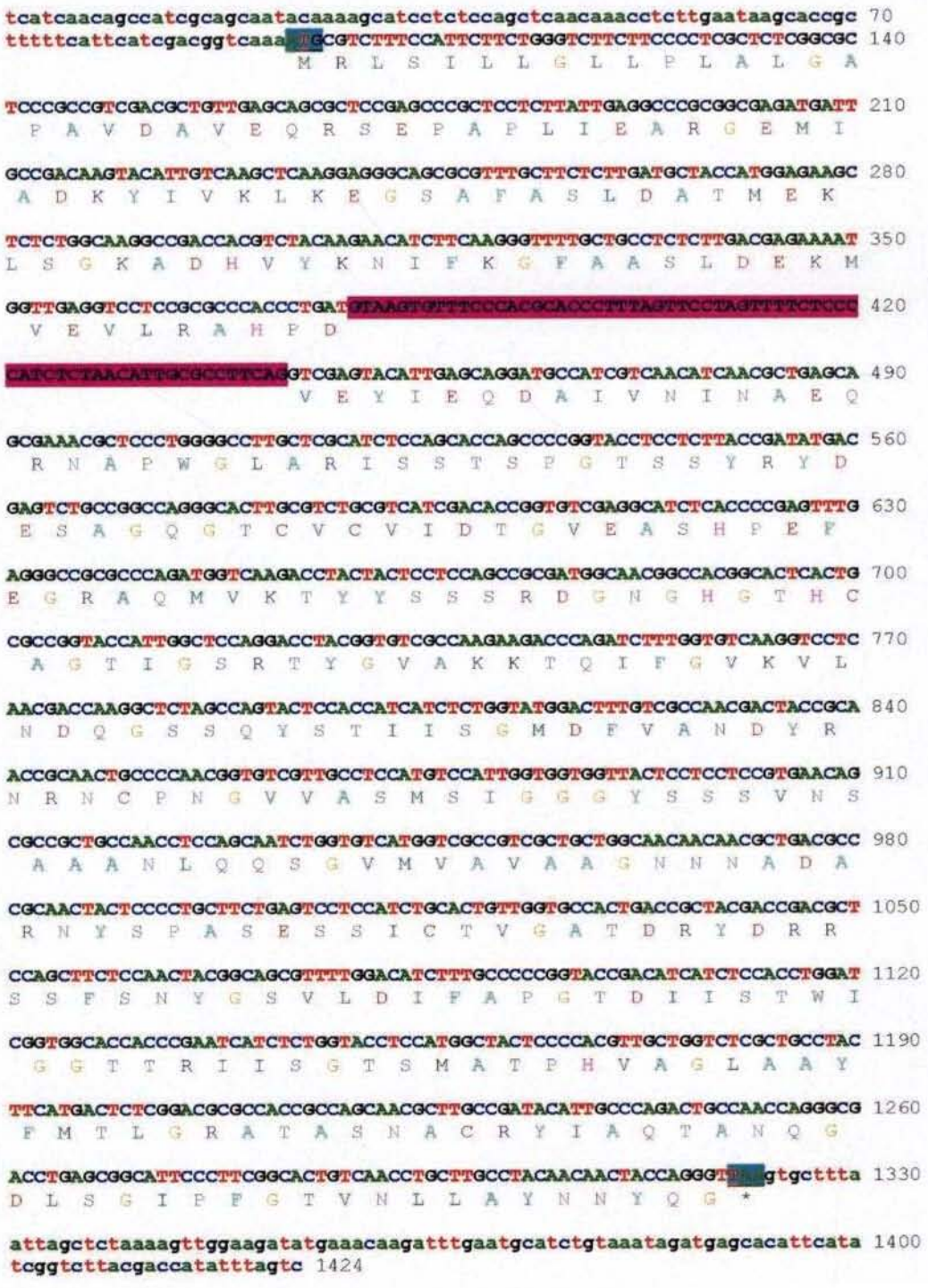
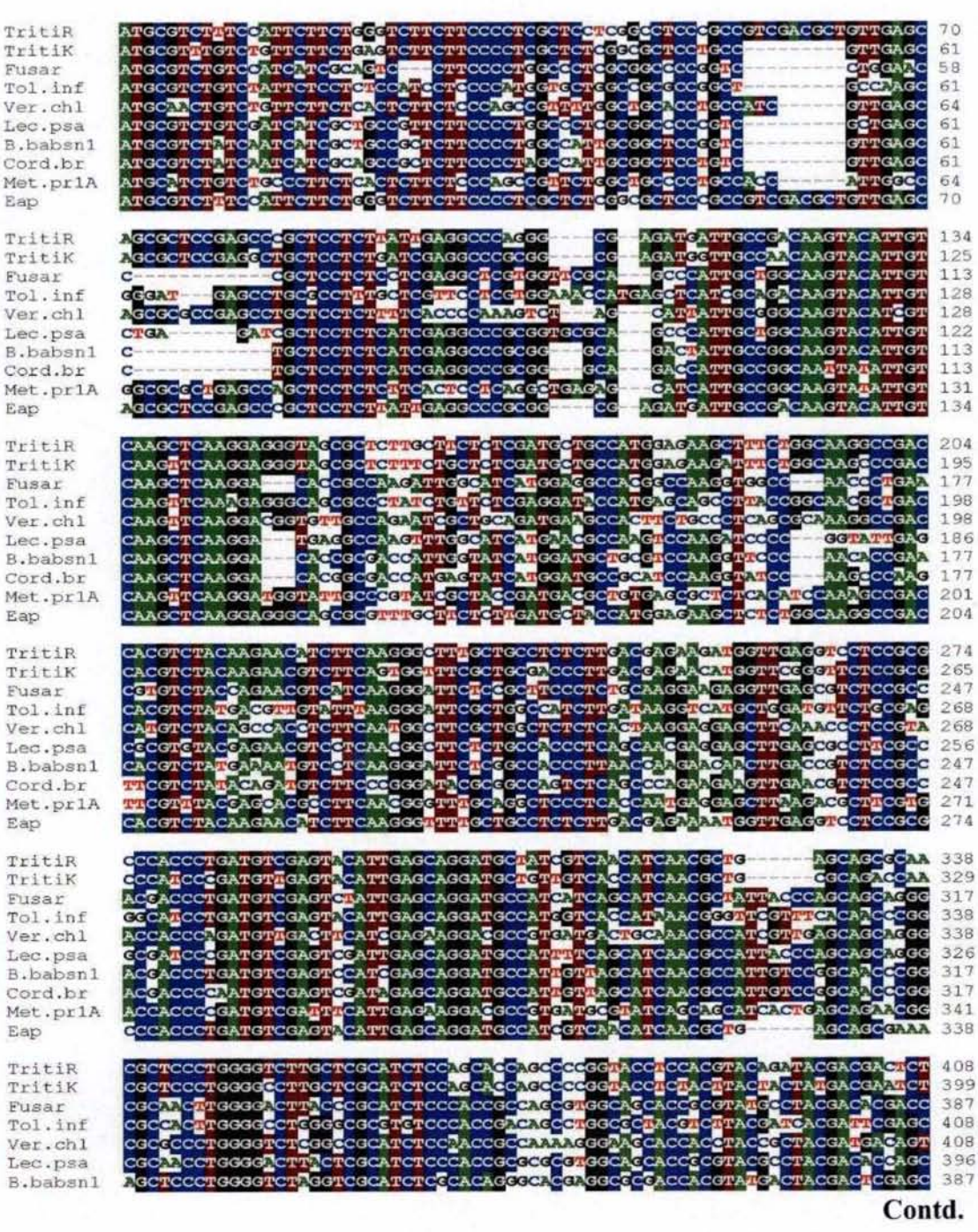
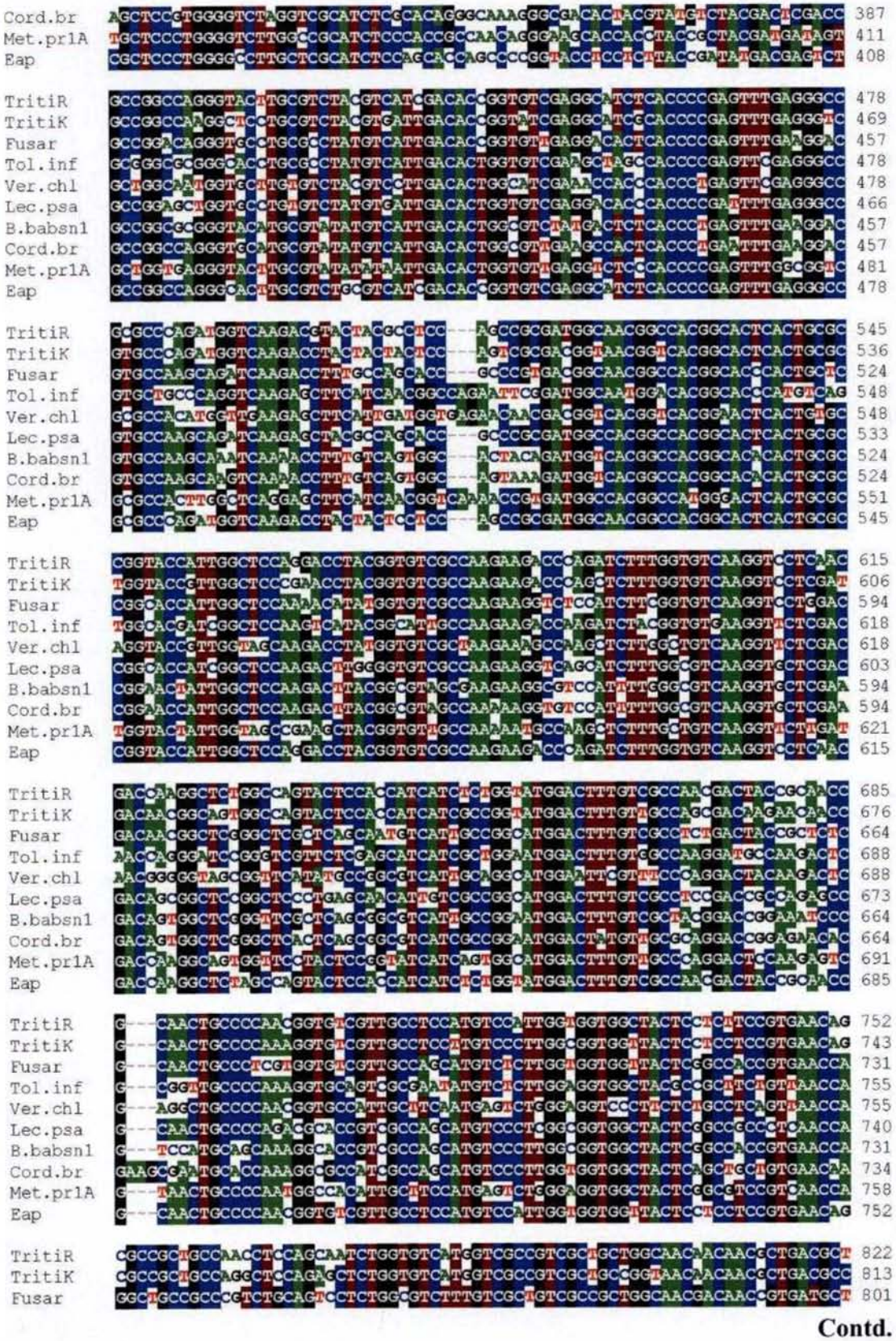


Table 4.2: Sequence identity of *Eap* with other fungal serine protease nucleic acid sequences calculated by sequence identity matrix programme of BioEdit

Organism and gene	GenBank accession	% Identity between EAP (DQ 268654) and other fungal serine protease nucleic acid sequences
Tritirarchium album proteinase R	X56116.1	**96**
Tritirarchium album proteinase K	X14688.1	**87.5**
Fusarium sp. S-19-5 Alp gene	S71812.1	67.0
Tolypocladium inflatum prots	AF467983.1	63.2
Verticillium chlamydosporium pl	AJ427460.1	63.0
Lecanicillium psalliotae ver112	AY692148.1	67.5
Beauveria bassiana bsn1	AF154118.1	64.1
Cordyceps brongniartii Prl	AY520815.1	63.8
Metarhizium anisopliae prlA	AJ251925.1	65.9
Paecilomyces lilacinus ser Prot	L29262.1	62.5
Leptosphaeria maculans secreted prot	AY422213.1	56.1
Phaeosphaeria nodorum subtilisin-like ser Prot	AY135714.1	57.2
Asprgillus oryzae alp	X17561.1	47.9
Trichoderma atroviride vacuolar ser Prot	DQ661007.1	39.6
Trichophyton verrucosum subtilisin like ser Prot	AY439105.1	49.5
Aspergillus flavus alkaline prot.	AF324246.1	47.9
Monacrosporium elegans cuticle-degrading prot.	AY859781.1	47.0

Figure 4.19: Clustal W multiple alignment of *Eap* gene sequence with other fungal subtilase gene sequences which showed maximum sequence identity. Nucleotides are numbered with respect to the Start codon (ATG). Gaps are introduced to optimize the alignment. TritiR (*Tritirarchium album* proteinase R; X56116.1), TritiK (*Tritirarchium album* proteinase K; X14688.1), Fusar (*Fusarium* sp. S-19-5 Alp; S71812.1), Tol.inf (*Tolypocladium inflatum*; AF467983.1), Ver.chl (*Verticillium chlamydosporium* AJ427460.1), Lec.psa (*Lecanicillium psalliotae* ver112; AY692148), B.babsn1 (*Beauveria bassiana* bsn1; AF154118.1), Cord.br (*Cordyceps brongniartii* Pr1; AY520815.1), Met.prlA (*Metarhizium anisopliae* pr1A; AJ251925.1), Eap (*Engyodontium album* alkaline serine protease; DQ 268654)




Cord.br
Met.pr1A
Eap
TritiR
TritiK
Fusar
Tol.inf
Ver.chl
Lec.psa
B.babsn1
Cord.br
Met.pr1A
Eap
TritiR
TritiK
Fusar
Tol.inf
Ver.chl
Lec.psa
B.babsn1
Cord.br
Met.pr1A
Eap
TritiR
TritiK
Fusar
Tol.inf
Ver.chl
Lec.psa
B.babsn1
Cord.br
Met.pr1A
Eap
TritiR
TritiK
Fusar
Tol.inf
Ver.chl
Lec.psa
B.babsn1
Cord.br
Met.pr1A
Eap
TritiR
TritiK
Fusar

Contd.

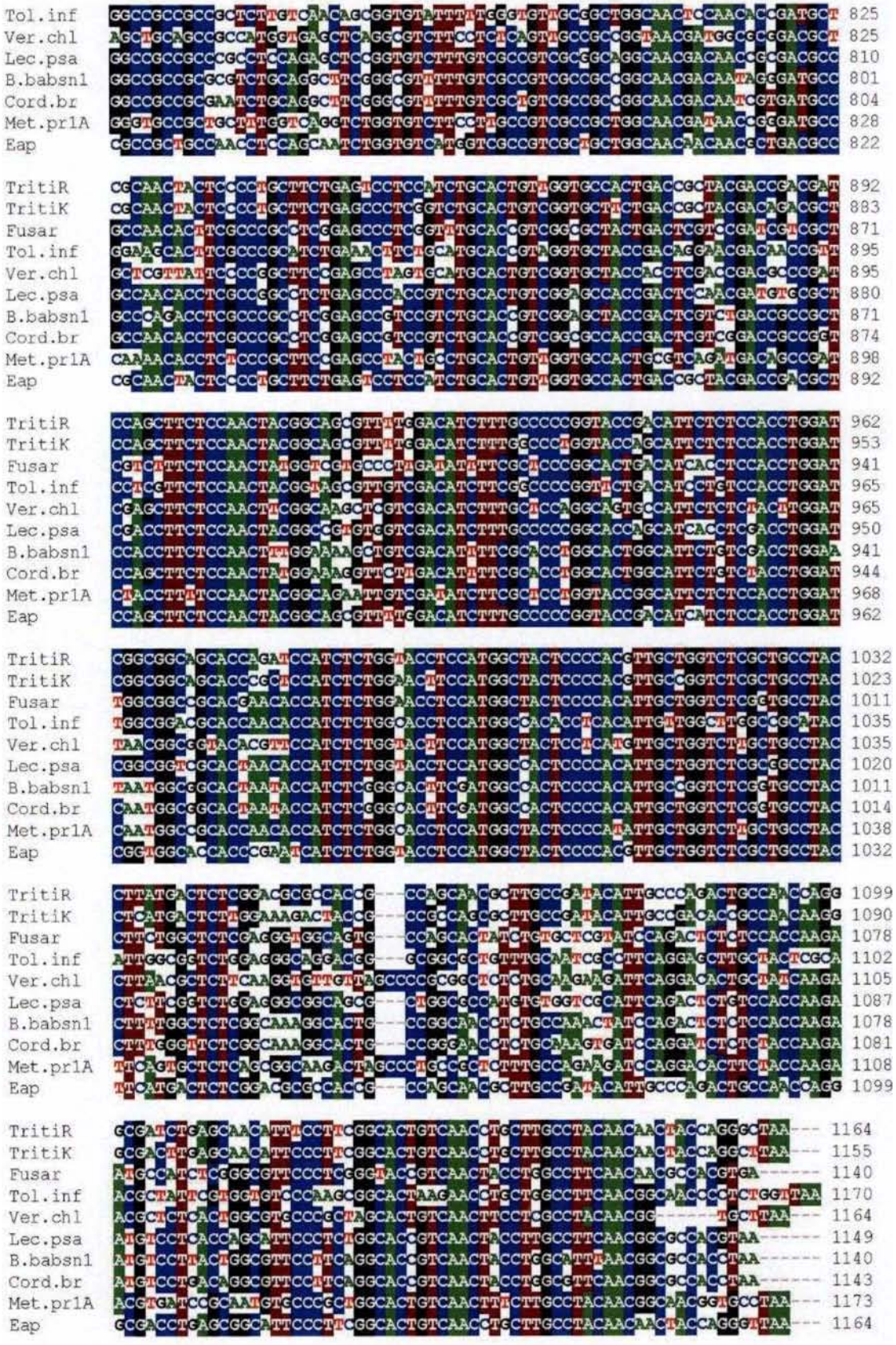

Tol.inf 825
Ver.chl 825
Lec.psa 810
B.babsn1 801
Cord.br 804
Met.pr1A 828
Eap 822
TritiR 892
TritiK 883
Fusar 871
Tol.inf 895
Ver.chl 895
Lec.psa 880
B.babsn1 871
Cord.br 874
Met.pr1A 898
Eap 892
TritiR 962
TritiK 953
Fusar 941
Tol.inf 965
Ver.chl 965
Lec.psa 950
B.babsn1 941
Cord.br 944
Met.pr1A 968
Eap 962
TritiR 1032
TritiK 1023
Fusar 1011
Tol.inf 1035
Ver.chl 1035
Lec.psa 1020
B.babsn1 1011
Cord.br 1014
Met.pr1A 1038
Eap 1032
TritiR 1099
TritiK 1090
Fusar 1078
Tol.inf 1102
Ver.chl 1105
Lec.psa 1087
B.babsn1 1078
Cord.br 1081
Met.pr1A 1108
Eap 1099
TritiR 1164
TritiK 1155
Fusar 1140
Tol.inf 1170
Ver.chl 1164
Lec.psa 1149
B.babsn1 1140
Cord.br 1143
Met.pr1A 1173
Eap 1164

Table 4.3: Codon usage for the *Eap* gene

aa	Codon	Number
Met	ATG	10.00
Leu	TTG	1.00
	TTA	0.00
	CTG	3.00
	CTA	0.00
	CTT	9.00
	CTC	9.00
Arg	AGG	1.00
	AGA	0.00
	CGG	0.00
	CGA	5.00
	CGT	1.00
	CGC	12.00
Ser	AGT	0.00
	AGC	10.00
	TCG	0.00
	TCA	0.00
	TCT	11.00
	TCC	20.00
Ile	ATA	0.00
	ATT	9.00
	ATC	15.00
Thr	ACG	0.00
	ACA	0.00
	ACT	8.00
	ACC	15.00
Gly	GGG	0.00
	GGA	1.00
	GGT	19.00
	GGC	19.00
Ala	GCG	1.00
	GCA	1.00
	GCT	19.00
	GCC	25.00

aa	Codon	Number
Pro	CCG	0.00
	CCA	0.00
	CCT	3.00
	CCC	10.00
Tyr	TAT	1.00
	TAC	17.00
Trp	TGG	2.00
	TGA	0.00
Lys	AAG	11.00
	AAA	1.00
Gln	CAG	11.00
	CAA	2.00
Asn	AAT	0.00
	AAC	22.00
His	CAT	0.00
	CAC	6.00
Val	GTG	1.00
	GTA	0.00
	GTT	6.00
	GTC	19.00
Phe	TTT	6.00
	TTC	4.00
Glu	GAG	16.00
	GAA	0.00
Asp	GAT	4.00
	GAC	15.00
Cys	TGT	0.00
	TGC	6.00
End	TAG	0.00
	TAA	1.00

4.10.2 The deduced aminoacid sequence of *Eap*

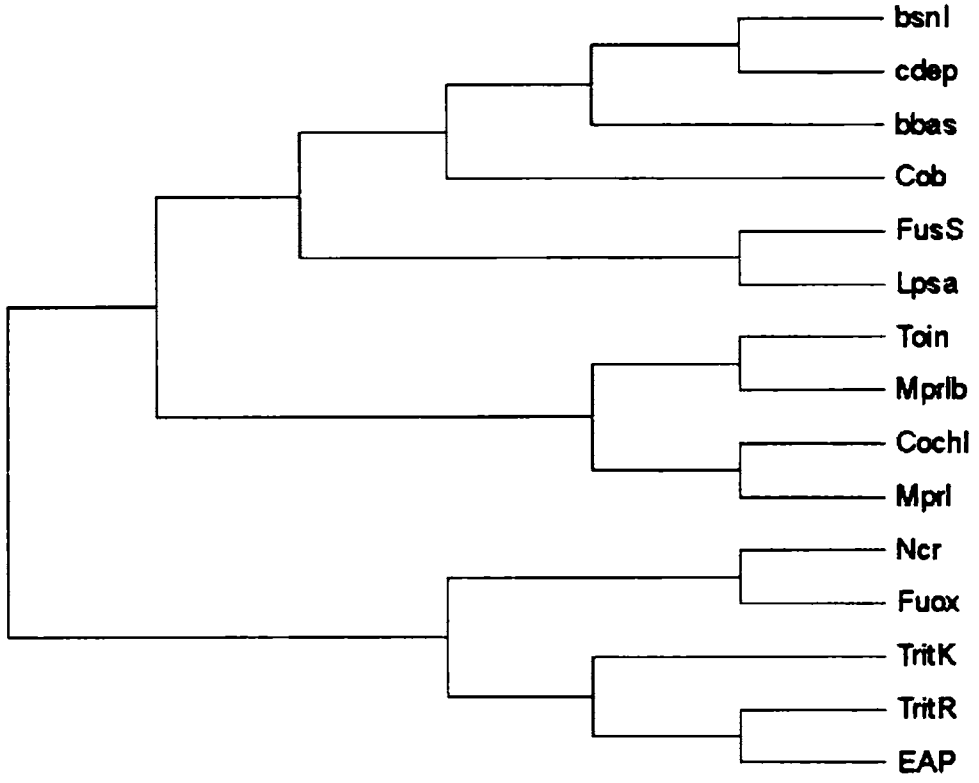
The *Eap* gene encoded a polypeptide consisting of 387 amino acids with a calculated molecular mass of 40.923kDa. A BLASTP (Altschul et al. 1990, Altschul et al. 1997) comparison with the GenBank database revealed that the encoded protein had homology with members of the subtilisin family of serine proteases. The deduced EAP protein exhibited 96.0%, 84.0%, 62.3%, 62.0%, 59.8%, 58.6% and 49.3% identity respectively to proteinase R of *Tritirarchium album* (Samal et al. 1990); proteainase K of *Tritirarchium album* (Gunkel & Gassen 1989); alkaline protease of *Fusarium* (Morita et al. 1994); Prl of *Beauveria brongniartii* (Sheng et al. 2006); bsnl of *Beauveria bassiana* (Joshi et al. 1995); prl of *Metarhizium anisopliae* (St. Leger et al. 1992) and prlB of *Metarhizium anisopliae* (Joshi et al. 1997). The percentage identity of the deduced amino acid sequence of EAP with other fungal subtilases is given in Table 4.4.

A dendrogram representing the measure of sequence homology was created from the sequence alignment of various members of subtilase family using the program, *MEGA* version 3.1 (Kumar et al. 2004b). The tree analysis revealed that EAP belongs to the *Tritirarchium album* serine protease cluster (Fig 4.20).

Table 4.4: Sequence identity of EAP with other fungal serine protease aminoacid sequences calculated by sequence identity matrix programme of BioEdit

Organism and Protein	GenBank accession	% Identity between EAP and other fungal serine protease aminoacid sequences
Tritirarchium album proteinase R	P23653	**95.6**
Tritirarchium album proteinase K	P06873	**84.2**
Fusarium sp. (S-19-5) ALP	AAC60571.2	62.3
Tolypocladium inflatum protease	AAL75579.1	60.6
Lecanicillium psalliotae ver112	Q68GV9	59.2
Beauveria bassiana bsn1	AAD29255.1	59.8
Cordyceps brongniartii Pr1	AAR97273.1	62
Metarhizium anisopliae pr1A	P29138	58.6
Metarhizium anisopliae pr1B	AAC49831.1	49.3
Neurospora crassa probable endopeptidase K	CAD71122.1	53.1
Fusarium oxysporum serine protease precursor	BAD72940.1	51.6
Beauveria bassiana subtilisin-like protease precursor	AAC48979.1	55
Cordyceps chlamydosporia alk ser prot	CAD20578.1	56.3
Beauveria bassiana CDEP-1	AAK70804.1	58.6

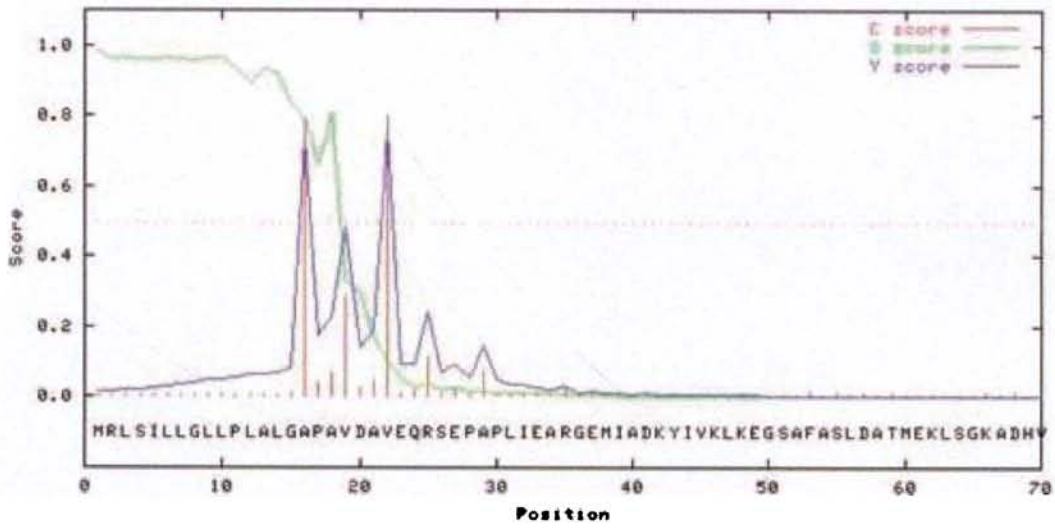
Figure 4.20: Unrooted neighbour-joining tree showing the relationship between EAP and other fungal subtilases.
TritR (*Tritirarchium album* proteinase R; P23653), TritK (*Tritirarchium album* proteinase K; P06873), FusS (*Fusarium* sp. (S-19-5) ALP; AAC60571.2), Toin (*Tolypocladium inflatum* protease; AAL75579.1), Lpsa (*Lecanicillium psalliotae* ver112; Q68GV9), bsnl (*Beauveria bassiana* bsn1; AAD29255.1), Cob (*Cordyceps brongniartii* Pr1; AAR97273.1), Mprl (*Metarhizium anisopliae* pr1A; P29138), Mprlb (*Metarhizium anisopliae* pr1B; AAC49831.1), Ncr (*Neurospora crassa* probable endopeptidase K; CAD71122.1), Fuox (*Fusarium oxysporum* serine protease precursor; BAD72940.1), bbas (*Beauveria bassiana* subtilisin-like protease precursor; AAC48979.1), Cochl (*Cordyceps chlamydosporia* alkaline serine protease; CAD20578.1), cdep (*Beauveria bassiana* CDEP-1; AAK70804.1), EAP (*Engyodontium album* alkaline serine protease precursor; DQ 268654)



Amino acid sequence homology analysis of the EAP protease with other subtilisin-like proteases indicated that the EAP protease is translated as a precursor protein. The analysis of the protein sequence with TargetP v1.1 (http://www.cbs.dtu.dk) predicted that the coded protein is destined to the secretory pathway, which is directed to the endoplasmic reticulum by a signal peptide (also called as leader peptide) of 21aa. Later, analysis of the signal peptide cleavage site using signalP 3.0 web server (http://www.cbs.dtu.dk) confirmed the presence of the signal peptide at the N-terminal

end of the deduced protein using neural networks (NN) and hidden Markov models (HMM) (Fig 4.21 and 4.22). It also indicates that the most likely cleavage site of the signal peptide is between Ala^{21} and Val^{22} residues.

Figure 4.21: SignalP-NN result



 # Data:EAP

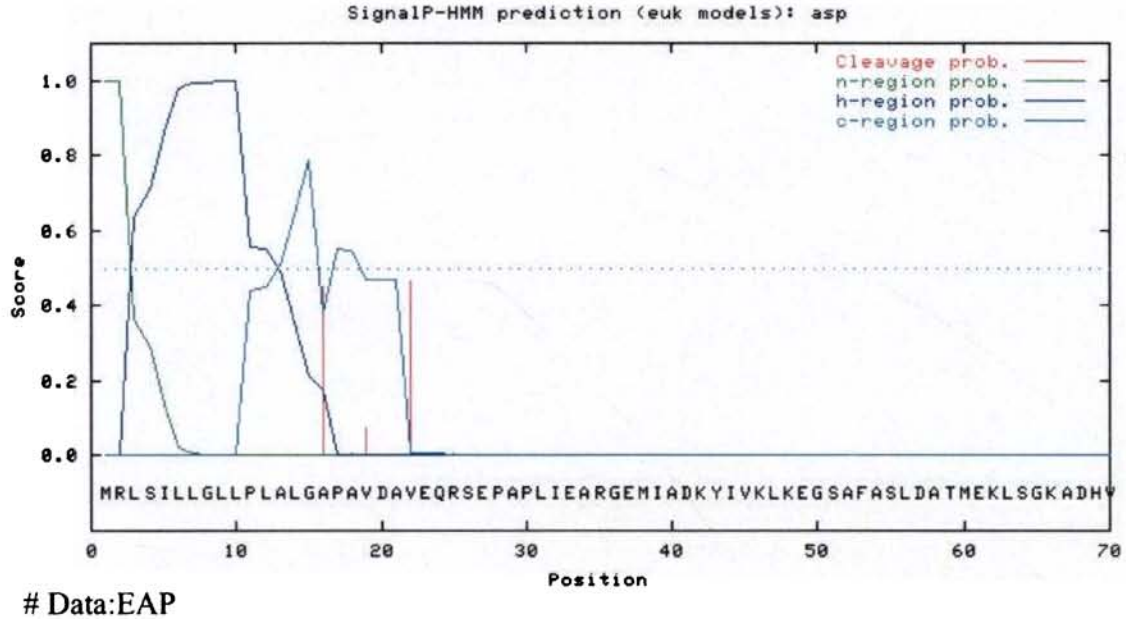
# Measure	Position	Value	Cutoff	signal peptide?
max. C	22	0.802	0.32	YES
max. Y	22	0.761	0.33	YES
max. S	1	0.984	0.87	YES
mean S	1-21	0.820	0.48	YES
D	1-21	0.790	0.43	YES

 # Most likely cleavage site between pos. 21 and 22: VDA-VE

A high *S-score* indicates that the corresponding amino acid is part of a signal peptide, and a low score indicate that the amino acid is part of a mature protein. The *C-score* is the "cleavage site" score, which should only be significantly high at the cleavage site. *Y-max* is a derivative of the C-score combined with the S-score resulting in a better cleavage site prediction than the raw C-score alone. This is due to the fact that multiple high-peaking C-scores can be found in one sequence, where only one is the true cleavage site. The cleavage site is assigned from the Y-score where the slope of the

S-score is steep and a significant C-score is found. The *S-mean* is the average of the S-score, ranging from the N-terminal amino acid to the amino acid assigned with the highest Y-max score. Thus the S-mean score is calculated for the length of the predicted signal peptide. The *D-score* is a simple average of the S-mean and Y-max score. The score shows superior discrimination performance of secretory and non-secretory proteins. For non-secretory proteins all the scores represented in the SignalP-NN output should ideally be very low.

Figure 4.22: SignalP-HMM result: The hidden Markov model calculates the probability of whether the submitted sequence contains a signal peptide or not. The eukaryotic HMM model also reports the probability of a signal anchor, previously named uncleaved signal peptides. Furthermore, the cleavage site is assigned by a probability score together with scores for the n-region, h-region, and c-region of the signal peptide, if such one is found



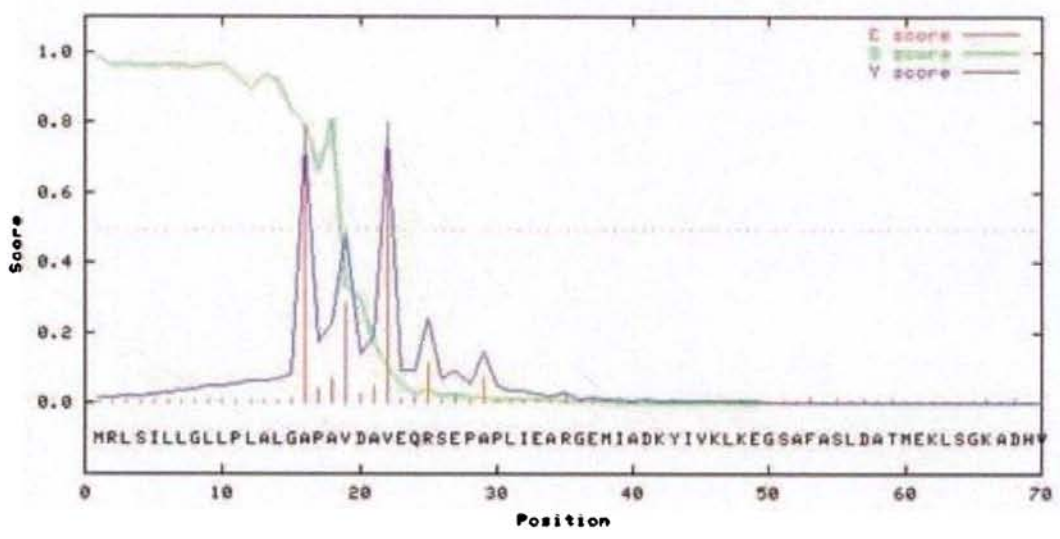
Data:EAP

Prediction: Signal peptide. Signal peptide probability: 1.000
Signal anchor probability: 0.000
Max cleavage site probability: 0.463 between pos. 21 and 22

The signal peptide consists of the initial methionine followed by a charged residue (Arg), a core of eight hydrophobic residues a helix breaking residue (Pro) and four residues before a signal peptidase cleavage site (Ala-Pro-Ala). This is consistent with the empirical rules of fungal pre-secretary sequences (Perlman & Halvorson 1983).

end of the deduced protein using neural networks (NN) and hidden Markov models (HMM) (Fig 4.21 and 4.22). It also indicates that the most likely cleavage site of the signal peptide is between Ala^{21} and Val^{22} residues.

Figure 4.21: SignalP-NN result



 # Data:EAP

# Measure	Position	Value	Cutoff	signal peptide?
max. C	22	0.802	0.32	YES
max. Y	22	0.761	0.33	YES
max. S	1	0.984	0.87	YES
mean S	1-21	0.820	0.48	YES
D	1-21	0.790	0.43	YES

 # Most likely cleavage site between pos. 21 and 22: VDA-VE

A high *S-score* indicates that the corresponding amino acid is part of a signal peptide, and a low score indicate that the amino acid is part of a mature protein. The *C-score* is the "cleavage site" score, which should only be significantly high at the cleavage site. *Y-max* is a derivative of the C-score combined with the S-score resulting in a better cleavage site prediction than the raw C-score alone. This is due to the fact that multiple high-peaking C-scores can be found in one sequence, where only one is the true cleavage site. The cleavage site is assigned from the Y-score where the slope of the

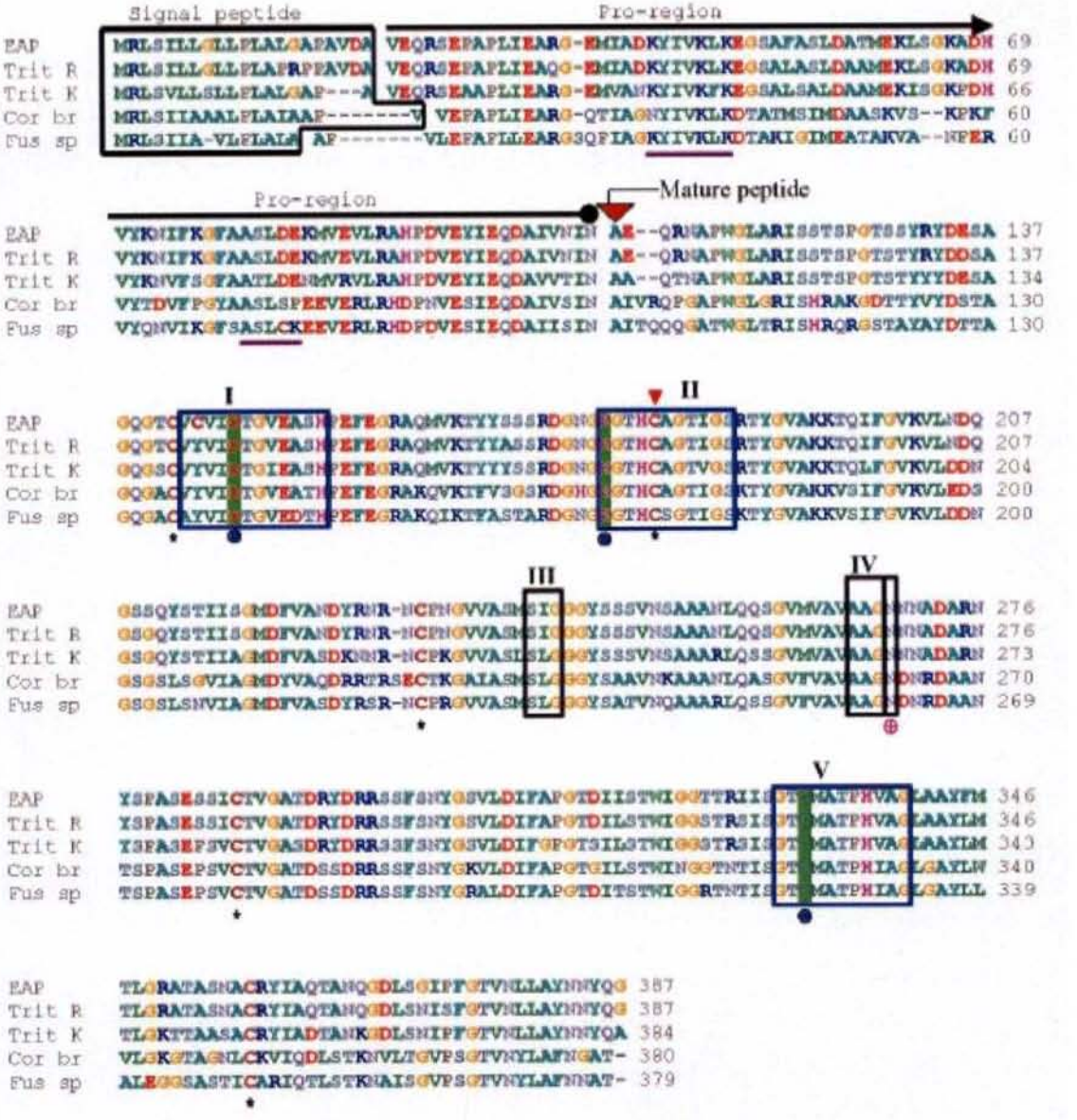
Next to the signal peptide, a putative propeptide region was identified from sequence alignments with fungal subtilases. It functions by binding to the enzyme active site as an inhibitor and are required for proper folding and secretion of the enzyme (Valls et al. 1987, Klionsky et al. 1988, Tatsumi et al. 1989). Although the N-terminal amino acid sequence of the mature EAP was not known, the comparison of the sequence of this protease with that of other known subtilisin-like proteases enabled it to postulate that the cleavage site of the N terminal propeptide localized between residues Asn^{108} and Ala^{109}. Thus the N-terminal propeptide of EAP was assumed to be composed of 87 amino acid residues with a calculated molecular mass of 9570 Da, and the mature protease domain consisted of 279 amino acid residues with a calculated molecular mass of 29.129kDa.

The propeptide and coding sequence of the mature EAP were ascertained by comparing mainly with the amino acid sequence of proteinase K and R (Jany et al. 1986, Samal et al. 1990) based on the high sequence similarity (84-96%)of the EAP with Proteinase K and R (*Tritirarchium album*) at both nucleic acid and amino acid level (Table 4.5).

Table 4.5: Comparison of *Eap* protease coding sequence with other reported fungal protease sequences showing high sequence identity

Organism and Gene	Cellular location	Genbank accession	Coding region		Signal peptide (aa)	Pro-peptide (aa)	Mature peptide (aa)
			Amino acids	Nucleo tides			
E. album, EAP	Extra cellular	DQ268654.	387	1164	21	87	279
T.album, Proteinase R	„	P23653	387	1164	21	87	279
T. album, Proteinase K	„	P06873	384	1155	15	90	279
C. brongniartii Prl	„	AAR97273.1	380	1143	18	81(?)	281(?)
Fusarium sp. Alp	„	AAC60571.2	379	1140	14	85	280

Figure 4.23: Clustal W multiple alignment of predicted aminoacid sequences of fungal subtilase genes that show maximum sequence identity with *Eap* sequences. Amino acids are numbered with respect to the first amino acid of the sequence. Gaps are introduced to optimize the alignment. The signal peptide, proregion and starting of mature peptide are marked. The two conserved stretches of residues in the pro region are underlined. Boxes (I to V) indicate conserved stretches assigned to subtilisin type serine protease motifs. * Marks represent the highly conserved cysteine residues present in all the five sequences. The three blue circles represent the catalytic triad residues of the proteases. The pink circle indicates the highly conserved asparagine residue and the small red triangle indicate the conserved free cysteine residue, both of which have important roles in catalysis. EAP (*E. album* protease), Trit. R (*T. album* proteinase R), Trit. K (*T. album* proteinase K) Cor. br (*C. Brongniartii Pr*1) Fus sp. (*Fusarium* sp *alp*). The genbank accession of these organisms are described in table 4.5



When the sequences of selected subtilisin like serine protease were aligned (Fig 4.23), five regions of major similarity were obtained. Three (I, II and V) of them correspond to the peptides that constitute the catalytic center of subtilisin like serine proteases. The two other identified regions are (III and IV) part of the specific crevice in subtilisin (sides of the S1 pocket) within which the P1 site of the substrate side chain fits. The putative upper and lower sides of the crevice are ^{240}Ser- ^{241}Ile - ^{242}Gly and ^{266}Ala - ^{267}Ala- ^{268}Gly respectively. The corresponding upper part of the crevice in proteinase K (^{237}Ser-^{238}Leu-^{239}Gly) forms hydrogen bonds with the P2 and P3 residues on the substrate (Kraut 1977, Betzel et al. 1986).

The occurrence of Gly-Thr-Ser-Met-Ala- residue (the signature sequence in the subtilisin like serine proteases) at position 330 to 334 of the protein confirms the fact that EAP is a subtilisin like enzyme. Moreover, the protein presents the typical features of the active site of subtilisin-type serine proteinases. The active site residues are at positions 147 (aspartic acid, D), 177 (histidine, H) and 332 (serine, S). From the first residue of the putative mature protein the positions of the active site residues are 39, 69 and 224 (Fig 4.23).

A free cysteine residue near the active site His in proteinase K has been postulated to play a role in catalysis (Betzel et al. 1986), which was also found conserved in EAP (Cys^{181}). A highly conserved Asn residue which is important in subtilisins for stabilization of the reaction intermediate formed during proteolysis (Kraut 1977) was observed in EAP (^{269}N) also (Fig 4.23).

Among the internal helix like structure near active site residues, the internal helix hC and hF, which are highly conserved in subtilases, are also being observed in EAP [177-184:HGTHCAGT; 330-344:GTSMATPHVAGLAAY] (Siezen et al. 1991).

There are five cysteine residues at positions 142, 181, 231, 286 and 357 in the mature region of the EAP. These residues are found perfectly conserved in the other four

proteases also (Fig 4.23). In proteinase K, four of these highly conserved cysteine molecules had been shown to form two disulphide bonds. The cysteine at 139 binds to that at 228 and the 283 to 354 (Jany et al. 1986). Considering the close similarity of EAP with that of proteinase K, presence of two disulphide bonds was proposed in EAP also.

The amino acid sequence was analysed for potential post-translational modification signals using various computer algorithms. Analysis of the deduced mature protein with NetNGlyc 1.0 and NetOGlyc 3.1(http://www.cbs.dtu.dk) predicted that EAP does not contain any potential N-glycosylation sites or O-glycosylation sites. For N-glycosylation, the general rule is Asn-X-Thr/Ser, where X is any residue except perhaps aspartate and for O-glycosylation, Ser-Thr residue pair represents the potential sites. The putative pro-peptide also not have potential glycosylation sites. One potential tyrosine sulfation site was being predicted in the EAP at Tyr^{98} (HPDVEYIE-QDAIVNIN) by the Sulfinator tool (http://www.cbs.dtu.dk) (Monigatti et al. 2002).

4.10.2.1 Protein Motif search

A protein sequence motif is broadly defined as a set of conserved amino acid residues that are important for protein function and located within a certain (short) distance from one another. These motifs can often provide clues to the functions of otherwise uncharacterized proteins.

The motif search for amino acid patterns in the EAP protein was done using a Scan Prosite tool in EXPASY proteomics server (Baroch et al. 1997). It detected three distinct patterns, which are characteristic of subtilases. These three patterns represent the long stretch of homologous residues in which the catalytic triad residues (Asp, His and Ser) are embedded. They are used as 'signatures' specific to the subtilisin family of serine proteases due to their highly conserved nature in such type of proteases. The three signatures include:

i) The sequence pattern around the aspartic acid active site (143-154aa: VCVIDTGVeasH) ii) The sequence pattern around the histidine active site (177-187aa: HGThCAGtIGS) and iii) the region around the serine active site (330-340aa: GTSmAtPhVAG) (Fig 4.23). If a protein includes at least two of the three active site signatures, the probability of it being a serine protease from the subtilase family is 100% (Siezen & Leunissen 1997).

The Scan Prosite tool also detected multiple potential phosphorylation sites in the deduced amino acid sequence. It predicted the presence of four classes of phosphorylation sites in the EAP. One group of sites is a potential substrate for cAMP- and cGMP-dependent protein kinases ([RK] [2]. [ST]), another group for protein kinase C ([ST]. [RK]), while the third group of sites is a potential substrate for Casein kinase II ([ST]. {2}[DE]) and the last group for Tyrosine kinase ([RK] .{2,3}[DE] .{2,3}Y). The pattern and positions of these sites in the EAP protein is given in the Table 4.6. In addition to this, a number of possible N-myristoylation sites (G [^EDRKHPFYW]. {2} [STAGCN] [^P]) are also detected in the EAP protein sequence by the Scan Prosite tool (Table 4.7).

Table 4.6: Predicted phosphorylation sites in EAP

Type of phosphorylation site	**Position**	**aa pattern**
cAMP-and cGMP-dependent protein kinase phosphorylation site	296 - 299	RRsS
Tyrosine kinase phosphorylation site	91-98	RahpDve.Y
Protein kinase C phosphorylation site	64 - 66	SgK
	130 - 132	SyR
	170 - 172	SsR
	291 - 293	TdR
	324 - 326	TtR
Casein kinase II phosphorylation site	81 - 84	SldE
	148 - 151	TgvE
	153 - 156	ShpE
	170 - 173	SsrD
	217 - 220	SgmD
	305 - 308	SvlD

Table 4.7: Predicted N-myristoylation sites in EAP

Site	aa pattern
50 - 55	GSafAS
138 - 143	GQgtCV
140 - 145	GTcvCV
149 - 154	GVeaSH
174 - 179	GNghGT
178 - 183	GThcAG
183 - 188	GTigSR

Site	aa pattern
234 - 239	GVvaSM
242 - 247	GGgySS
243 - 248	GgysSS
260 - 265	GVmvAV
268 - 273	GNnnAD
330 - 335	GTsmAT
371 - 376	GIpfGT

4.10.2.2 Protein Domain searches

Protein domains are the structurally compact, independently folding units that form stable three-dimensional structures and show a certain level of evolutionary conservation. Usually, a conserved domain contains one or more motifs. Many proteins consist of a single protein domain; whereas others contain several domains or include additional, non-globular parts, e.g. signal peptides in membrane and secreted proteins.

The EAP was checked for its protein domains using InterPro Scan (Integrated Resource of Protein Families, Domains and Sites; http://www.ebi.ac.uk/interpro). The analysis showed the presence of highly conserved domain of Peptidase _S8 subtilase family (pfam 00082) at 120 to 377 aa positions and the Subtilisin N-terminal region (pfam 05922) at 40 - 108 aa positions of EAP respectively. Subtilisin N-terminal region is the domain family usually found at the N-terminus of a number of subtilisins, which get cleaved prior to activation of the enzyme. In addition to this catalytic triad domains characteristic of subtilisin type serine protease was also reported by InterPro Scan at 138 - 157 (Aspartic acid active site domain), 173 - 186 (Histidine active site domain) and 329 - 345 (Serine active site domain) respectively.

4.10.2.3 Secondary structure analysis

As an initial attempt to compare the secondary structure and folding patterns of the predicted precursor protein EAP with four other selected proteases of fungal origin, the hydrophobicity of the proteins was determined by Kyte and Doolittle algorithm (Kyte & Doolittle 1982, Hall 1999) (Fig 4.24).

Later, the secondary structure prediction of the deduced amino acid sequence of *E. album* alkaline proteinase (EAP) was done with "Antheprot" (http://antheprot-pbil.icbp.fr). In this program, EAP sequence was analyzed using four different methods: GOR I (Garnier et al. 1978) GOR II (Gibrat et al. 1987) PHD (Rost & Sander 1993) and by the SOPMA (Geourjon & Deléage 1994).

The graphical representations of the analysis are presented in the figures 4.25a, 4.26a, 4.27a, and 4.28a respectively. Helices are represented as blue colored blocks and the sheets are represented as green colored blocks. Similarly the turns are represented in grey and coils in saffron colors. The percentage of each of the structural features as predicted by the different methods is given in figures 4.25b, 4.26b, 4.27b, and 4.28b respectively.

Figure 4.24: Hydrophobicity plot of Proteinase R and K of *Tritirarchium* (Trit R & TritK), Serine protease of *Fusarium* sp (FusS), pr1 of *Cordyceps brongniartii* (Cob) and *Engyodontium album* alkaline serine protease (EAP) determined Kyte and Doolittle algorithm

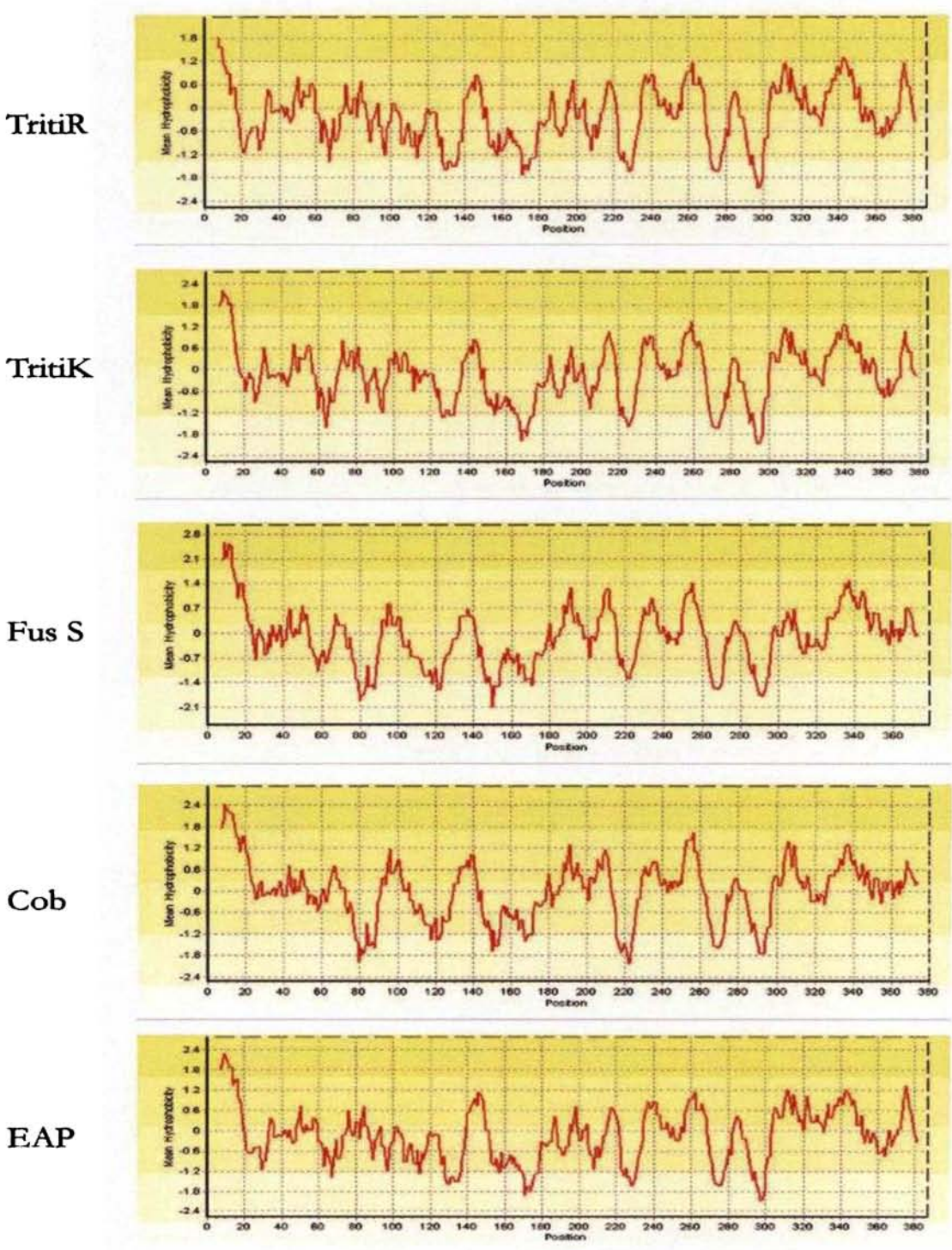


Figure 4.25a: Secondary structure prediction of the mature EAP using GOR I method

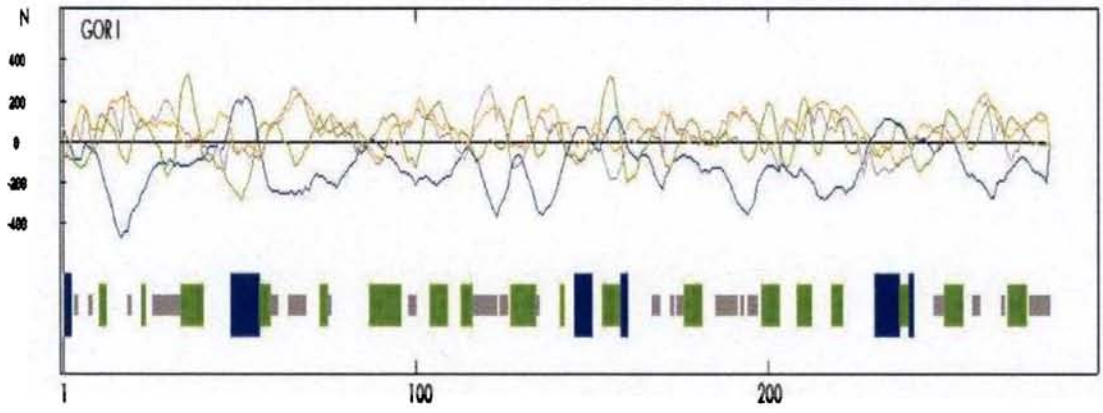


Figure 4.25b: Percentage representation of structural features of the mature EAP obtained by the GOR I method

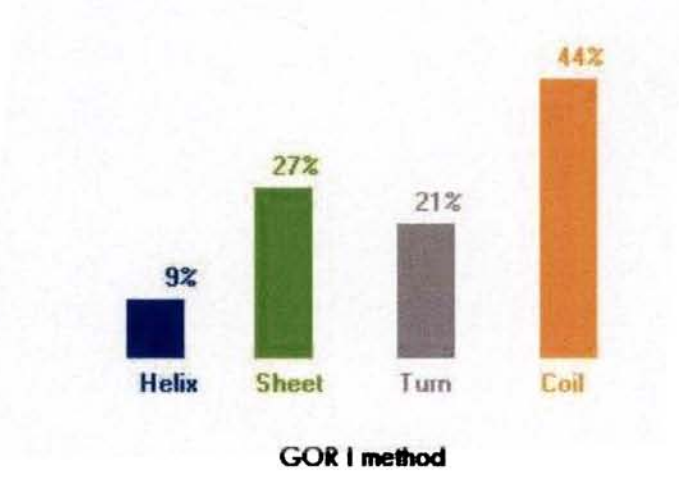


Figure 4.26a: Secondary structure prediction of the mature EAP using GOR II method

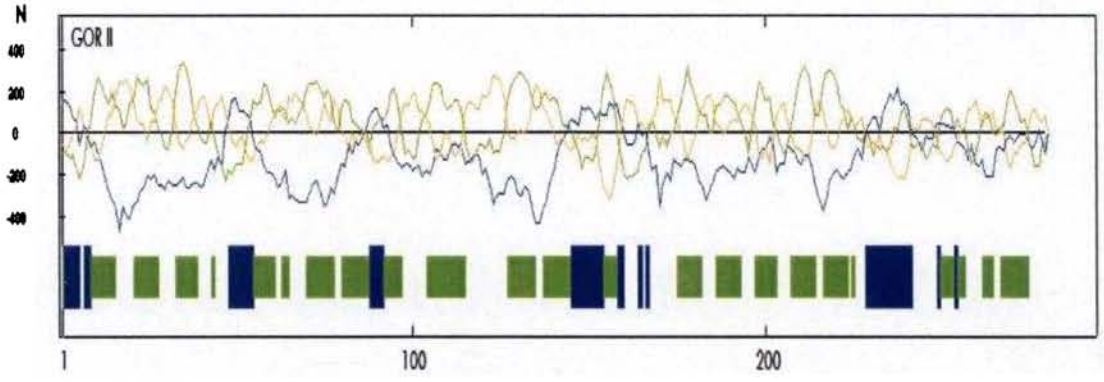


Figure 4.26b: Percentage representation of structural features of the mature EAP obtained by the GOR II method

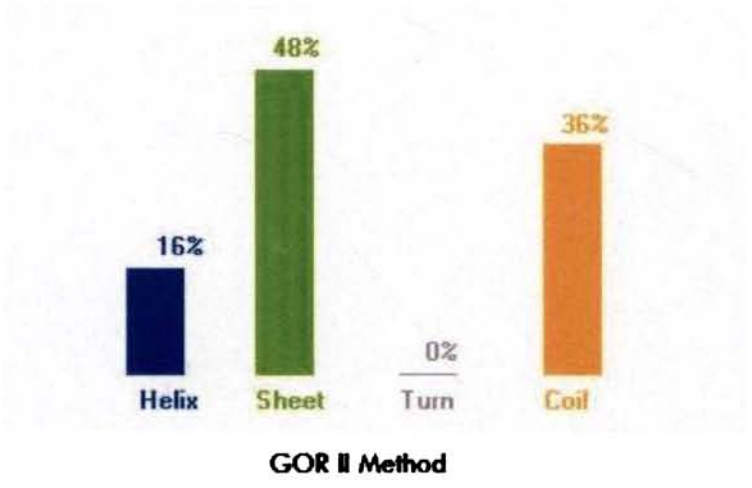


Figure 4.27a: Secondary structure prediction of the mature EAP using PHD method

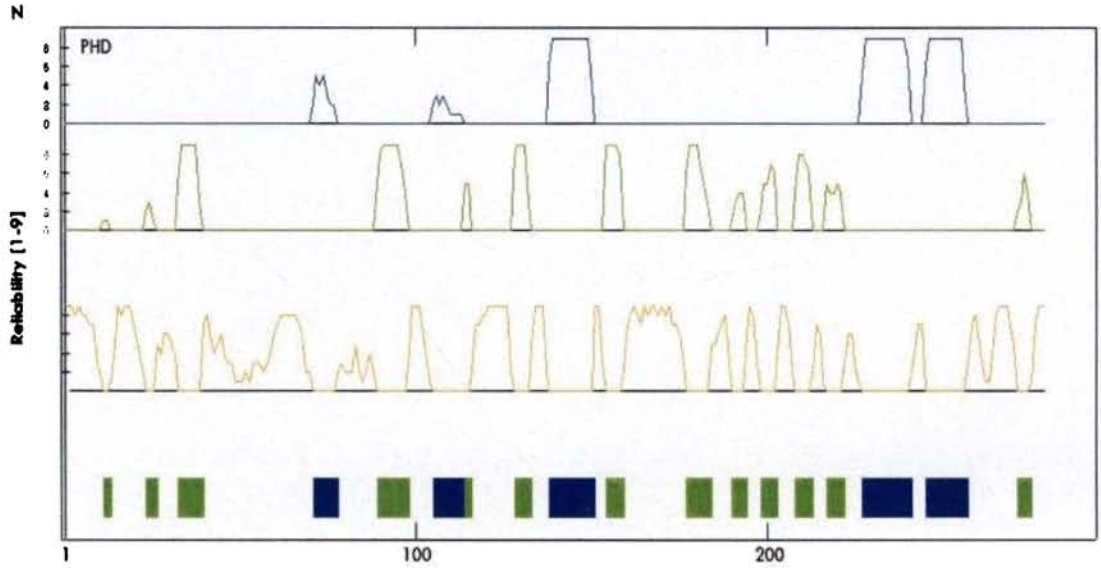


Figure 4.27b: Percentage representation of structural features of the mature EAP obtained by the PHD method

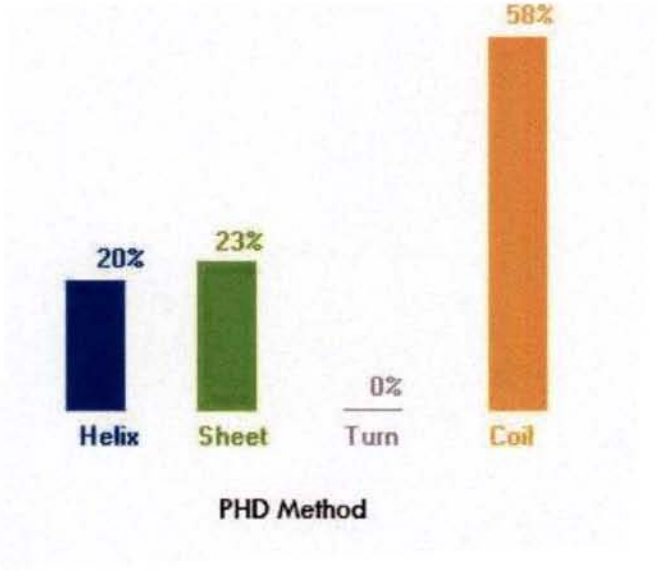


Figure 4.28a: Secondary structure prediction of the mature EAP using SOPMA method

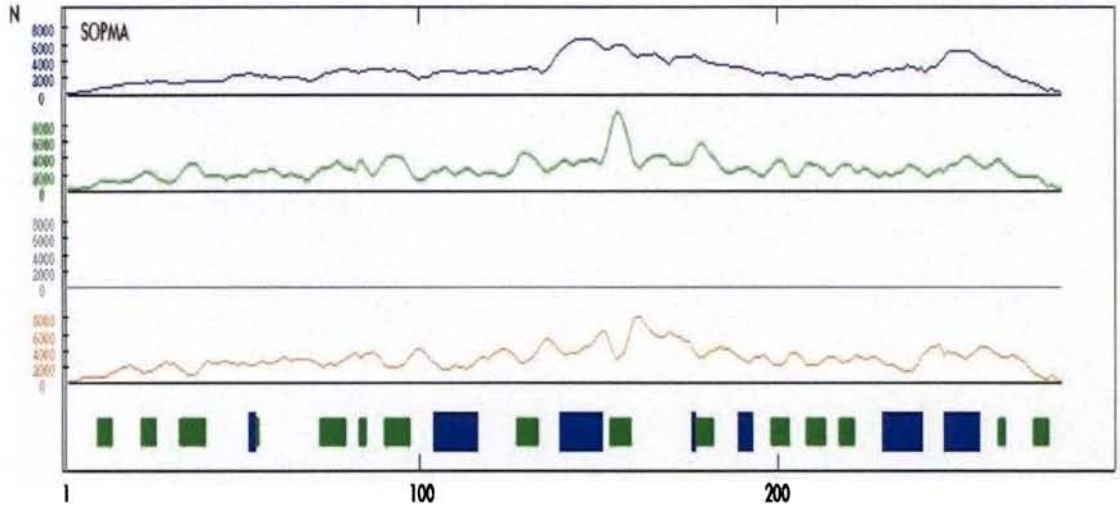
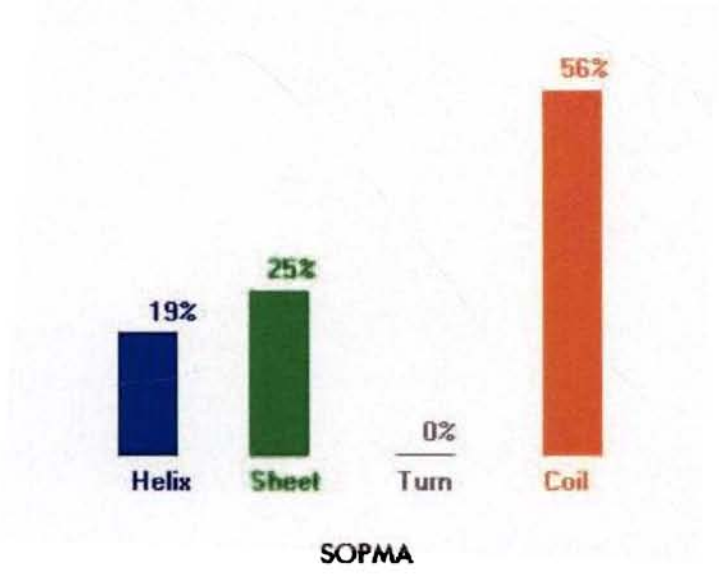
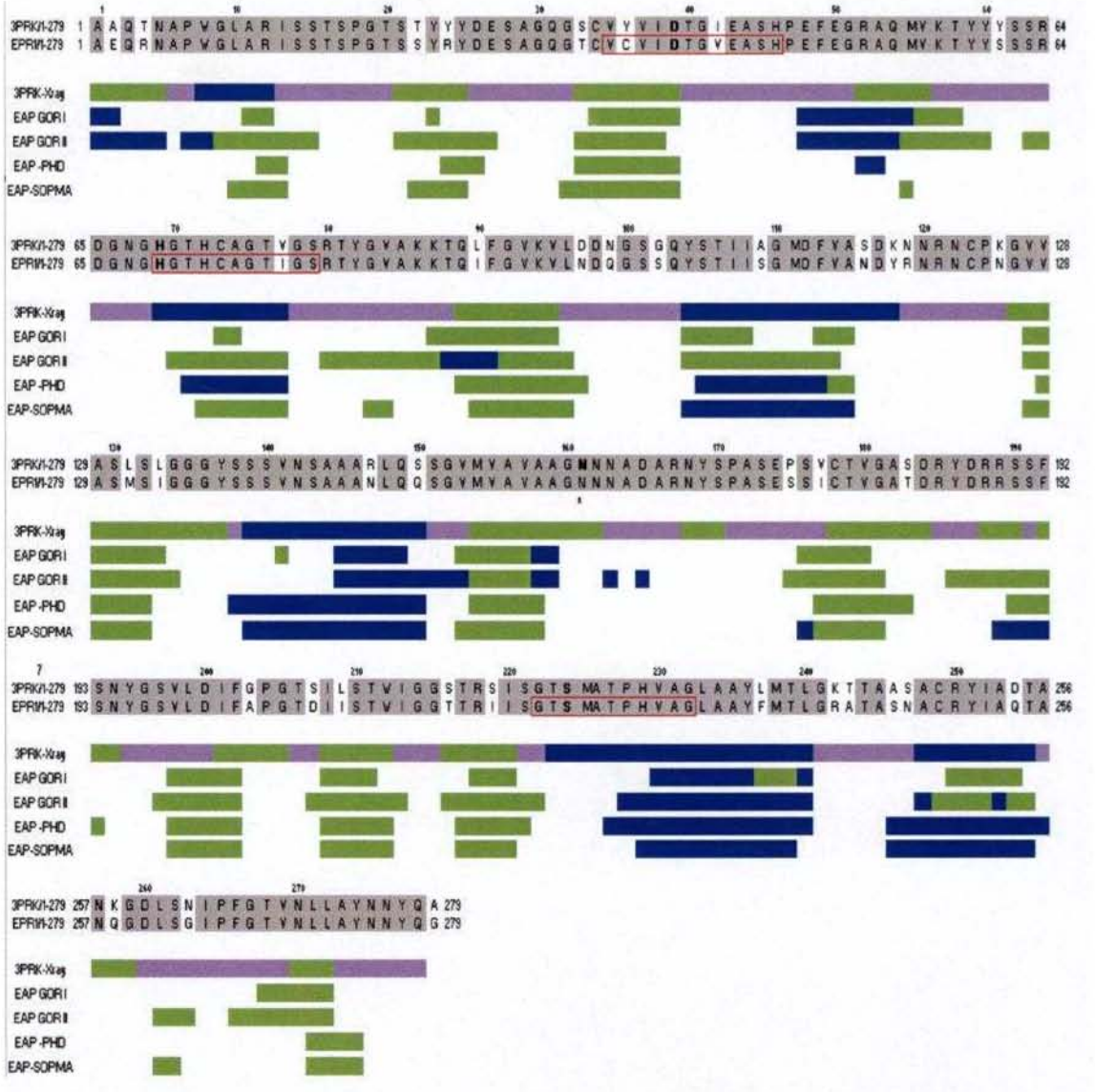


Figure 4.28b: Percentage representation of structural features of the mature EAP obtained by the SOPMA method



The secondary structure of EAP predicted by the above four methods (GOR I, GOR II, PHD and SOPMA) are compared with the actual secondary structure elements of Proteinase K determined by X ray crystallography (Wolf et al. 1991) (Fig .4.29).

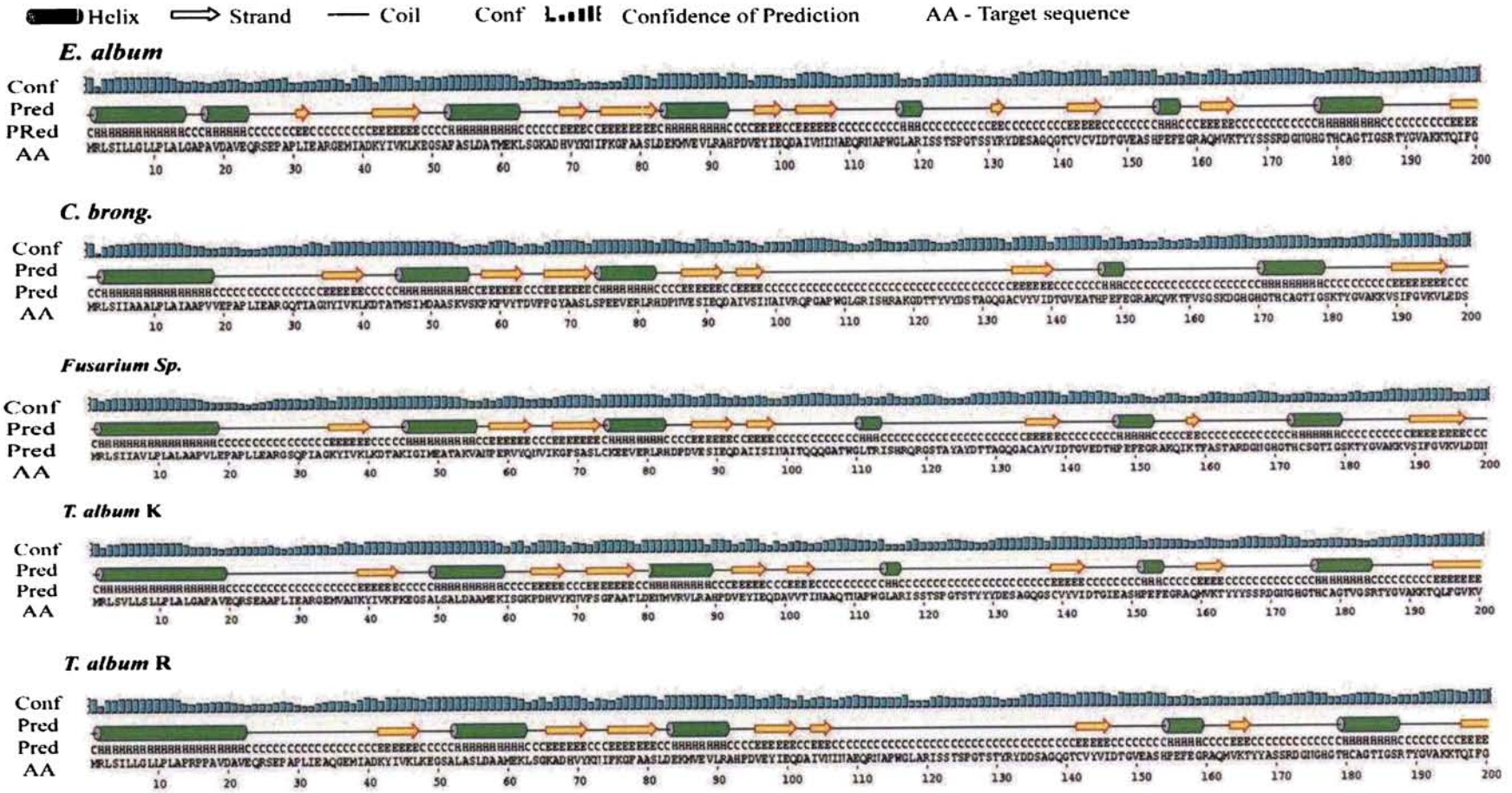
Figure 4.29: Alignment of EAP with proteinase K sequence obtained by X-ray crystallography. The sequence blocks, which are identical, are shaded grey. Regions corresponding to the catalytic triad for subtilases are boxed in red. Region 35-46 contain Active site Aspartate residue (D at 39), Region 69-79 contain Active site Histidine residue (H at 69) and Region 222-232 contain Active site Serine residue (S at 224). Catalytic residues are in shown in bold



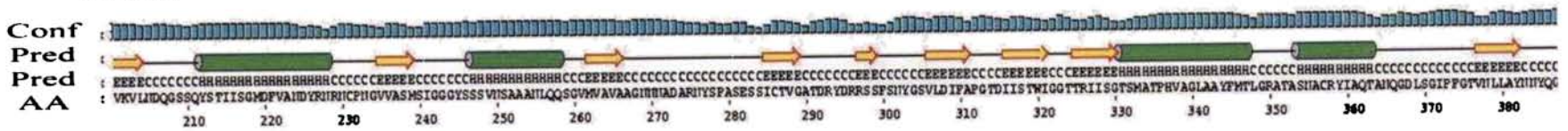
Asp 161 in Proteinase K which forms the "oxanion hole" along with the residues of the catalytic triad is indicated by a * below the alignment. Immediately below the alignment, is given the graphical representation of actual secondary structure elements of Proteinase K as determined by X-ray crystallography (Wolf et al. 1991). Following lines contain graphical representations of secondary structure features of EAP as determined by analysis using various methods in the order GOR I, GOR II, PHD and SOPMA. Helices are indicated by blue, sheets by green and coils are indicated by violet (coils are indicated by violet in 3PRK only).

Further, comparative secondary structure analysis of selected fungal serine protease sequences obtained by PSI-PRED algorithm are given in the Fig 4.30. All the important motifs taking part in the catalysis have similar secondary structure.

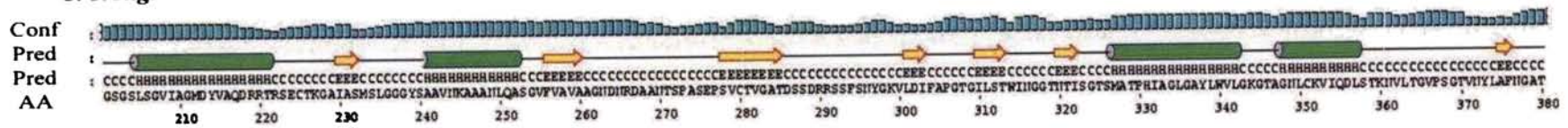
Figure 4.30: Comparative analysis of the secondary structure of EAP (DQ268654) with subtilisin like serine protease sequences of *T. album prot K* (P06873), *T. album prot R* (P23653), *C. brongniartii* Prl (AAR97273) and *Fusarium* sp.(strain S-19-5) Alp (AAC60571.2). The deduced precursor protein sequences are considered for the analysis. The secondary structure analysis was carried out in PSIPRED protein structure prediction server (http://bioinf.cs.ucl.ac.uk/psipred)



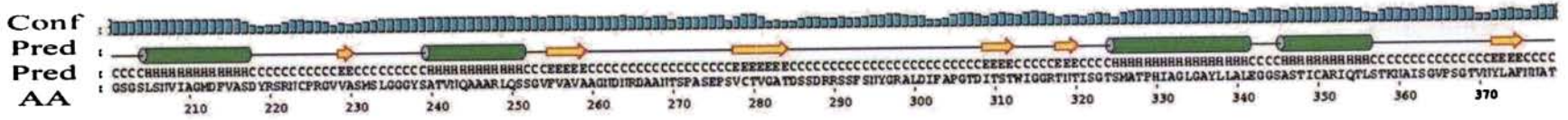
E. album



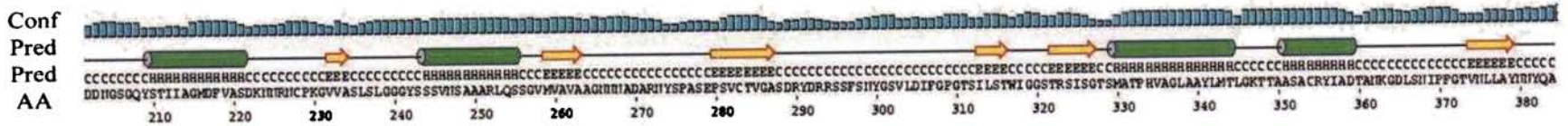
C. brong.



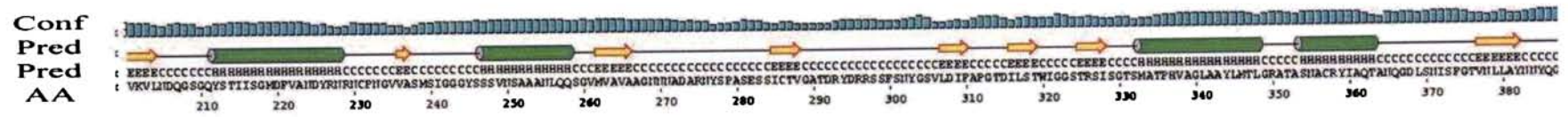
***Fisarium* Sp**



***T. album* K**



***T. album* R**



4.10.2.4 Modelling of EAP

Homology modeling of the *Engyodontium album* alkaline protease (EAP) was performed at the Swiss Modeller web server using "Swiss PDB viewer" (DeepView). *Tritirarchium album* serine protease "Proteinase K" (Betzel et al. 2001) (Fig. 4.31a) was used as the template for modeling.

Figure 4.31a & b: The cartoon representations of proteinase K (PRK1) from *T. album* and the modeled structure of EAP

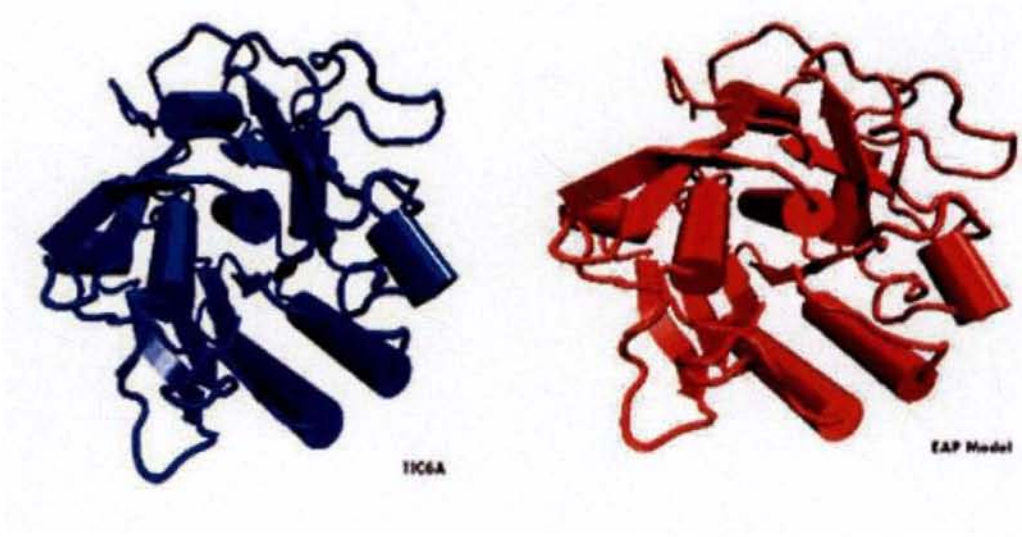


Figure 4.31a **Figure 4.31b**

The model representations were rendered using the molecular graphics tool VMD (http://www.ks.uiuc.edu/Research/vmd/) (Fig.4.31b). The atomic co-ordinates for PRK1 was obtained from EXPDB template 1IC6A derived from 1ic6.pdb (Betzel et al. 2001).

Sequence alignments and homology model of EAP shows that the protein is almost identical to the *T. album* proteinase K. The predicted 3D model of EAP was analyzed using PROCHEK (Laskowski et al. 1993) for checking the stereochemical quality of the model.

To assess the structure, the program makes use of a number of parameters that have been found to be good indicators of stereo chemical quality. Based on these parameters, it produces a number of plots, which describe the quality of the structure. The various parameters include: Ramachandran plot for the aa residues, the main chain and side chain parameters, the residue properties, the main chain bond length and bond angle distributions, RMS distance from planarity and Distorted geometry plots.

As per the standards of the PROCHEK, a good quality model would be expected to have over 90% in the most favored regions. From the plot (Fig 4.32) it is clear that 89.5% of the residues (excluding Gly and Pro) had Phi-Psi angles in the most favored region and 10.5% of residues in the allowed region while none of the residues were observed in the disallowed regions of the plot or even in the generously allowed regions. Hence this plot result assures that the model is of good quality. In Ramachandran plot for Glycine (32 residues) & Proline (8 residues), all the residues are being observed to have favourable conformations except one (^{78}Gly) (Fig 4.33).

Figure 4.32: Ramachandran plot for the aa residues in EAP (except Gly and Pro)
X axis –Phi angle (degrees), Y axis –Psi angle (degrees), Red - most favored region (A, B, L), Yellow – allowed region (a, b, l, p), Cream- generously allowed regions (~a, ~b, ~l, ~p), Black squares – residues, White – Disallowed regions, Black triangles – Gly residues

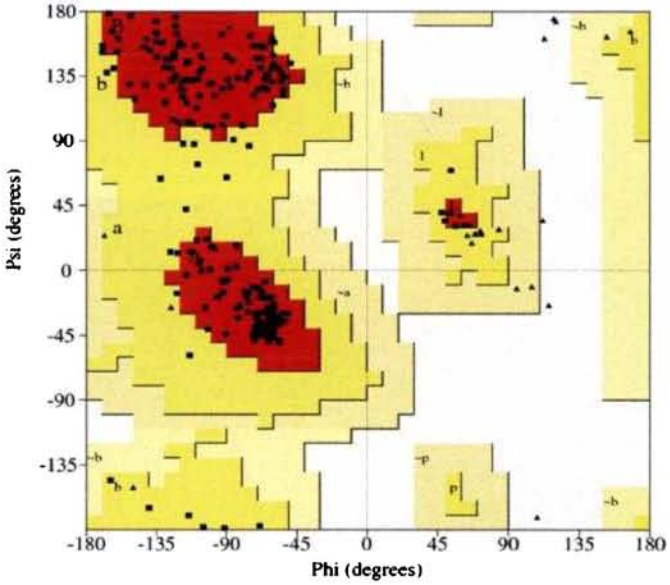
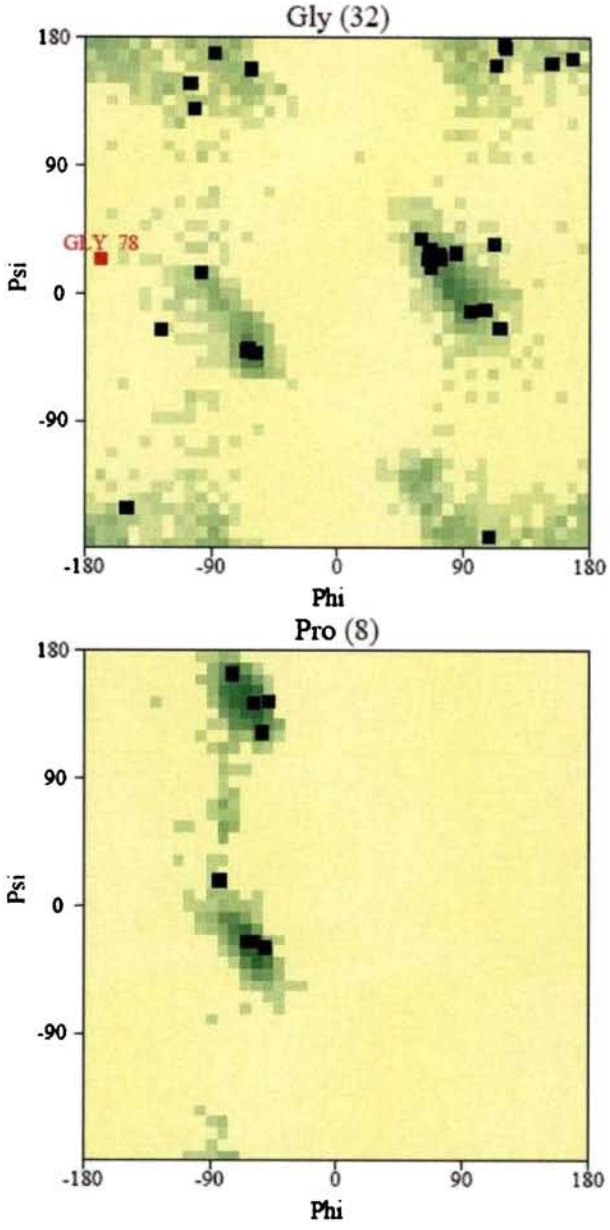


Figure 4.33. Ramachandran plots for Glycine & Proline: Numbers of Glycine and Proline residues are shown in brackets. Those in unfavorable conformations are labeled in Red



The plot produced for only those amino acids (total 130 residues) having the Chi1-Chi2 torsion angles also indicated that all of them belong to the favorably shaded regions of the plot. Main chain parameters of the model were all within acceptable limits, and sometimes even better than the typical values while the plots for the side chain parameters indicated that all of them were better than the typical values used.

In the context of the main chain bond lengths, the plot obtained for the model revealed that 100% of them were within the acceptable limit but when the main chain bond angle values were plotted, 94.3% of them were within the acceptable limits and the rest 5.7% were highlighted as 'unusual'. The plot for the planar groups also revealed some deviations (20.7%) where 79.3% were found to be within the limits. Thus considering the overall positive results obtained for different parameters plotted, PROCHEK proposed that this model is of good quality.

4.10.2.5 Finding disulphide linkages in EAP

From the primary sequence the total number of cysteine residues in the protein EAP was found to be six. The cysteines were located at positions 34, 36, 73, 123, 178 and 249. From the 3D model of the protein it is observed that the cysteines 34 and 123 come close enough to suspect a SS linkage. Similar was the case with cysteines 178 and 249 (Fig 4.34). Prediction of disulphide linkages was performed using "DISULFIND web server" (Ceroni et al. 2006) at http://disulpfind.dsi.unifi.it and these suspected SS bonds were predicted with maximum confidence (fig 4.35). Figure 4.36 shows the two predicted SS linkages in EAP three dimensional model.

Figure 4.34: Three dimentional Stereo image of the protein showing all of the six cycteine residues in EAP. Four of them (^{34}Cys -^{123}Cys) and (^{178}Cys - ^{249}Cys) are found to be close enough to suspect the formation of two SS linkages.

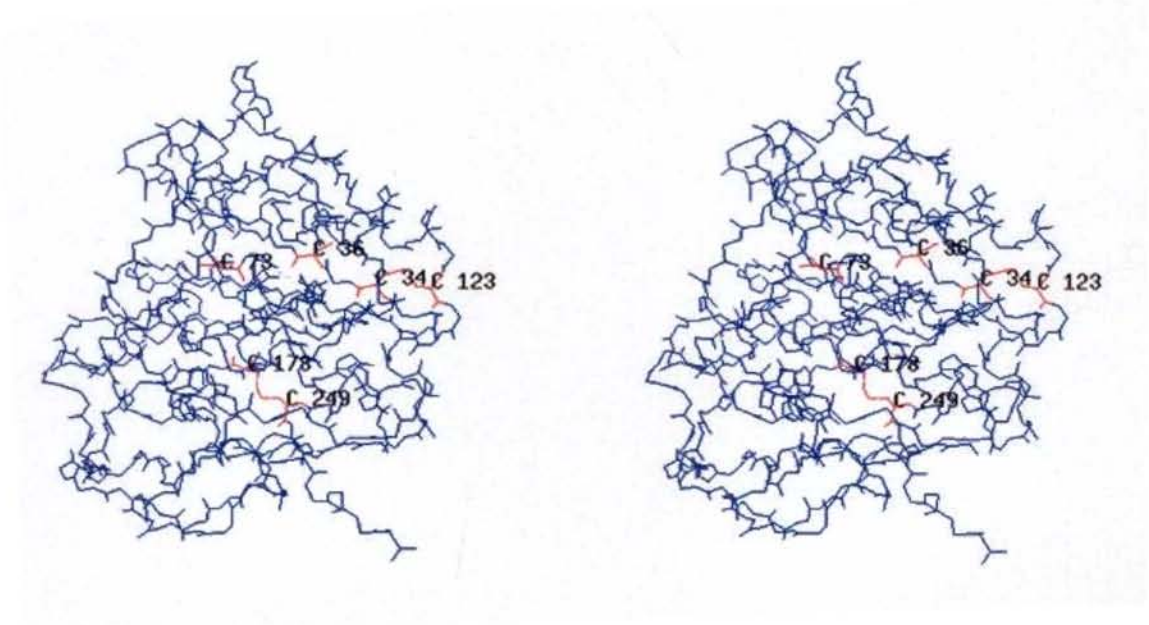
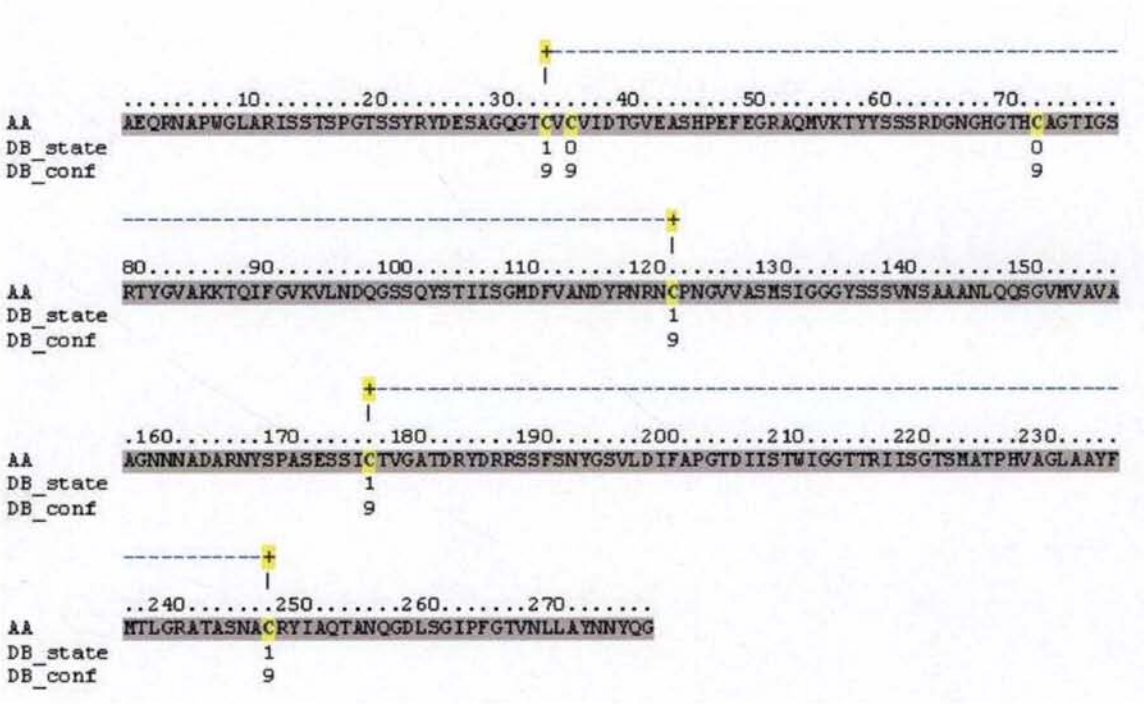


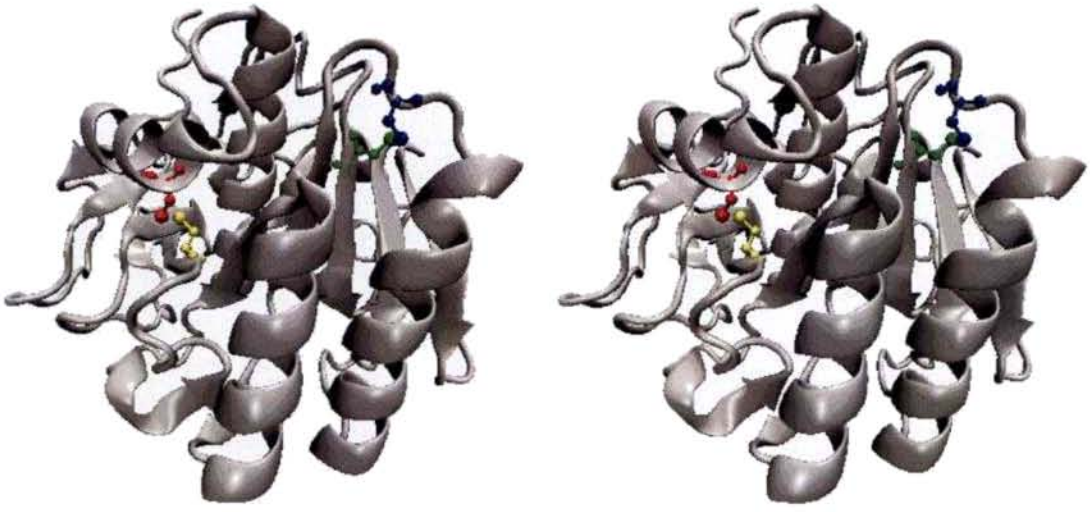
Figure 4.35: Results for DISULFIND analysis
AA: amino acid sequence, DB_state: predicted disulfide bonding state (1=disulfide bonded, 0 = not disulfide bonded), DB_conf: confidence of disulfide bonding state prediction (0=low to 9=high), DB_bond : position in sequence of a pair of cysteines predicted to be forming a disulfide bridge. Conn_conf: confidence of connectivity assignment given the predicted disulfide bonding state (real value in [0, 1])




```
DB_bond    bond (34,123)
DB_bond    bond (178,249)

Conn_conf 0.795061
```

Figure 4.36: Three-dimensional stereo image of EAP showing the two-disulphide bonds visualized using VMD, the molecular graphics tool. The protein structure is represented as a cartoon. The cysteines predicted to form disulphide linkages are indicated as ball and stick models and colored differently to identify them. ^{34}Cys - green ^{123}Cys - blue ^{178}Cys - yellow ^{249}Cys - red



4.10.2.6 Finding Ca^{++} binding sites in EAP

The 3D model for *E. album* proteinase (EAP) was also analyzed for putative "Ca" binding sites using the program GGv1.0 at the website http://www.chemistry.gsu.edu/faculty/Yang/GG.html (Deng et al. 2006). Out of the 8 putative pseudo Ca ion locations detected, closely located points were refined to obtain four locations, which could be nearer to the locations. The merged predictions are given below.

Ca001				
Oxygen atom Ids	226	243	245	650
Residues	ALA11	SER14	SER14	SER22
Ca coordinates	42.775	49.516	14.445	

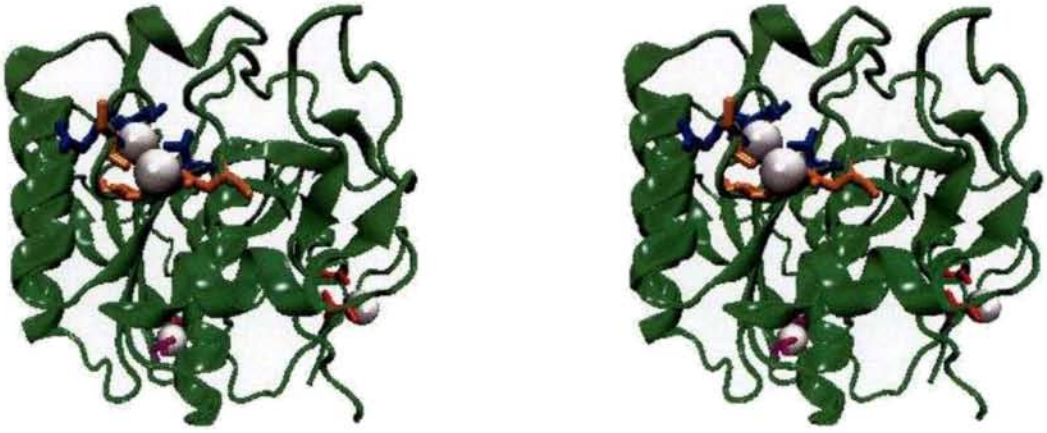
Ca002				
Oxygen atom Ids	1272	1284	1466	1467
Residues	GLY30	THR33	THR33	THR88
Ca coordinates	54.039	33.147	24.912	
Ca-O distance	2.455			

Ca003				
Oxygen atom Ids	1245	1252	1263	1300
Residues	PRO171	ALA172	GLU174	THR179
Ca coordinates	51.095	35.091	29.473	
Ca-O distance	2.571			

Ca004				
Oxygen atom Ids	84	108	110	157
Residues	SER175	ILE177	ASP200	ASP200
Ca coordinates	62.959	55.185	14.580	
Ca-O distance	2.991			

Based on above prediction, the putative four Ca binding sites were visualized in EAP three-dimensional model by editing the model coordinates file of EAP to include the coordinates of the predicted Ca ions (Fig.4.37).

Figure 4.37: The three dimensional stereo image of EAP showing four putative Ca binding sites: The protein chain is represented as cartoon with green color. The predicted Ca binding sites are shown with Ca-ions occupying the positions so as to form bonds with oxygen atoms of the residues that form the Ca binding site. Residues forming Ca 001 is indicated in red, Ca 002 in magenda, Ca 003 in blue and Ca 004 in orange. The Ca ions are shown as grey spheres



4.10.2.7 Active Site and Peptide binding regions of EAP

To understand more about the active site and the peptide-binding region, these regions were modeled and visualized using Deepview and VMD. Figure 4.38a and 4.38b shows the stereo view of the Catalytic triad of Proteinase K (Wolf et al. 1991) and EAP overlapped.

Figure 4.38a & b: The stereo view of the superimposition of the Catalytic triad EAP over that of Proteinase K. Red color shows the residues in 3PRK while blue indicates those from EAP. The residues forming the Subtilase_ Asp (35-46), Subtilase_His (69-79) and Subtilase_ Ser (22-232) and the Asn161 residue, which forms part of the Oxanion hole in Proteinase K also shown.

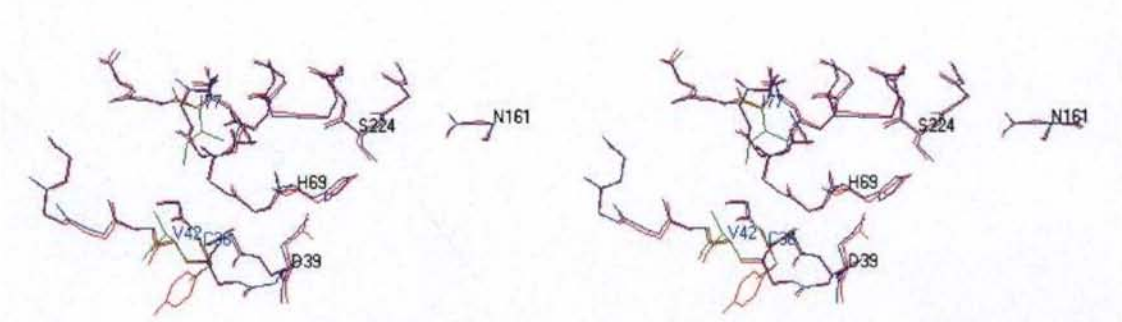


Figure 4.38a.

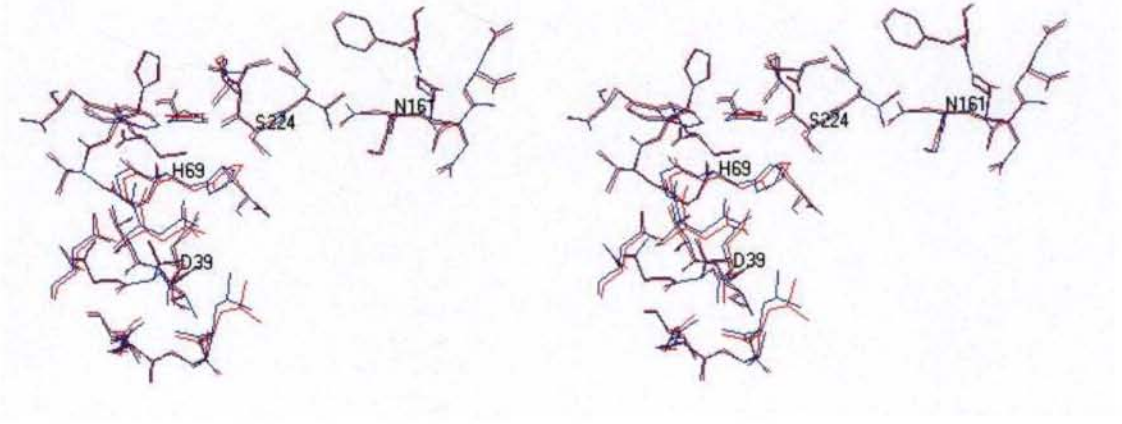


Figure 4.38b.

The catalytic triad overlaps almost exactly as is the Asn 161 residue, indicating the conservation of the active site in EAP. The substitutions Y-C at position 36, I-V at position 42 and V-I at position 77 have not affected the overall topology of the active site in EAP. While, there was no substitutions in the residues closer (within 4 A^{o}) to the catalytic triad as can be observed from figure 4.38b.

Substrate recognition site in proteinase K (PRK1) is formed by two peptide chains at 99-104 and 132-136 respectively. These peptide chains are oriented approximately parallel and are directly connected only by a hydrogen bond formed between ^{104}Tyr and ^{136}Gly (Fig 4.39a). The arrangement of the catalytic triad and the substrate-binding domain of EAP are almost exactly similar to that of PRK1 (Fig 4.39b).

Figure 4.39a & b: Representation of the substrate recognition site in proteinase K and EAP

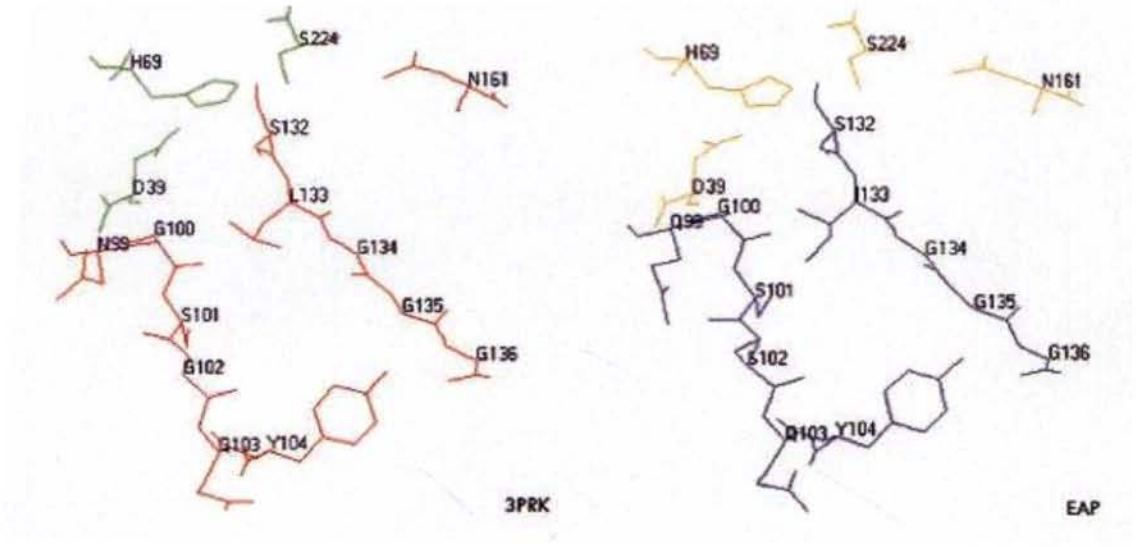


Fig 4.39a **Fig 4.39b**

The Asn at position 99 in PRK1 is replaced by Glu and the Gly at position 102 is replaced by Ser in EAP, but as can be observed from the figures (4.39a & 4.39b) above this substitution do not place any significant changes in the conformation of the substrate binding domain.

Figures 4.40a and 4.40b show another representation of the active site and substrate binding domains of PRK1 and EAP respectively. Here PRK 1 (1IC6.pdb) is shown without the binding site being occupied by inhibitor unlike in 3PRK, which is X-ray crystallographic structure of proteinase K bound to an inhibitor (Wolf et al. 1991), so that the binding sites may be compared.

Figure 4.40a & b: Representation of the substrate recognition site in proteinase K and EAP: The proteins are shown as surface representations with the hydrophobic core tinted red and the protein surface blue with a color gradation depending on the hydrophobicity. The active site residues are shown as ball and stick models. Coloration scheme for residues is: Blue -D39, Green -H69, Red -S224, Cyan -N161. The substrate binding surfaces formed by residues 99-104 and 132-136 are colored green and yellow respectively.

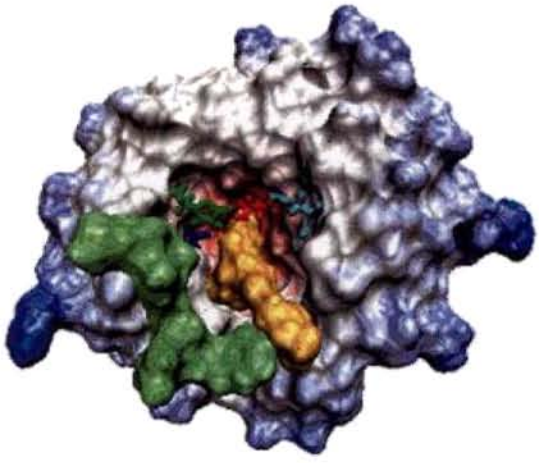


Figure 4.40a

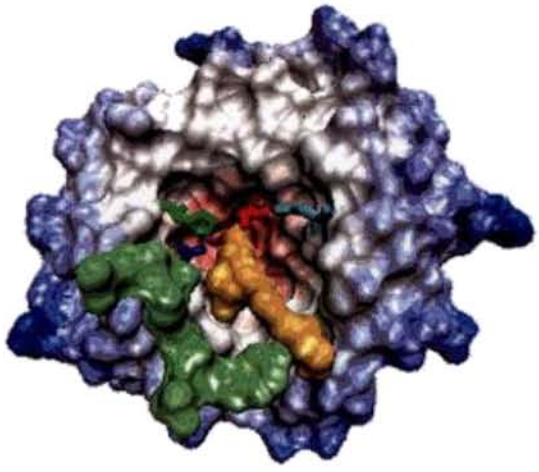


Figure 4.40b

It is known that the binding of substrate does not produce any systemic and significant conformational changes in the PRK1 protein except for the binding region (Wolf et al. 1991). Studies using inhibitor Methoxy Succinyl –Ala-Ala-Pro-Ala-chloromethyl ketone have shown that the binding of this inhibitor at the substrate-binding pocket in PRK1 results in the shifting of strand 100-104 1A° to open up the binding cleft. (Fig.4.41a & 4.41b)

Figure 4.41 a & b: Three dimensional representation of PRK1 substrate binding pocket: (a)PRK1 in the unbound state (1IC6.pdb) **(b)** PRK1 bound to the inhibitor Methoxysuccinyl – Ala-Ala-Pro-Ala-$COCH_2Cl$ (The inhibitor is hidden to expose the binding surface).Protein is represented as green ribbons. Strands 99-104 and 132-136 are colored yellow and the residues are labeled. The active site residues are represented in lines and are labeled

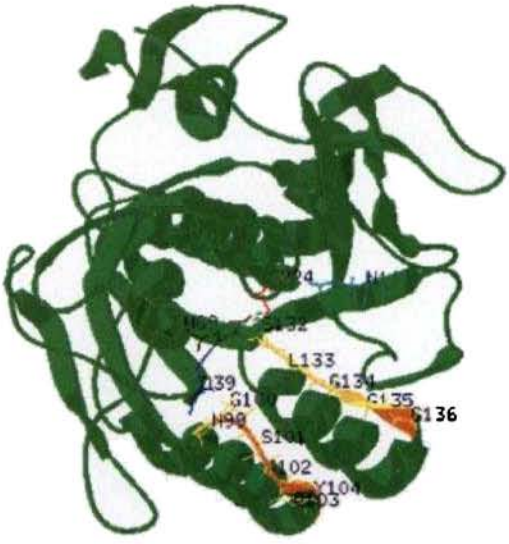


Figure 4.41a

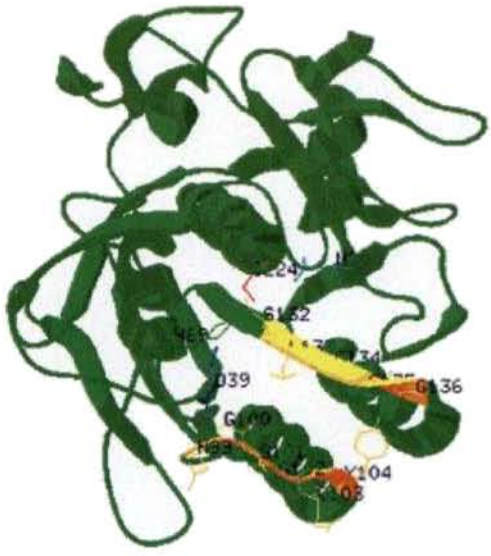


Figure 4.41b

The position of strand 132-136 is not significantly affected since it is tightly anchored to the bulk of proteinase K structure. The N terminus of this strand is linked to a nine-stranded beta-sheet through strand βII4 with the aminoacids 168-170 (Betzel et al. 1988b). The segment 9(100-104) although connected at its N terminal with strand βII3 and at its C terminus to alpha helix a3 (Betzel et al. 1988b) is not involved in secondary structure hydrogen bonding. The strand is exposed to solvent and can move according to stearic requirements.

It is speculated that EAP shares the same features with PRK1 and the conformations adopted by the catalytic site and the substrate binding cleft might be similar to that of PRK1 since the former is almost identical in its structural features at the catalytic core and the substrate-binding region. The model of EAP was superimposed on

the structure of PRK1 bound to N-Ac-Pro-Ala-Pro-Phe-DAla-Ala-Ala-Ala-NH (2) (1PFG.pdb) (http://www.rcsb.org/pdb/) to identify features of the EAP substrate binding cleft and the positions of the subsites where the substrate is bound.

It is well known from the crystallographic studies on Proteinase K (Betzel et al. 1988a) that the substrate recognition site is formed by two peptide chains ^{99}Asn - ^{104}Tyr and ^{132}Ser-^{136}Gly. EAP has an exactly similar structure for the substrate-binding site but with Asn-99 replaced with Gly and Gly-102 with Ser and Leu-133 replaced by Ile. Neither of these substitutions alters the topology of the substrate-binding region significantly. The cleavage site for a protein (substrate) involve residues both N-terminal (numbered P3, P2, P1) and the C-terminal (numbered P1', P2', P3'...) to the scissile bond with the cleavage site for a peptide defined as .....P3-P2-P1-P1'-P2'-P3' ....., and the cleavage occurs between the P1 and P1' residues (Schechter & Berger 1967). The corresponding sites on the enzyme are called .....S3-S2-S1-S1'-S2'-S3'..... respectively.

Superimposing of the octapeptide, which is in the hydrolyzed form (already cleaved at the peptide bond between Phe284 and Ala285) shows that the inhibitor can occupy the entire substrate-binding region. It identifies the S1 specificity subsite, which corresponds to the P1 residue of the octapeptide (Phe284). It follows that the regions corresponding to the N terminal region of the octa peptide, i.e. the residues from 281-283 forms the P4, P3 and P2 sites and defines the specificity subsites S4, S3 and S2 respectively. Prime regions S1', S2', S3', and S4' are occupied by the C terminal tetrapeptide D Ala-Ala-Ala-Ala-NH_2. Figure 4.42 and 4.43 shows the substrate-binding region of EAP with the hydrolyzed octapeptide superimposed and surface representation respectively.

Figure 4.42: Representation of the substrate binding region of EAP superimposed with the hydrolysed octapeptide. The substrate binding residues and the catalytic triad of EAP are shown in blue color. This is super imposed on corresponding residues of PRK1 (1PFG.pdb)(http://www.rcsb.org/pdb/) shown in light blue. Octapeptide residue is shown in red. The octapeptide inhibitor is cleaved at the scissile bond between ^{284}Phe and ^{285}Ala, and has moved apart. The residues with respect to the scissile bond are labeled as .... P1, P2, P3, P4 and P1', P2', P3' and P4'. Corresponding specificity subsites are labeled as S1, S2, S3, S4 and S1', S2', S3' and S4' respectively. The Cleavage site is indicated by arrow. The position of the cleaved peptide bond is indicated by a violet line connecting ^{284}F (at C-terminal end) and ^{285}A (at the N-terminal end). Hydrogen bonds are indicated as green dashed lines

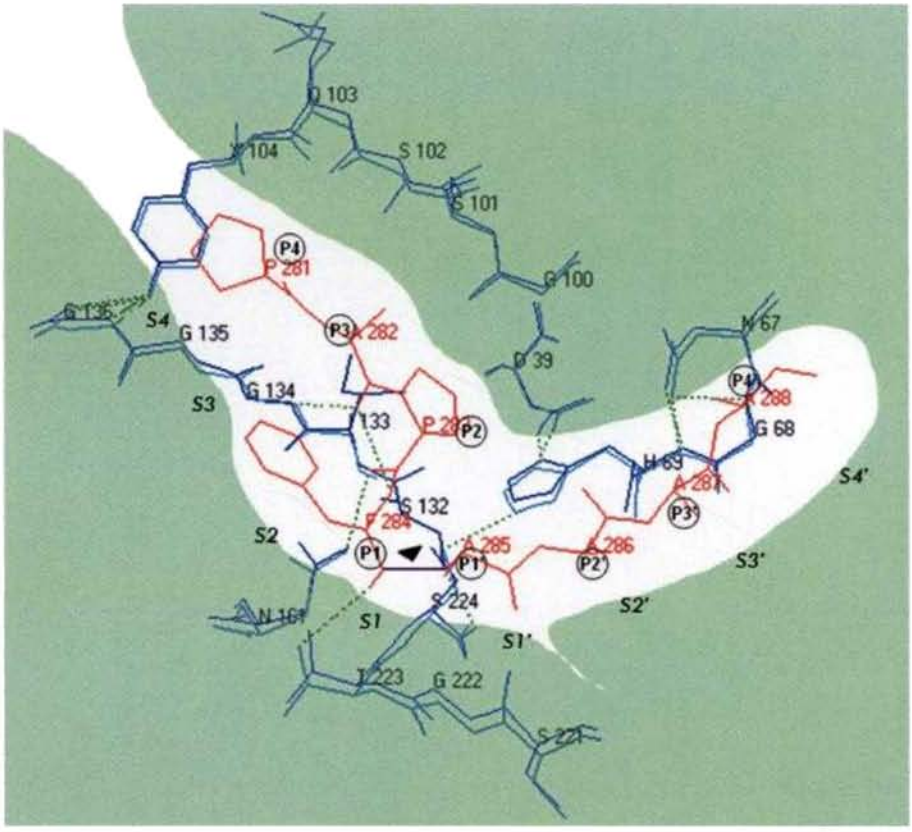
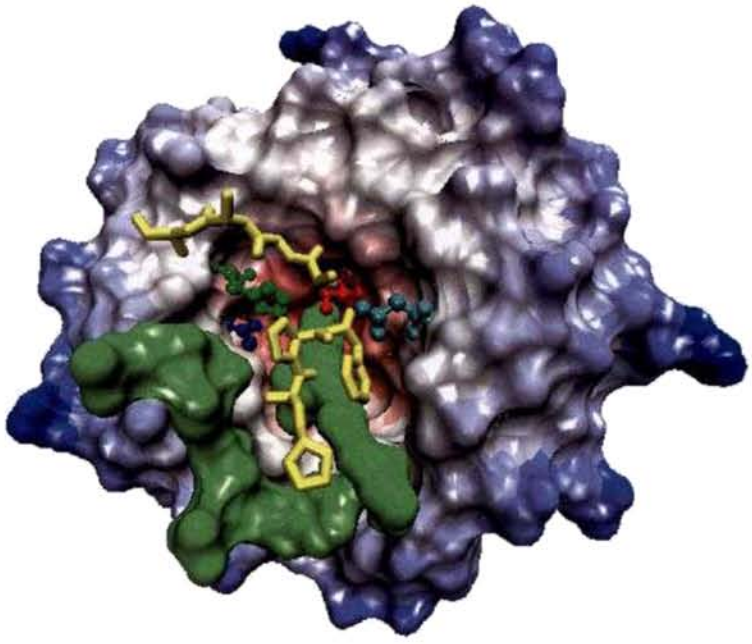


Figure 4.43: The surface representation of the substrate binding region of EAP bound with the inhibitor .The green regions indicate the substrate binding residues 99-104 (green glossy) and 132-136 (green diffuse). Catalytic residues are shown as ball and stick models where Asp39 is blue, His69 is green, Ser 224 is red and Asn 161 is cyan. The N terminal tetrapeptide and the C-terminal tetrapeptide of the cleaved octa-peptide inhibitor are shown as licorice models in yellow color



Based on homology modeling and sequence alignment, it can be concluded that in EAP and Proteinase K, the substrate binding is predominantly determined by the binding of the PI and P4 residues in two pockets or clefts on either side of the backbone strand 132 - 135. The two sides of the P I cleft are formed by the backbone segments 132 - 135 and 158 -162. The P4 pocket, between the strands 101-104 and 133-135, is lined with hydrophobic side chains, i.e. residues 96, 104, 107 and 126.

Chapter 5

DISCUSSION

In the present study, initial attempts made to isolate the gene from the cDNA library constructed for *E. album* and screening of the cDNA library with the polyclonal antibody raised against the enzyme and also with prepared heterologous probe (*A. fumigatus* serine protease gene) failed to yield any positive clone in spite of genuine efforts. A similar case was reported earlier in the case of *Trichophyton rubrum* where the cDNA library screened with high titer pool of antibodies, failed to identify any positive clones (Woodfolk et al. 1998). The success of the screening of the cDNA library with the antibody probe strongly depends on the specificity and titer of the antibody used. Further, the signal/ background ratio obtained during the antibody screening was frequently low, a situation which would have led to the non-identification of truly positive clones (Marin et al. 2003).

Negative results in the case of cDNA library screening with heterologous probe was also reported earlier where the human islet cDNA library was screened using the mouse IGRP probe (Martin et al. 2001). Later, the human IGRP (Islet specific Glucose-6-phosphatase catalytic subunit-related protein) gene sequence was established using PCR technique. Another similar case was reported by Caballero et al. (1997), where attempt was made to isolate the urate oxidase gene from the chickpea cDNA library with a heterologous probe, which showed problems like cross hybridization and high stringency background with the vector used for cloning.

In the present study during the course of isolation of the full gene encoding alkaline serine protease from *E. album* (*Eap*), initially the partial coding region was successfully isolated using the degenerate primer pair (ASP1) designed from the conserved regions of other fungal subtilase gene sequences available in the literature.

This 878bp PCR product, cloned in pMOS vector was used as the homologous probe sequence for further isolation of full coding region.

In fact, the degenerate primers used in protease gene isolation studies were designed based on the N-terminal amino acid sequence information (Wilson & Barbara 2005) or on the internal peptide sequence information obtained for the protein (Gerem´ıa et al. 1993, Faraco et al. 2005). When these informations are not available, then primers are designed based on the highly conserved amino acid sequences in the protein. In *Aphanomyces astaci,* the protease gene-specific probe was constructed by PCR application using degenerated primers synthesized according to the highly conserved amino acid sequences of several known subtilisin like serine protease genes (Bangyeekhun et al. 2001). Further, the highly conserved amino acid stretch of subtilisin like serine proteases were used to design degenerate oligo-probes which were then directly used for screening the libraries to fish out the gene (Samal et al. 1989, Kim et al. 1999).

The Southern analysis of fungus genome using the homologous probe (878bp PCR fragment) revealed that there is only one copy of *Eap* gene in the *E. album* genome. Similar results were obtained in the southern blotting experiments for *Magnaporthe grisea* vacuolar serine protease gene SPM1 (Fukiya et al. 2002), VCP1 gene in *Pochonia chlamydosporia* (Morton et al. 2003) and protease gene in *P. brasiliensis* (Venancio et al. 2002).

Usually the genomic libraries are constructed adopting the shotgun cloning method, where a huge number of clones are to be screened, since the probability of occurrence of a single copy gene in a comparatively large genome of fungi will be of very less number. For instance the genomic DNA library for *Yarrowia lipolytica* from the *Sau*3 AI digested genomic DNA fragments (3-15kb) was constructed with an aim to screen for an alkaline serine protease gene (Davidow et al. 1987). In the present study in

order to reduce the number of candidates, which had to be screened, a gene-enriched library was constructed for *E. album* following the earlier report on *Tritirarchium* proteinase K gene (Gunkel & Gassen 1989). In the case of *T. album*, the sample that showed prominent signal in southern analysis was selected for the library construction, the sample was size fractionated and DNA in the corresponding region of the southern band was eluted and used for library construction. Similarly the cloning of the proteinase A gene from *Hansenula polymorpha* was reported (Bae et al. 2005), where the positive clone was isolated after screening a total of 500 clones. A gene-enriched partial genomic DNA library was reported to be constructed for *Streptomyces albogriseolus* serine protease also (Taguchi et al. 1995).

In the present study a sub genomic library for *E. album* was also constructed by ligating the genomic *Pst* I fragments (1.5-5kb) with the pUC18 plasmid vector, which resulted in 750 recombinant clones. pUC18 is reported to be used as the vector for genomic DNA library construction in several cases: for the cloning of pectate lyase gene from *Fusarium solani* (Guo et al. 1995) and for cloning the β lactam genes in the marine fungus, *Kallichroma tethys* (Kim et al. 2003).

The recombinants were further screened using non-radioactively labeled gene specific probe and ended up in the selection of 40 clones, which showed intense signals. These 40 clones were further screened by colony PCR technique using the ASP1 primers, which resulted in the selection of two positive clones, one with 1.4kb insert (pEap23) and other with 5kb (pEap288) insert. Nucleotide sequencing of the clone, pEap288 did not yield satisfactory results due to the high GC content of the insert, while in the case of the clone pEap23 the full coding region of the *E. album* alkaline serine protease gene (*Eap*) could be deduced. Based on the sequence information obtained, a set of primers was designed from the 5' and 3' UTR of the *Eap* and was successfully utilized to raise the cDNA of *Eap* gene.

Comparison of the nucleic acid sequence information (pEap23) and the deduced amino acid sequence with the known proteins from GenBank indicated that this gene could encode an extracellular protease that belongs to the subtilisin family of serine protease. This inference is supported by the data obtained from the enzyme inhibition studies of the enzyme (EAP) (Sreeja Chellappan 2005). Among the inhibitors tested, PMSF and aprotinin were observed to inhibit the protease activity of EAP. Since both of these are serine protease inhibitors, it was concluded that the enzyme (EAP) belongs to the serine protease group (Sreeja Chellappan 2005).

The subtilisin like serine proteases family is the second largest serine protease family characterized to date. Over 200 subtilases are presently known, more than 170 of which are with complete amino acid sequence (Siezen & Leunissen 1997). It is widespread, being found in eubacteria, archaebacteria, eukaryotes and viruses. The vast majority of the family is endopeptidases, although there is an exopeptidase and tripeptidyl peptidase. Structures have been determined for several members of the subtilisin family: they exploit the same catalytic triad as the chymotrypsins, although the residues occur in a different order (HDS in chymotrypsin and DHS in subtilisin), but the structures show no other similarity (Rawlings & Barrett 1993, 1994).

A total of 1,424bp of the *Eap* locus was sequenced from the genomic clone pEap23 and the sequence analysis revealed an open reading frame consisting of 1161bp. A putative ATG initiation triplet was found near the 5' end of the *E. album* genomic DNA. This triplet was preceded by a 93bp UTR. The bases surrounding the putative initiation site are: 5'- AAAATGC -3'. This sequence shows homology to the eukaryotic translation consensus sequence 5'-ANNAUGN-3' (Brown & Lithgow 1987) where the sequence 5'-ACCAUGG-3' is the best context for initiation (Kozak 1986). The sequence downstream of the ATG coded for a continuous peptide except for one small 65bp intron. This sequence continued in frame until a TTA stop triplet was encountered near the

3' end. The UTR that follows the stop triplet is AT-rich, a characteristic of genes from filamentous fungi (Gurr et al. 1988).

Comparison of the nucleotide sequence of the genomic clone (pEAP23) with that of the cDNA confirmed the presence of the single intron (65bp) in the coding region of *Eap*. The intron has the characteristic features of intron sequences observed in other filamentous fungi (Gurr et al. 1988). First it is small in size, 65bp. Second, the boundaries of the intron is similar to those of *S. cerevisiae* and filamentous fungi in possessing the nucleotide sequences GTNNGT and YAG bordering the 5' and 3'extremities respectively. Third, within each intron lies a sequence close to the consensus sequence YGCTAACN thought to be necessary for intron splicing (Gurr et al. 1988). Similar types of introns were observed in the Alp gene of *Aspergillus oryzae*. It contained three introns one of which was in the pro region and two in the mature coding region. The size of all the three introns were small (between 50-59 bp) (Tatsumi et al. 1991). The serine protease gene from *Acremonium chrysogenum* was reported to contain two introns (Isogai et al. 1991) while the *pepC* gene from *A. niger* contain a single intron (70nt) with all the characteristic features of a fungal intron (Frederick et al. 1993).

In *Eap*, 44 out of 61 possible codons are used and whenever there is a choice between a purine or pyramidine at 3rd position, in 83% of codons pyramidine is used. When a purine is used, 85% of the codons contain guanosine and when pyramidine is used 70% of them are cystosines. This codon bias is very similar to *pepC* gene in *A. niger* reported by Frederick and colleagues (Frederick et al. 1993). Similar codon bias has been observed for lignin and manganese peroxidases produced by *Phanerochaete chrysosporium* (de Boer et al. 1987, Ritch & Gold 1992).

The highly expressed or constitutive genes in filamentous fungi prefer codons ending in C and avoid codons ending in A (Ballace 1991). The order of preference for the 3rd codon in a codon family for such genes will be C>T>G>A. The *Eap* gene showing

this characteristic feature ensures the highly expressed nature of this gene. This observation is in agreement with the result obtained for the same enzyme in production studies (Sreeja Chellappan 2005, Sreeja Chellappan et al. 2005).

Comparison of *Eap* with other serine proteases in the databank showed that it shared extensive similarities to fungal members of the subtilisin family. *Eap* was found to have the highest level of homology with proteinase R (*Tritirarchium album*) (Kolvenbach et al. 1990). Homologies between amino acid as well as the nucleic acid sequences of the gene are about 96%. The homology extends to the pre-pro-region in addition to the sequence of the mature proteinase. However, there are a number of differences between EAP and proteinase R amino acid sequences. There are 17 aminoacid substitutions in EAP precursor sequence compared to that of proteinase R. A similar study was reported earlier by Peng and colleagues, where they observed a close similarity of 96.7% between mature peptides of subtilisin DFE and Subtilisin BPN (Peng et al. 2004). EAP also exhibited a considerable homology with proteinase K of *Tritirarchium album* at nucleotide (87.5%) as well as amino acid (84.2%) level.

The dendrogram, which indicate the phylogenetic relationship of various fungal subtilases with that of EAP, revealed that it belongs to the cluster of *T. album* serine proteases. Further it appears that the proteinase R and EAP are almost identical. The proteases of *F. oxysporum* and *Neurospora crassa* are found to be more closer to this cluster than that of *Beauveria* and *Metarhizium* spp. The higher degree of sequence identity of EAP with the extracellular serine proteases from different plant pathogenic (*Fusarium* sp), entomopathogenic (*Metarhizium* and *Beauveria* sp) and nematophagous (*Lecanicillum psaliotale*) fungi suggest that they might have evolved from a common ancestral subtilisin-like protease gene.

Generally, the extracellular subtilisin-like proteases are usually synthesized as precursors comprising a signal peptide, an N-terminal propeptide and a mature protease domain. The signal peptide is to assist in the secretion of the enzyme across the cytoplasmic membrane and is cleaved by a signal peptidase. The N-terminal propeptide generally functions as an intramolecular chaperone, and is auto-processed from the mature domain after the completion of enzyme folding (Eder et al. 1993, Yabuta et al. 2001). The N-terminal propeptides of subtilisin BPNP and subtilisin E also have the ability to inhibit their cognate mature enzymes (Li et al. 1995, Kojima et al. 1997).

The first 21 amino acid residues of the EAP polypeptide possessed the feature of a signal peptide with a high content of hydrophobic residues followed by a typical signal peptidase cleavage site (^{16}Ala-^{17}Pro-^{18}Ala) like that of proteinase K (Gunkel & Gassen 1989). But it is predicted based on Signal P algorithm that the EAP is presumably cleaved after ^{21}Ala. A similar situation was reported in Proteinase R (*T. album*) where in spite of a signal peptidase cleavage site present at ^{16}Pro-^{17}Pro-^{18}Ala, the signal peptide cleaves between ^{21}Ala and ^{22}Val.

It may be noted that SignalP algorithm is widely being used for the prediction of signal sequences present in various protease precursor molecules (Garcia-Sanchez et al. 2004, Pozo et al. 2004, Faraco et al. 2005, Wilson & Barbara 2005). In the previous study, EAP was reported as an extracellular enzyme (Sreeja Chellappan 2005). Analysis of EAP primary sequence with the target P algorithm further confirms that this protein is destined to the secretory pathway.

In the case of proteinase K, the 279 amino acid mature protein is generated by a cleavage between asparagine and alanine residues (Gunkel & Gassen 1989). This was further confirmed by comparing the deduced Proteinase K precursor sequence with the amino acid sequence data of the mature proteinase K (Jany et al. 1986). The initial 105aa (pre-propeptide) of the precursor protein was found missing in the mature protein

sequence (Gunkel & Gassen 1989). While the deduced protein sequence of proteinase R precursor possessed a putative pre-proregion of 108 aa, the length of the mature polypetide was reported to be 279 aa similar to that of proteinase K. Here, the coding sequence for the mature protein was ascertained by comparison with the N-terminal aminoacid sequence of proteinase K (Jany et al. 1986).

Since the N-terminal amino acid sequence of the mature EAP was not available, the comparison of the deduced amino acid sequence of this protease was made with that of other known subtilisin-like proteases, especially with that of Proteinase K and R of *Tritirarchium album* which showed high overall similarity, enabled it to postulate that the cleavage site of the N terminal propeptide localized between residues Asn^{108} and Ala^{109}. Thus the N-terminal propeptide of EAP is assumed to be composed of 87 amino acid residues, with a calculated molecular mass of 9570 Da.

In EAP and proteinase K, the propeptide shows a sequence similar to prokaryotic subtilisins but not with the properties of prochymotrypsinogen, protrypsinogen or other serine protease zymogens (which have limited similarity to each other). In both proteinases two stretches of aa show strong similarity to regions which were fully conserved in their sequences, and location within the subtilisin propeptides (Jany & Mayer 1985, Jany et al. 1986, Kolvenbach et al. 1990). The proproteinase K sequences (Tyr-Ile-Val-Lys-Phe-Lys and Ala-Thr-Leu-Asp-Glu) and the EAP propeptide sequences (Tyr-Ile-Val- Lys-Leu-Lys and Ala-Ser-Leu-Asp-Glu) correspond to the prosubtilisin E sequences Tyr-Ile-Val-Gly-Phe-Lys and Ala-Thr-Leu-Asp-Glu respectively, indicating that all these sequences may have involved in protein interaction during maturation of the proprotein and thus are under strong evolutionary pressure. It seems likely that the propeptide has a role in the activation like those of the mammalian enzymes (Gunkel & Gassen 1989). However, further experiments are needed to determine the kinetics and cellular location of this presumptive activation process.

A series of experiments had shown that the pro-sequence is essential for producing active subtilisin in vivo as well as in vitro, and it functions by facilitating correct folding of the protease domain (Ikemura et al. 1987, Shinde & Inouye 1993). This pro-sequence functions as an intramolecular chaperone, but is not required for enzymatic functioning of the folded protein (Shinde & Inouye 1993). Interestingly, point mutations within the pro-sequence can affect its folding function (Li et al. 1995).

The pro-peptide cleavage site is before the N-terminus of the secreted protein, the final residue of the pro-peptide is an asparagine (N), position 108 in EAP and proteinase R (Samal et al. 1990), 105 in proteinase K (Gunkel & Gassen 1989), 99 in cuticle degrading protease, Prl (Sheng et al. 2006) and protease of *Fusarium* (Morita et al. 1994). The deduced mature EAP (279 aa) has a calculated molecular mass of 29.129kDa. This prediction is supported by the MALDI-MS data for the EAP (Sreeja Chellappan 2005), which indicated the size of EAP as 28.6kDa.

Three consensus stretches of residues are being observed in most of the subtilisin like serine proteases in which the catalytic triad residues are embedded (Siezen & Leunissen 1997). In EAP this typical signature patterns were observed in the positions: 143-154aa (VCVI**D**TGVeasH), 177-187aa (**H**GThCAGtIGS) and 330-340aa (GT**S**mAtPhVAG) and the catalytic triad residues are the following (^{39}Asp, ^{69}His and ^{224}Ser). This feature of subtilisin like serine proteases is reported earlier for *Pen* c1 gene from *P. citrinum* (Su et al. 1999), subtilase genes from Sapstaining fungi (Hoffman & Breuil 2002), *kex*2 gene from Paracoccidioides *brasiliensis* (Venancio et al. 2002), SPMI gene of *Magnaporthe grisea* (Fukiya et al. 2002) and also in SUB genes from *Trichophyton rubrum* (Jousson et al. 2004). In addition to this, a highly conserved oxyanion hole residue, which is important in subtilisins for stabilization of the reaction intermediate formed during proteolysis, was also being observed in EAP (^{269}Asn). Similar instance was reported in Pen c1 gene from *P. citrinum* (Su et al. 1999) and kex2 gene from *Paracoccidioides brasiliensis* (Venancio et al. 2002).

A free cysteine residue near the active site His which has been postulated to play a role in catalysis was found conserved in many serine proteases including Proteinase R from *T. album* (Samal et al. 1990), *Alp* from *Fusarium* spp (Morita et al. 1994) and *Prl* from *Metarhizium anisopliae* (St. Leger et al. 1992). This conserved residue is being observed in EAP (^{181}Cys) also. Conversely, this highly conserved cysteine residue is being substituted by a Valine in Subtilase gene of *Fusarium oxysporum* (Pietro et al. 2001) and in *Trichoderma harzianum* protease gene (Gerem´ıa et al. 1993).

Two stretches of residues (Ser-Leu-Gly and Ala-Ala-Gly) which form part of the specific crevice (sides of the S1 pocket) in functional subtilisin molecule, within which the P1 site of the substrate side chain fits, was reported as highly conserved in several proteases like alkaline protease of *A. oryzae* (Tatsumi et al. 1989), Proteinase K and T of *Tritirarchium album* (Gunkel & Gassen 1989, Samal et al. 1989). In EAP also, these stretches were found well conserved (^{240}Ser- ^{241}Ile - ^{242}Gly) and (^{266}Ala - ^{267}Ala- ^{268}Gly) but with a single substitution of Leu with Ile. The same substitution was being reported in the context of Proteinase R from *Tritirarchium album* (Samal et al. 1989). The analysis of EAP primary sequence for the presence of protein domains enabled the recognition of the highly conserved Peptidase S8 domain (pfam 00082) between position 120-377 and the Subtilisin N-terminal domain (pfam 05922) between positions 40-108.

Many of the extracellular proteases are found to be glycosylated. The serine proteases of *Metarhizium anisopliae* (St. Leger et al. 1992) and *Arthrobotrys oligospora* (Ahman et al. 1996), contain potential N-glycosyltion sites while two possible O-glycosylation sites were being reported in tvsp1 gene of *Trichoderma virens* (Pozo et al. 2004). However, in EAP, none of these potential glycosylation sites (N-linked/ O-linked) were detected. Likewise, pr1 gene of *Beauveria bassiana* (Joshi et al. 1995) and vcp1 gene of *Pochonia chlamydosporia* (Morton et al. 2003) were also reported to lack any potential glycosylation sites. Glycosylation play key role in targeting the protein to different locations, serving as biological labels. The glycosylated proteins usually get

secreted or get localized to the cell surface. The lack of glycosylation sites in EAP which can lead it in to the secretary pathway can be justified by the presence of the signal peptide in the precursor protein which may otherwise direct the protein to ER lumen and then to the secretory pathway.

In EAP one potential Tyrosine sulfation site was recognized at ^{78}Tyr of the precursor EAP while two putative sulfated tyrosines were reported in the serine protease gene of *Trichoderma virens* (Pozo et al. 2004). Tyrosine sulfation is an important post-translational modification found on many secreted and transmembrane-proteins. It was suggested to be important for increased stability and or correct folding (O¨nnerfjord et al. 2004).

In EAP, fourteen possible myristoylation sites were recognized. Protein N-myristoylation refers to the covalent attachment of myristate (a 14-carbon saturated fatty acid) to the N-terminal glycine of eukaryotic and viral proteins. N-myristoylated proteins have diverse functions and intracellular destinations. They include proteins involved in a wide variety of signal transduction cascades. The N-myristoylation of the protein promotes weak and reversible protein-membrane and protein-protein interactions. Typically myristate moiety attached to the protein act in concert with other mechanisms to regulate the protein targeting and function (Farazi et al. 2001). It can be suggested that the myristoylated EAP may get targeted to the plasma membrane where it reversibly (and non-covalently) bind with the myristoyl receptors on the membrane and later get secreted out of the membrane efficiently.

In addition to this, several phosphorylation sites, which are substrates for four different classes of protein kinases, were also recognized in the EAP primary sequence. Out of the 13 phosphorylation sites predicted, six of them belong to Casein kinase II type, five belong to Protein kinase C type while one site each were recognized for Tyrosine kinase and cAMP- and cGMP-dependent protein kinases. Multiple phosphorylation sites

were also reported in serine protease of *Trichoderma virens* (Pozo et al. 2004). In the S3a ribosomal protein of *Eimeria tenella* three different types of phosphorylation sites were detected and it was proposed that these proteinase kinases are involved in the regulation of function of this ribosomal protein (Ouarzane et al. 1998). Phosphorylation provide an allosteric mode of protein regulation which activate or inhibit (reversibly) a wide variety of cellular proteins in response to environmental signals.

Presence of all these kind of post-traslational modification signals in EAP suggest that this protein may have some other important role in fungi in addition to its well known role in nutrition acquisition through extracellular protein scavenging, such as in signal transduction cascades. This view is supported by the point that in many cases particularly in higher eukaryotes, the subtilases have developed in to highly specialized enzymes of biosynthetic pathways where they are involved in processing and maturation of pro-proteins (Siezen & Leunissen 1997).

The results obtained for the comparative analysis of various secondary structure elements of EAP were not conclusive since the GOR I and II based predictions showed much variation in the pattern of secondary structure elements compared to the other two methods (PHD and SOPMA), which gave matching results. Later, the secondary structure of EAP predicted by the above four methods was also compared with the actual secondary structure elements of proteinase K obtained by X-ray crystallography which gave a reliable picture of the various secondary structure elements of EAP. In fact the secondary structure prediction of EAP by PHD method was closer to the X-ray crystallography based secondary structure of proteinase K. Certainly an accurate picture of the various secondary structure elements of EAP is possible only if X-ray crystallographic analysis of the protein is done.

In addition to this, a comparative analysis of the secondary structure of EAP with four other serine protease (proteinase R and K from *T. album*, alkaline protease of *Fusarium* sp and *Prl* of *C. brongniartii*) was done with the program PSI-PRED (http://www.us.expasy.org/). This analysis revealed that the secondary structure patterns of these enzymes are largely similar.

The hydrophobicity plot for the EAP show marked similarity with the other four plot patterns (for proteinase R and K from *T. album*, alkaline protease of *Fusarium* sp and *Prl* of *C. brongniartii*) indicating that the folding patterns of proteases are identical to a great extends. It also indicate that the N-terminal portion of all these polypeptides are hydrophobic, a characteristic feature of signal sequences. Thus, this analysis also confirms the SignalP prediction of the signal sequence in the case of EAP precursor. In a similar study, hydropathy and surface plot patterns of mature prl of *M. anisopliae* and proteinase K were compared to show that the folding pattern of both the proteins are similar (St. Leger et al. 1992). The similarity of folding patterns in *prl*B protease of *M. anisopliae* with that of the *T. album* proteinase K based on the surface probability and hydrophilicity plots was reported (Joshi et al. 1997). All these studies indicated that though the primary sequences of these proteases are showing much variation, their secondary structures and folding patterns are seem to be largely similar.

EAP is identified as a serine protease belonging to the Class II subtilases, which includes Proteinase K like proteinases. The enzyme shares a very significant homology with the Proteinase K from *T. album* and majority of the residues (86% - 241/279) are identical. There is about 92% homology between the two proteins and the sequences align without gaps. This is indicative of the fact that the modeled enzyme from marine *E. album* strain might be the same protein identified in *T. album* (*E. album*) but with minor substitutions, which generally have not changed the protein to a great extent.

Homologies modeling of subtilases (class II) are generally performed using the crystal structure of Proteinase K as the template (Siezen et al. 1991). Thus, the protease VCP1 of *Pochonia chlamydosporia* was modeled based on Proteinase K as structural template (Morton et al. 2003). Previous to this study, the homology modelling of stetterlysin and pyrolysin were reported, where the modeling was based on the crystal structures of subtilisin BPN9 and thermitase (Voorhorst et al. 1997). The homology modelling led to the prediction of three dimensional model of the catalytic domain core of stetterlysin and pyrolysin for which over 92% of the residues (excluding Gly, Pro) had phi-psi angles in most favoured or allowed regions in a Ramachandran plot. Moreover, side chain parameters were all within acceptable limits as assessed by PROCHECK (Laskowski et al. 1993). Further, the alignment and the predicted three-dimensional models were used to analyze amino acid composition and structural features of the catalytic domain of the two enzymes.

The three-dimensional structure model for EAP was constructed based on the crystal structure of proteinase K. The reliability and stereochemical quality of the predicted 3D model was analyzed using the program PROCHEK (Laskowski et al. 1993). Sequence alignments and homology model of EAP show that the protein is almost identical to the *T. album* proteinase K.

When the three dimensional structure of the active site and the peptide binding region of EAP was modeled and superimposed on the corresponding regions of Proteinase K model, it was inferred that the catalytic triad overlaps almost exactly as is the Asn 161 residue, indicating the conservation of the active site in EAP and there was no substitutions in the residues closer (within 4 A°) to the catalytic triad. Except for the three substitutions at position 99,102 and 133, the substrate-binding domain of EAP is exactly similar to that of PRK1. But from the 3D model superimposition it was obvious that these substitutions do not place any significant changes in the conformation of the substrate-binding domain of EAP when compared to that of proteinase K.

In general, the substrate-binding region in subtilases can be described as a surface channel or crevice capable of accommodating at least six amino acid residues (P4-P2') of a polypeptide substrate or inhibitor (pseudo-substrate). Both main-chain and side chain interactions between enzyme and substrate/inhibitor contribute to binding (Siezen & Leunissen 1997). In EAP and proteinase K, the P4-P1 specificity side of the substrate line up between the extended enzyme backbone segment 100-103 and 132-135, forming the central strand of a three-stranded antiparallel β sheet.

In subtilisins and thermitase P1-P4 pockets are large and hydrophobic, which explains the broad specificity of these enzymes with a preference for aromatic or large nonpolar PI and P4 substrate residues (Gron et al. 1992). Variations in the substrate specificity of naturally occurring subtilases should be due to (and could be modified by) modulation of the residues in the substrate-binding region and in particular those residues whose side chains interact with PI and P4 substrate residues. Engineering studies of subtilisin BPN' have demonstrated that P1 specificity can be dramatically modulated by substitutions of G166 at the bottom of the P1 cleft (Siezen et al. 1991), in addition charged substitutions at positions 166 or 156 (at the P1 cleft entrance) shift the specificity to oppositely charged P1 residues (Siezen et al. 1991). Proteinase K and several related class II subtilases have a Tyr at position 169 (with respect to Proteinase K), but since it is rotated away, the PI cleft remains wide and hydrophobic (Betzel et al. 1988a, Betzel et al. 1988b, Siezen et al. 1991). This explains why proteinase K also has a broad PI specificity with a slight preference for aromatic and hydrophobic residues. The residue at the corresponding position of EAP also is a Tyr (^{169}Tyr) indicating that the PI cleft in EAP is wide and hydrophobic and it may have a broad PI specificity with a slight preference for aromatic and hydrophobic residues similar to that of proteinase K.

In P4 pocket, the hydrophobic side chains are found generally conserved in wild type subtilases. Residue 104 at the pocket entrance varies considerably from hydrophobic (Y, A, L) to hydrophilic (T, S, N, D), and this exposed residue is known to exhibit

conformational flexibility in mutant subtilisins (McPhalen & James 1988). In subtilisin (BPN'), Tyr at position 104 is thought to form a flexible lid to the S4 pocket (Takeuchi et al. 1991). In proteinase K and EAP this tyrosine residue is found to be conserved at position 104.

Five positionally conserved cysteine residues are observed in EAP at aa positions 142,181,231,286 and 357. These cysteine residues are found highly conserved in proteinase K and R of *T. album*, Alp of *Fusarium* sp and *Prl* of *C. brongniartii*. In proteinase K, four of these highly conserved cysteine molecules had been shown to form two disulphide bonds. The cysteine at 139 binds to that at 228 and the 283 to 354 (Jany et al. 1986). The thermostable nature of the enzyme proteinase K is attributed to the presence of two-disulphide bonds in it. Considering the close similarity of EAP with proteinase K, presence of two disulphide bonds was proposed in EAP also.

Disulfide bridges can contribute to the overall stability of a protein, and the introduction of new **S-S** bonds can enhance the thermal stability, as demonstrated in, for example, phage T4 lysozyme (Matsumura et al. 1989). Initial attempts to stabilize subtilisin by introduction of **S-S** bonds 22-87, 24-87, 26-232, 29-1 19, 36-210, 41-80, and 148-243 were not successful (Wells & Powers 1986, Pantoliano et al. 1987, Mitchinson & Wells 1989, Katz & Kossiakoff 1990); All of these cross links were designed by inspection of the three-dimensional structure of subtilisin. The first successful thermal stabilization of subtilisin was the introduction of the 61-98 **S-S** bond (Takagi et al. 1990), which occurs naturally in aqualysin. It stands to reason, therefore, that naturally occurring **S-S** bonds should provide a better choice for stabilization of subtilisins than previously designed disulfides.

After homology modeling, in the 3D model of the protein it was observed that the cysteines at 34 and 123 come close enough to suspect a SS linkage. Similar was the case with cysteines at 178 and 249. Prediction of disulphide linkages between these residues

was performed using the "DISULFIND" program (Ceroni et al. 2006) and these suspected SS bonds were predicted with maximum confidence. Experimental evidences also suggest that EAP is thermostable to a good extent. It is being reported that EAP is active over a broad range of incubation temperature with maximal activity at 60°C (Sreeja Chellappan 2005).

The 3D model for *E. album* proteinase (EAP) was analyzed for putative "Ca" binding sites using the program GG v 1.0 (Deng et al. 2006) and this led to the identification of four putative Ca ion locations, Ca 001(^{11}Ala, ^{14}Ser, ^{14}Ser, ^{22}Ser), Ca 002 (^{30}Gly, ^{33}Thr, ^{33}Thr, ^{88}Thr), Ca 003 (^{171}Pro, ^{172}Ala, ^{174}Glu, ^{179}Thr), and Ca 004(^{175}Ser, ^{177}Ile, ^{200}Asp, ^{200}Asp).

Binding of Ca^{2+} ions at specific sites, often in external loops, increases the stability of many proteases by reducing the flexibility of the molecule and hence the denaturation and/or autolysis rate (Siezen et al. 1991). Four calcium-ion binding sites are known from crystal structures of subtilisins, thermitase, and proteinase K; these calcium ions are essential for stability and activity. From previous sequence alignments and homology modeling it was predicted that the Cal (strong) and Ca3 (weak) sites are most common in members of the subtilase family, whereas the Ca2 (medium-strength) site is less common (Siezen et al. 1991). The weak Ca4 site has only been found in proteinase K (Betzel et al. 1988a, Betzel et al. 1988b). Experimentally it has been shown that Ca^{2+} is having the role of a stabilizer promoting thermal stability during incubation of the enzyme (EAP) at various temperatures, with maximal residual activity at 60°C (Sreeja Chellappan 2005). This testifies the prediction that EAP contain Ca binding sites in it structure.

Chemical oxidation can be a significant source of enzyme inactivation, particularly for enzymes, which function extracellularly. Oxidative stability can be improved by replacement of oxidatively sensitive residues, mainly cysteine and

methionine, particularly those near the active site (Siezen et al. 1991). In subtilisin BPN' the residues M124 and M222, both near the active site, are susceptible to oxidation by hydrogen peroxide; the bulky sulphoxide derivative near the catalytic site leads to a reduction of enzyme activity (Siezen et al. 1991). Protein engineering studies showed that the best alternatives for M222 are the non-oxidizable residues Ala, Ser and Leu; this led to mutant subtilisins with retention of 12-53% activity and complete resistance to oxidation (Siezen et al. 1991). Such a mutant M222A (and G 195E) subtilisin called Durazyme R is presently used in a commercial detergent (Siezen et al. 1991). Previously it has been reported (Sreeja Chellappan 2005) that hydrogen peroxide has a drastic inhibitory effect on protease (EAP) even at the lowest concentrations tried. In EAP, three residues (^{36}Cys, ^{131}Met, ^{225}Met) were suspected to be of this category, out of which the ^{36}Cys and ^{225}Met residues, which lie close to the Aspartate and Serine residues of the catalytic triad, are more prone to oxidation. Further studies are needed to confirm the role of these residues in oxidative susceptibility of the enzyme. Once the vulnerable residues are identified, engineering of EAP with oxidative stable residues at the corresponding positions can be accomplished by protein engineering strategies.

Structural comparison of proteins has proven to be very useful in creating proteins with desired properties. Molecular techniques such as site-directed mutagenesis have been used successfully to increase its thermostability (Takagi et al. 1990) and activity (Takagi et al. 1988) and to alter its pH optimum (Thomas et al. 1985). Higher thermal stability is a desirable trait for industrial application of the *Conidiobolus* protease. Since the *E. album* protease (EAP) and proteinase K are structurally very similar, existing knowledge about proteinase K can be gainfully employed to improve the thermostability of this fungal enzyme.

Chapter 6

SUMMARY AND CONCLUSION

The present study included mainly isolation of the intact full gene encoding alkaline serine protease from *Engyodontium album* (*Eap*) by molecular cloning and characterization of the same. Various cloning strategies were tried to isolate the full gene, which include cDNA library construction, genome library construction through shotgun cloning and screening the library using probes. Of the various strategies employed the degenerate oligo-primers and homologous probe based screening of sub genomic DNA library was successful to clone the *Eap* gene encoding alkaline protease from *E. album*. Partial fragment of the gene was amplified by PCR using degenerate primers designed based on fungal alkaline serine protease sequences. The amplified (878bp) fragment was cloned in pMOS and sequenced. The nucleic acid sequence was analyzed by various computer algorithms and confirmed its similarity with different fungal alkaline serine proteases especially with proteinase R and proteinase K from the fungus *Tritirarchium album*.

Genomic southern hybridization, using partial gene fragment as the probe revealed that the protease gene occur as a single copy in the genome of the fungus. A gene enriched sub-genomic DNA library was constructed based on the genomic southern data. The library was screened with the partial gene probe and also by using colony PCR technique. The positive clone obtained was sequenced to confirm the presence of the full coding region. Restriction enzyme profile of the positive clone was also created. A set of primers was designed from the 5' and 3' un-translated region of the genomic clone, and the cDNA of the protease gene was raised by RT-PCR.

A total of 1,424bp of the EAP locus were sequenced, which consist of an open reading frame of 1161bp flanked by 93bp in the 5'non-coding region and 102bp in the 3' non-coding region. Comparison of the nucleotide sequence of the genomic clone with that of the cDNA confirmed the presence of the single intron (65bp) in the coding region of EAP.

Homology comparison of the deduced amino acid sequence of *Eap* with other known proteins indicated that gene *Eap* encode an extracellular protease that belongs to the subtilisin family of serine protease. The putative peptide encoded by the *Eap* gene is a prepropeptide consisting of 387 amino acid residues with a calculated molecular mass of 40kDa. The presence of aa putative signal peptide sequence of 21 residues at the N terminus indicates that EAP could be secreted. Moreover a highly conserved subtilisin N-terminal domain (pfam 05922) was observed sub-terminally. The cleavage site of this putative propeptide (between N-108 and A-109) was predicted according to sequence similarity. Based on these facts, it is proposed that the *E. album* protease undergoes two processing events: One after amino acid 21 to remove the signal sequence and the second process after amino acid 108 to yield the mature enzyme. The putative mature protein (residues 109-387) is of 279aa in length with the calculated molecular mass of 29kDa.

The sequences surrounding the catalytic triad (^{39}Asp, ^{69}His, and ^{224}Ser), signatures of the subtilisin proteinase family were found well conserved in EAP. It also shows similar structural features to subtilisin enzymes, such as (i) two stretches; (^{240}Ser- ^{241}Ile - ^{242}Gly) and (^{266}Ala- ^{267}Ala- ^{268}Gly) which form part of the specific crevice in subtilisin within which the P1 site of the substrate side chain fits (ii) the oxyanion hole residue (Asn-269), which helps stabilize oxyanion in a transition state in most subtilisins, and (iii) the internal helix hC and hF which are highly conserved in subtilases [177-184:HGTHCAGT;330-344:GTSMATPHVAGLAAY in EAP].

The analysis of EAP primary sequence for the presence of protein domains enabled the detection of the highly conserved Peptidase S8 domain (pfam 00082) between position 120-377. Analysis of EAP with the Prosite program revealed consensus sequences for one tyrosine sulfation site, one tyrosine kinase phosphorylation site, one cAMP- and cGMP-dependent protein kinase phosphorylation site, five protein kinase C phosphorylation site, six casein kinase II phosphorylation site and fourteen myristoylation sites. Presence of all these post-traslational modification signals in EAP suggest that this protein may have some other important role besides its well known role in nutrient acquisition through extracellular protein scavenging.

The three dimensional structural model of the EAP was predicted based on the crystal structure of Proteinase K. The model was assessed by PROCHECK and its reliability and stereochemical quality was assured. Based on the three dimensional model, predictions were made for EAP concerning essential conserved residues, allowable substitutions, disulfide bonds, Ca^{2+} binding sites and substrate binding site residues. Out of the six cysteine residues present in the primary sequence of EAP, four were suggested to be involved in two disulfide bond formation (^{34}Cys -^{123}Cys) and (^{178}Cys - ^{249}Cys). Four possible Ca^{2+} binding sites were also predicted; both of which may be contributing to the thermostable nature of the protein.

Based on the present study the full gene sequence of the alkaline serine protease of the marine fungus *E. album* was elucidated besides understanding of the structure and function of the enzyme. The sequence alignments and homology modelling would provide the basis for developing a protein-engineering strategy aimed at modulating stability, catalytic activity, or substrate specificity of the enzyme.

Chapter 7

LITERATURE CITED

A°hman J, Johansson T, Olsson M, Punt PJ, Hondel CAvd, Tunlid A (2002) Improving the Pathogenicity of a Nematode-Trapping Fungus by Genetic Engineering of a Subtilisin with Nematotoxic Activity Applied and Environmental Microbiology 68:3408 - 3415

Ahman J, B E, Rask L, Tunlid A (1996) Sequence analysis and regulation of a gene encoding a cuticle-degrading serine protease from the nematophagous fungus *Arthrobotrys oligospora*. Microbiology 142:1605 - 1616

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403 - 410

Altschul S, Madden T, Schäffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Research 25:3389 - 3402

Anwar A, Saleemuddin M (1998) Alkaline proteases: A review. Bioresource Technology 64:175-183

Argos P (1987) A sensitive procedure to compare amino acid sequences. Journal of Molecular Biology 193:385 - 396

Bae J-H, Sohn J-H, Rhee S-K, Choi E-S (2005) Cloning and characterization of the *Hansenula polymorpha* PEP4 gene encoding proteinase A. Yeast 22:13 - 19

Bagga S, Hu G, Screen S, St. Leger R (2004) Reconstructing the diversification of subtilisins in the pathogenic fungus *Metarhizium anisopleae*. Gene 324:159 - 169

Ballace D (1991) Transformation systems for filamentous fungi and an overview of fungal gene structure. In: Leong S, Berka R (eds) Molecular Industrial Mycology. Dekker, New York, p 1-29

Banerjee R, Bhattacharya B (1993) Kinetic properties of extracellular alkaline proteases of *Rhizopus oryzae*. Journal of Fermentation and Bioengineering 75:380 - 382

Bang M, Villadsen I, Sandal T (1999) Cloning and characterization of an endo-b-1,3(4)glucanase and an aspartic protease from *Phaffia rhodozyma* CBS 6938. Applied Microbiology and Biotechnology 51:215 - 222

Bangyeekhun E, Cerenius L, Söderhäll K (2001) Molecular Cloning and Characterization of Two Serine Proteinase Genes from the Crayfish Plague Fungus, *Aphanomyces astaci*. Journal of Invertebrate Pathology 77:206 -216

Barindra Sana, Debashish Ghosh, Malay Saha, Mukherjee J (2006) Purification and characterization of a salt, solvent, detergent and bleach tolerant protease from a new gamma-Proteobacterium isolated from the marine environment of the Sundarbans. Process Biochemistry 41:208 - 215

Baroch A, Bucher P, Hofmann K (1997) The PROSITE database, its status in 1997. Nucleic Acid Research 25:217 - 221

Benito M, Connerton I, Cordoba J (2006) Genetic characterization and expression of the novel fungal protease, EPg222 active in dry-cured meat products. Canadian Journal of Microbiology 52:550 - 559

Bennett J (1998) Mycotechnology: the role of fungi in biotechnology. Journal of Biotechnology 66:101 - 107

Berka O, Carmona C, Hayenga K, Thompson S, Ward M (1993) Isolation and characterization of the *Aspergillus oryzae* gene encoding aspergillopepsin Gene 125:195 - 208

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

Betzel C, Belleman M, Pal G, Bajorath J, Saenger W, Wilson K (1988a) X-ray and model-building studies on the specificity of the active site of proteinase K. Proteins structural and Functional Genetics 4:157 - 164

Betzel C, Gourinath S, Kumar P, Punit Kaur, Perbandt M, Eschenburg S, Singh TP (2001) Structure of a Serine Protease Proteinase K from *Tritirarchium album* limber at 0.98 Å Resolution. Biochemistry 40:3080 - 3088

Betzel C, Pal G, Saenger W (1988b) Three-dimensional structure of proteinase K at 0.15 nm resolution. European Journal of Biochemistry 78:155 - 171

Betzel C, Pal G, Struck M, Jany K, Saenger W (1986) Active site geometry of proteinase K. FEBS Letters 197:105 - 110

Bhosale S, Rao M, Deshpande V, Srinivasan (1995) Thermostability of high activity alkaline serine protease from *Conidiobolus coronatus* (NCL 86.8.20). Enzyme and Microbial Technology 17:136 - 139

Bidochka M, Melzer M (2000) Genetic polymorphisms in three subtilisin-like protease isoforms (Pr1A, Pr1B, and Pr1C) from *Metarhizium* strains. Canadian Journal of Microbiology 46:1138 - 1144

Bindschedle LV, Sanchez P, Dunn S, Mikan J, Thangavelu M, Clarkson JM, Cooper RM (2003) Deletion of the SNP1 trypsin protease from *Stagonospora nodorum* reveals another major protease expressed during infection. Fungal Genetics and Biology 38:43 - 53

Bonants P, Fitters P, Thijs H, den Belder E, Waalwijk C, Henfling J (1995) A basic serine protease from *Paecilomyces lilacinus* with biological activity against *Meloidogyne hapla* eggs. Microbiology 141:775 - 784

Brown AJP, Lithgow G (1987) The structure and expression of nuclear genes in *Saccharomyces cerevisiae*, Vol. IRL Press, Oxford

Caballero JL, Redondo-Nevado J, Cardenas J, Pineda M (1997) A procedure for cloning genes from genomic DNA using weakly hybridizing heterologous probes and a polymerase chain reaction based screening:cloning of the chickpea urate oxidase gene. Analytical Biochemistry 244:167 - 169

Ceroni A, Passerini A, Vullo, Frasconi P (2006) DISULFIND: a Disulfide Bonding State and Cysteine Connectivity Prediction Server. Nucleic Acid Research 34 (Web server issue):W177 - W181

Chakrabarti S, Matsumura N, Ranu R (2000) Purification and characterization of an extracellular alkaline serine protease from *Aspergillus terreus* (IJIRA 6.2). Current Microbiology 40:239 - 244

Cheevahanarak S, Renno D, Saunders G, Holt G (1991) Cloning and selective overexpression of an alkaline protease-encoding gene from *Aspergillus oryzae*. Gene 108:151 - 155

Chen E, Seeburg P (1985) Supercoil sequencing : A fast and simple method for sequencing plasmid DNA. DNA 4:165 - 170

Chi Z, Ma C, Wang P, Li H (2006) Optimization of medium and cultivation conditions for alkaline protease production by the marine yeast *Aureobasidium pullulans*. Bioresource Technology Article in Press

Chomcynski P, Sacchi N (1987) Single step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162:156 - 159

Chou H, Lai H, Tam M, Chou M, Wang S, Han S, Shen H (2002) cDNA cloning, biological and immunological characterization of the alkaline serine protease

major allergen from *Penicillium chrysogenum*. International Archieves in Allergy and Immunology 127:15 - 26

Chung C, Goldberg A (1981) The production of the lon (capR) gene in *Escherichia coli* is the ATP dependent protease, protease LaT. Proceedings of National Academy of Science USA 78:4931 - 4935

Clark S, Templeton M, Sullivan P (1997) A secreted aspartic proteinase from *Glomerella cingulata*: purification of the enzyme and molecular cloning of the cDNA. Microbiology 143:1395 - 1403

Cohen B (1977) The proteases of *Aspergillus*. In: Smith J, Pateman J (eds) Genetics and physiology of *Aspergillus*. Academic Press, New York, p 281-292

Croocker P, Sako Y, Uchida A (1999) Purification and characterization of an intracellular heat-stable proteinase (Pernilase) from the marine hyperthermophilic archaeon *Aeropyrum pernix* K1. Extremophiles 3:3-9

Damare S, Raghukumar C, Muraleedharan UD, Raghukumar S (2006) Deep-sea fungi as a source of alkaline and cold-tolerant proteases. Enzyme and Microbial Technology 39:172 - 181

Davidow L, O'Donnel M, Kaczmarek F, Pereira D, DeZeeuw J, Franke A (1987) Cloning and sequencing of the alkaline extracellular protease gene of *Yarrowia lipolytica*. Journal of Bacteriology 169:4621 - 4629

de Boer H, Zhang Y, Collins C, Reddy C (1987) Analysis and nucleotide sequence of two ligninase cDNAs from a white-rot filamentous fungus *Phanerochaete chrysosporium*. Gene 60

Delgado-Jarana J, Rincon AM, Benitez T (2002) Aspartyl protease from *Trichoderma harzianum* CECT 2413; cloning and characterization. Microbiology 148:1305 - 1345

Deng H, Chen G, Yang W, Jenny J. Yang (2006) Predicting Calcium Binding Sites in Proteins-A Graph Theory and Geometry Approach. http://chemistry.gsu.edu/faculty/Yang/GG.htm

Descamps F, Brouta F, Monod M, Zaugg C, Baar D, Losson B, Mignon B (2002) Isolation of a *Microsporum canis* gene family encoding three subtilisin-like proteases expressed in vivo. Journal of Investigative Dermatology 119:830 - 835

Deshpande M (1992) Proteinases in fungal morphogenesis. World Journal of Microbiology and Biotechnology 8:242 - 250

Dienes D, Börjesson J, Hägglund P, Tjerneld F, Lidén G, Réczey K, Stålbrand H (In Press) Identification of a trypsin-like serine protease from *Trichoderma reesei* QM9414 Enzyme and Microbial Technology

Dobinson K, Lecomte N, Lazarovits G (1997) Production of an extracellular trypsin-like protease by the fungal plant pathogen *Verticillium dahliae*. Canadian Journal of Microbiology 43:227 - 233

Eder J, Rheinnecker M, Fersht A (1993) Folding of subtilisin BPNP: Role of the pro-sequence. Journal of Molecular Biology 233:293 - 304

Enderlin C, Ogrydziak D (1994) Cloning, nucleotide sequence and functions of XPR6, which codes for a dibasic processing endoprotease from the yeast *Yarrowia lipolytica*. Yeast 10:67 - 79

Estrade-Badillo C, Facundo J, Rocha M (2003) Effect of agitation rate on biomass and protease production by a marine bacterium *Vibrio harveyi* cultured in a fermentor. World Journal of Microbiology and Biotechnology 19:129 - 133

Fang W, Zhang Y, Yang X, Wang Z, Pei Y (2002) Yi Chuan Xue Bao. Cloning and characterization of cuticle degrading enzyme CDEP-1 from *Beauveria bassiana* 29:278 - 282

Faraco V, Palmieri G, Festa G, Monti M, Sannia G, Giardina P (2005) A new subfamily of fungal subtilases: structural and functional analysis of a *Pleurotus ostreatus* member. Microbiology 151:457 - 466

Farazi TA, Waksman G, Gordon JI (2001) The biology and enzymology of protein N-myristoylation. The Journal of Biological Chemistry 276:39501 - 39504

Farrel D, Crosa J (1991) Purification and characterization of a secreted protease from the pathogenic marine bacterium *Vibrio anguillarum*. Biochemistry 30:3434 - 3436

Feinberg A, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Analytical Biochemistry 132:6 - 13

Fox J, Shannon J, Bjarnason J (1991) Proteinases and their inhibitors in biotechnology. Enzymes in biomass conversion. ACS symposium series 460:62 - 79

Frederick GD, Rombouts P, Buxton FP (1993) Cloning and characterization of *pepC*, a gene encoding serine protease from *Aspergillus niger*. Gene 125:57 - 64

Fukiya S, Kuge T, Tanishima T, Sone T, Kamakura T, Yamaguchi I, Tomita F (2002) Identification of a putative vacuolar serine protease gene in the rice blast fungus, *Magnaporthe grisea*. Bioscience Biotechnology and Biochemistry 66:663 - 666

Ganesh Kumar C, Takagi H (1999) Microbial alkaline proteases: From a bioindustrial viewpoint. Biotechnology Advances 17:561–594

Garcia-Sanchez A, Cerrato R, Larrasea J, Ambrose NC, Parra A, Alonso JM, Hermoso-de-Mendoza M, Rey JM, Mine MO, Carnegie PR, Ellis TM, Masters AM, Pemberton AD, Hermoso-de-Mendoza J (2004) Characterization of an extracellular serine protease gen (*nasp* gene) from *Dermatophilus congolensis*. FEMS Microbiology Letters 231:53 - 57

Garnier J, Osguthorpe D, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. Journal of Molecular Biology 120:97 - 120

Gente A, Durand-Poussereau N, Fevre M (1997) Controls of the expression of aspA, the aspartyl protease gene from *Penicillium roqueforti* S. Molecular Genetics and Genomics 256:557 - 565

Geourjon C, Deléage G (1994) SOPM: a self-optimised method for protein secondary structure prediction. Protein Engineering 7:157 - 164

Gerem´ıa R, Goldman G, Jacobs D, Ardiles W, Vila S, Van Montagu M, Herrera Estrella A (1993) Molecular characterization of the proteinase-encoding gene, *prb1* related to mycoparasitism by *Trichoderma harzianum*. Molecular Microbiology 8:603 - 613

Gibrat J, Garnier J, Robson B (1987) Further developments of protein secondary structure prediction using information theory. Journal of Molecular Biology 198:425 - 443

Godfrey T, West S (eds) (1996) Industrial enzymology, Vol. Stockton Press, New York

Gomez M, Cutting S, Stragier P (1995) Transcription of spoIVB is the only role of jG that is essential for pro-jK procession during spore formation in *Bacillus subtilis*. Journal of Bacteriology 177:4825 - 4827

Gomi K, Arikawa K, Kamiya K, Kitamoto K, Kumagai C (1993) Cloning and nucleotide sequence of the acid protease-encoding gene (pepA) from *Aspergillus oryzae*. Bioscience Biotechnology and Biochemistry 57:1095 - 1100

Gottesman S, Maurizi M (1992) Regulation by proteolysis, energy dependent proteases and their targets. Microbiological Reviews:591 - 621

Gray GL, Hayenga K, Cullen D, Wilson LJ, Norton S (1986) Primary structure of *Mucor miehei* aspartyl protease: evidence for a zymogen intermediate. Gene 48:41 - 53

Gron H, Meldal M, Breddam K (1992) Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. Biochemistry 3:6011 - 6018

Gunkel F, Gassen H (1989) Proteinase K from *Tritirarchium album* limber. Characterization of the chromosomal gene and expression of the cDNA in *Escherichia coli*. European Journal of Biochemistry 179:185 - 194

Guo W, Gonzalex-Candelas L, Kolattukudy PE (1995) Cloning of a new pectate lyase gene *pelC* from *Fusarium solani* f. sp. *pisi* (*Nectria haematococca* mating type VI) and characterization of the gene product expressed in *Pichia pastoris*. Archives in Biochemistry and Biophysics 323:352 - 360

Gupta R, Beg Q, Lorenz P (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. Applied Microbial Biotechnology 59:15 - 32

Gurr S, Unkies S, Kinghorn J (1988) The structure and organization of nuclear gene of filamentous fungi, Vol 22. IRL Press, Oxford

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nuceic Acids Symposium Series 41:95 - 98

Hamlyn PF (1998) Fungal Biotechnology. British Mycological Society Newsletter

Hartley B (1960) Proteolytic enzymes. Annual Review of Biochemistry 29:45 - 72

Have At, Dekkers E, Kay J, Phylip LH, Kanl JALv (2004) An aspartic proteinase gene family in the filamentous fungus *Botrytis cinerea* contains members with novel features. Microbiology 150:2475 - 2489

Heerikhuisen M, Van den Hondel C, Punt P, Van Biezen N, Albers A, Vogel K (2000) Novel means of transformation of fungi and their use for heterologous protein production. In: Patent W (ed)

Hoffman B, Breuil C (2002) Cloning and genetic analysis of subtilases in sapstaining fungi. Current Genetics 41:168 - 175

Hoffman B, Breuil C (2004a) Analysis of the distribution and regulation of three representative subtilase genes in sapstaining fungi. Fungal Genetics and Biology 41:274 - 283

Hoffman B, Breuil C (2004b) Disruption of the subtilase gene, albin1, in *Ophiostoma piliferum*. Applied and Environmental Microbiology 70:3898 - 3903

Hofmeister A, Londono-Vallejo A, Harry E, Stragier P, Losick R (1995) Extracellular signal protein triggering the proteolytic activation of a developmental transcription factor in *B. subtilis*. Cell 83:219 - 226

Horiuchi H, Yanai K, Oka0zaki T, Takagi M, Yano K (1988) Isolation and Sequencing of a Genomic Clone Encoding Aspartic Proteinase of *Rhizopus niveus*. Journal of Bacteriology 12:272 - 278

Hou Y-H, Wang T-H, Long H, Zhu H-Y (2006) Novel cold-adaptive *Penicillium* strain FS 010 secreting thermo-labile xylanase isolated from Yellow Sea. Acta Biochimica et Biophysica Sinica 38:142

Hu G, St. Leger J (2004) A phylogenomic approach to reconstructing the diversification of serine proteases in fungi. Journal of Evolutionary Biology 17:1204 - 1214

Ikemura H, Tagai H, Inouye M (1987) Requirement of prosequence for the production of active subtilisin E in *Escherichia coli*. Journal of Biological Chemistry 262:7859 - 7864

Inoue H, Kimura T, Makabe O, K T (1991) The gene and deduced protein sequences of the zymogen of *Aspergillus niger* acid proteinase. American Journal of Biological Chemistry 266:19484-19489

Isogai T, Fukagawa M, Kojo H, Kohsaka M, Aoki H, Imanaka H (1991) Cloning and nucleotide sequences of the complementary and genomic DNAs for the alkaline protease from *Acremonium chrysogenum*. Agriculture Biology and Chemistry 55:471 - 477

Jalving R, Peter J, vondervoort Vd, visser J, Schaap PJ (2000) Characterization of the kexin-like maturase of *Aspergillus niger*. Applied and Environmental Microbiology 1:363 - 368

Jany K-D, Lederer G, Mayer B (1986) Aminoacid sequence of proteinase K from the mold *Tritirarchium album* limber. FEBS 199:139 - 144

Jany K, Mayer B (1985) Proteinase K from *Tritirarchium album* limber. I. Molecular mass and sequence around the active site serine residue. Biological Chemistry Hoppe-Seyler 366:485 - 492

Jara P, Delmas P, Razanamparany V, Olsen L, Dupin P, Bayol A, Begueret J, Loison G (1995) Self-cloning in filamentous fungi: application to the construction of

endothiapepsin overproducers in *Cryphonectria parasitica*. Journal of Biotechnology 40:111 - 120

Jara P, Gilbert S, Delmas P, Guillemot J, Kaghad M, Ferrara P, Loison G (1996) Cloning and characterization of the eapB and eapC genes of *Cryphonectria parasitica* encoding two new acid proteinases, and disruption of eapC. Molecular Gene and Genetics 250:97 - 105

Jarai G, Kirchherr D, Buxton FP (1994a) Cloning and characterization of the pepD gene of *Aspergillus niger* which codes for a subtilisin-like protease. Gene 139:51 - 57

Jarai G, van den Hombergh H, Buxton F (1994b) Cloning and characterization of the pepE gene of *Aspergillus niger* encoding a new aspartic protease and regulation of pepE and pepC. Gene 145:171 - 178

Jasmin C, Sreeja C, Chandrasekaran M (2006a) *Engyodontium album* 5.8S rRNA gene and Internal transcribed spacer1 and 2, complete sequence. NCBI accession No: DQ 872370

Jasmin C, Sreeja C, Chandrasekaran M (2006b) *Engyodontium album* large subunit ribosomal RNA gene partial sequence. NCBI accession No: DQ 872371

Jasmin C, Sreeja C, Chandrasekaran M (2006c) *Engyodontium album* large subunit ribosomal RNA gene, partial sequence. NCBI accession No: DQ 872372

Jaton-Ogay K, Suter M, Crameri R, Falchetto R, Fatih A, Monod M (1992) Nucleotide sequence of a genomic and a cDNA clone encoding an extracellular alkaline protease of *Aspergillus fumigatus*. FEMS Microbiology Letters 71:163 - 168

Joha J-H, Kimb B-G, Kongb W-S, Yoob Y-B, Kima N-K, Parka H-R, Choa B-G, Leea C-S (2004) Cloning and developmental expression of a metzincin family metalloprotease cDNA from oyster mushroom *Pleurotus ostreatus*. FEMS Microbiology Letters 239:57 - 62

Johnson SM, Kerekes KM, Roger Zimmermann C, Williams RH, Pappagianis D (2000) Identification and cloning of an aspartyl proteinase from *Coccidioides immitis*. Gene 241:213 - 222

Joshi L, Leger RJS (1999) Cloning, Expression, and Substrate Specificity of MeCPA,a Zinc Carboxypeptidase That Is Secreted into Infected Tissues by the Fungal Entomopathogen *Metarhizium anisopliae*. The Journal of Biological Chemistry 274:9803 - 9811

Joshi L, Leger RJS, Bidochka MJ (1995) Cloning of a cuticle degrading protease from the entamopathogenic fungus, *Beauveria bassiana*. FEMS Microbiology Letters 125:211 - 218

Joshi L, Leger RJS, Roberts DW (1997) Isolation of a cDNA encoding a novel subtilisin like protease (Pr1B) from the entomopathogenic fungus, *Metarhizium anisopliae* using differential display- RT - PCR. Gene 197:1 - 8

Jousson O, Lechenne B, Bontems O, Mignon B, Reichard U, Barblan J, Quadroni M, Monod M (2004) Secreted subtilisin gene family in *Trichophyton rubrum*. Gene 339:79 - 88

Kanda S, Aimi T, Kano S, Ishihara S, Kitamoto Y, Morinaga T (Article in Press) Ambient pH signaling regulates expression of the serine protease gene (sprl) in pine wilt nematodetrapping fungus, *Monacrosporium megalosporum*. Microbiological Research

Katz B, Kossiakoff A (1990) Crystal structures of subtilisin BPN' variants containing disulfide bonds and cavities: Concerted structural rearrangements induced by mutagenesis. Proteins structural and Functional Genetics:7343 - 7357

Katz ME, Mcloon M, Burrows S, Cheetham BF (1998) Extreme DNA sequence variation in isolates of *Aspergillus fumigatus*. FEMS Immunology and Medical Microbiology 20:283 - 288

Katz ME, Rice R, Cheetham BF (1994) Isolation and characterization of an *Aspergillus nidulans* gene encoding an alkaline protease. Gene 150:287 - 292

Keniry C, Li D, Ashby A (2002) Cloning and expression studies during vegetative growth and sexual development of Psp2, a serine protease gene from *Pyrenopeziza brassicae*. Biochemical Biophysical Acta 1577:159 - 163

Kim C-f, Lee SKY, Price J, Jack RW, Turner G, Kong RYC (2003) Cloning and expression analysis of the *pcbAB-pcbC* b-lactam genes in the marine fungus *Kallichroma tethys*. Applied and Environmental Microbiology 69:1308 - 1314

Kim H-K, Hoe H-S, Suh DS, Kang SC, Hwang C, Kwon S-T (1999) Gene structure and expression of the gene from *Beauveria bassiana* encoding bassiasin I an insect cuticle-degrading serine protease. Biotechnology Letters 21:777 - 783

Kingsnorth C, Eastwood D, Burton K (2001) Cloning and postharvest expression of serine proteinase transcripts in the cultivated mushroom *Agaricus bisporus*. Fungal Genetics and Biology 32:165 - 144

Klionsky D, Banta L, Emr S (1988) Intracellular sorting signal and processing of a yeast vacuolar hydrolase: proteinase A propeptide contains vacuolar targeting information. Molecular Cell Biology 5:2105 - 2116

Kojima S, Minagawa T, Miura K (1997) The propeptide of subtilisin BPNP as a temporary inhibitor and e¡ect of an amino acid replacement on its inhibitory activity. FEBS Letters 411:128 - 132

Kolattukudy PE (1985) Enzymatic penetration of the plant cuticle by fungal pathogens. Annual Review of Phytopathology 23:233

Kolattukudy PE, Lee JR, LM, Zimmerman P, Ceselski S, Fox B, Stein B, Copelan E (1993) Evidence of possible involvement of an elastolytic serine protease in aspergillosis. Infection and Immunity 61:2357 - 2368

Kolvenbach C, Narhi L, Lazenby K, Samal B, Arakawa T (1990) Comparative study on proteinase R, T, and K from *Tritirachium album*. International Journal of Peptide and Protein Research 36:387 - 391

Komeda T, Saka Y, Kato N, Kondo K (2002) Construction of Protease-deficient *Candida boidinii* Strains Useful for Recombinant Protein Production: Cloning and Disruption of Proteinase A Gene (PEP4) and Proteinase B Gene. Bioscience Biotechnology and Biochemistry 66:628 - 631

Kothary M, Chase T, MacMillan J (1984) Correlation of elastase production by some strains of *Aspergillus fumigatus* with ability to cause pulmonary invasive aspergillosis inmice. Infection and Immunity 43:320- 325

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283 - 292

Kraut J (1977) Serine proteases:str and mechanism of catalysis. Annual Review of Biochemistry 46:331 - 358

Kumar C, Joo H, Koo Y, Paik S, Chang C (2004a) Thermostable alkaline protease from a novel marine haloalkalophilic *Bacillus clausii* isolate. World Journal of Microbiology and Biotechnology 20:351 - 357

Kumar S, Tamura K, Nei M (2004b) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5:150 - 163

Kutty G, Kovacs J (2003) A single-copy gene encodes Kex1, a serine endoprotease of *Pneumocystis jiroveci*. Infection and Immunity 71:571 - 574

Kwon BK, Han KH, Han KY, Ju SM, Hwang SG, Jeon BH, Han DM, Kim WS (2001) Molecular cloning of *kpcA* gene encoding a *Kex2p*-like endoprotease from *Aspergillus nidulans*. Molecules and Cells 12:142 - 147

Kwon S, Terada I, Matsuzawa H, Ohta T (1988) Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine protease) of *Thermus aquaticus* YT-1 and characterestics of the deduced primary structure of the enzyme. European Journal of Biochemistry 173:491 - 497

Kyte J, Doolittle R (1982) A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157:105 - 132

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a Programs to check the steriochemical quality of protein structures. Journal of Applied Crystallography 26:283 - 291

Lee B, Takeuchi M, Kobayashi Y (1995) Molecular cloning and sequence analysis of the scpZ gene encoding the serine carboxypeptidase of *Absidia zychae*. Current Genetics 27:159 - 165

Lee L, Gigliotti F, Wright T, Simpson-Haidaris P, Weinberg G, Haidaris C (2000) Molecular characterization of KEX1, a kexin-like protease in mouse *Pneumocystis carinii*. Gene 242:141 - 150

Li Y, Hu Z, Jordan F, Inouye M (1995) Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding: refolding and inhibitory abilities of propeptide mutants. The Journal of Biological Chemistry 270:25127 - 25132

Lu J, Inoue H, Kimura T, Makabe O, Takahashi K (1995) Molecular cloning of a cDNA for proctase B from *Aspergillus niger* var. macrosporus and sequence comparison with other aspergillopepsins I. Bioscience Biotechnology and Biochemistry 59:954 - 955

Lugli E, Allen A, Wakefield A (1997) A Pneumocystis carinii multi-gene family with homology to subtilisin-like serine proteases. Microbiology 143:2223 - 2236

Makino K, Koshikawa T, Nishihara T, Ichikawa T, Kondo M (1981) Studies on protease from a marine bacteria. Isolation of marine *Pseudomonas* 145 - 2 and purification of protease. Microbios 31:103 -112

Marin F, Groot Kd, Westbroek P (2003) Screening molluscan cDNA expression libraries with anti-shell matrix antibodies. Protein Expression and Purification 30:246 - 252

Martin CC, Bischof LJ, Bergman B, Hornbuckle LA, Hilliker C, Frigeri C, Wahl D, Svitek CA, Wong R, Goldman JK, Oeser JK, Leprêtre F, Froguel P, O'Brien RM, Hutton JC (2001) Cloning and characterization of the human and Rat islet-specific glucose-6-phospahatase catalytic subunit-related protein (IGRP) genes. Journal of Biological Chemistry 276:25197 - 25207

Matsumura M, Signor G, Matthews B (1989) Substantial increase of protein stability by multiple disulphide bonds. Nature 342:291 - 293

McPhalen C, James M (1988) Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin-C-subtilisin Carlsberg and CI-2-subtilisin Novo. Biochemistry 27:6582 - 6598

Meloun B, VBaudys M, Kostka V, Hausdorf G, Frommel C, Hohne W (1985) Complete primary structure of thermitase from *Thermoactinomyces vulgaris* and its structural feature related to the subtilisin type proteinases. FEBS Letters 183:195 - 200

Menon A, Goldberg A (1987) Protein substrates activate the ATP-dependent protease La by promoting nucleotide binding and release of bound ADP. Journal of Biological Chemistry 262:14929 - 14934

Mitchinson C, Wells J (1989) Protein engineering of disulfide bonds in subtilisin BPN'. Biochemistry 28:4807 - 4815

Mizuno K, Nakamura T, Ohshima T, Tanaka S, Matsuo H (1998) Yeast KEX2 genes encodes an endopeptidase homologous to subtilisin-like serine proteases. Biochemistry and Biophysics Research Communications 156:246 - 254

Monigatti F, Gasteiger E, Bairoch A, Jung E (2002) The Sulfinator: predicting tyrosine sulfation sites in protein sequences. Bioinformatics 18:769 - 770

Moon J, Shaw L, Mayo J, Potempa J, Travis J (2006) Isolation and properties of extracellular proteinases of *Penicillium marneffei*. The Journal of Biological Chemistry 387:985 - 993

Moralejo FJ, Cardoza RE, Gutierrez S, Lombran M, Fierro F, Martín JF (2002) Silencing of the Aspergillopepsin B (pepB) Gene of *Aspergillus awamori* by Antisense RNA Expression or Protease Removal by Gene Disruption Results in a Large Increase in Thaumatin Production. Applied and Environmental Microbiology 68:3550 - 3559

Morita S, Kuriyama M, Maejima K, Kitano K (1994) Cloning and nucleotide sequence of the alkaline protease gene from *Fusarium* sp. S-19-5 and expression in

Saccharomyces cerevisiae. Bioscience Biotechnology and Biochemistry 58:621 - 626

Morton C, Hirsch P, Peberdy J, BR K (2003) Cloning of and genetic variation in protease VCP1 from the nematophagous fungus *Pochonia chlamydosporia*. Mycology Research 107:38 - 46

Murakami K, Ishida Y, Masaki A, Tatsumi H, Murakami S, Nakano E, Tatsumi H, Murakami S, Nakano E, Motai H, Kawabe H, Arimura H (1991) Isolation and characterization of the alkaline protease gene of *Aspergillus oryzae*. Agriculture Biology and Chemistry 55:2807 - 2811

Murphy J, Walton J (1996) Three extracellular proteases from *Cochliobolus carbonum*: cloning and targeted disruption of ALP1. Molecular Biology of Plant Microbe Interaction 9:290 - 297

Nakagawa Y (1970) Alkaline proteinases from *Aspergillus*. Methods in Enzymology 19:581 - 591

Neurath H (1975) proteases and biological control, Vol. Cold Springer Harbor, New York

North M (1982) Comparative biochemistry of the proteinases of eucaryotic microorganisms. MIicrobiological Reviews:308 - 340

O¨nnerfjord P, Heathfield TF, Heinegård D (2004) Identification of tyrosine sulfation in extracellular leucine-rich repeat proteins using mass spectrometry. The Journal of Biological Chemistry 279:26 - 33

Ohi H, Ohtani W, Okazaki N, Furuhata N, Ohmura T (1998) *Pichia pastoris*,carboxypeptidase Y PRC1. Gene 12:31 - 40

OhsumiI K, Matsuda Y, Nakajima H, Kitamoto K (2001) Cloning and Characterization of the *cpy*A Gene Encoding Intracellular Carboxypeptidase from A*spergillus nidulans*. Bioscience Biotechnology and Biochemistry 65:1175 - 7780

Orozco I, Ortiz L, Elorza M, Ruiz-Herrera J, Sentandreu R (2002) Cloning and characterization of PRB1, a *Candida albicans* gene encoding a putative novel endoprotease B and factors affecting its expression. Research in Microbiology 153:611 - 620

Ouarzane M, Labbé M, Péry P (1998) *Eimeria tenella*:cloning and characterization of cDNA encoding a S3a ribosomal protein. Gene 225:125 - 130

Pantoliano M, Ladner R, Bryan P, Rollence M, Wood J, Poulos T (1987) Protein engineering of subtilisin BPN': Enhanced stabilization through the introduction of two cysteines to form a disulfide bond. Biochemistry 26:2077 - 2082

Paoletti M, Castroviejo M, Begueret J, Clave C (2001) Identification and characterization of a gene encoding a subtilisin-like serine protease induced during the vegetative incompatibility reaction in *Podospora anserina*. Current Genetics 39:244 - 252

Peng Y, Yang X-J, Xiao L, Zhang Y-Z (2004) Cloning and expression of a fibrinolytic enzyme (subtilisin DFE) gene from *Bacillus amyloliquefaciens* DC-4 in *Bacillus subtilis*. Research in Microbiology 155:167 -173

Perlman D, Halvorson H (1983) A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. Journal of Molecular Biology 167:391 - 409

Pietro AD, Dolores Huertas-González M, Felix Gutierrez-Corona J, Martínez-Cadena G, Méglecz E, Isabel G. Roncero M (2001) Molecular Characterization of a Subtilase from the Vascular Wilt Fungus *Fusarium oxysporum*. Molecular Plant-Microbe Interactions 14:653 - 662

Poldermans B (1990) Proteolytic enzymes, Vol 1. Wenheim, Germany

Pozo MJ, Baek J-M, Garcia JM, Kenerley CM (2004) Functional analysis of *tvsp 1*, a serine protease encoding gene in the biocontrol agent *Trichoderma virens*. Fungal Genetics and Biology 41:336 - 348

Rajamani S, Hilda A (1987) Plate assay to screen fungi for proteolytic activity. Current Science 56:1179 - 1181

Ramesh MV, Sirakova T, Kolattukudy PE (1994) Isolation, characterization, and cloning of cDNA and the gene for an elastinolytic proteinases from *Aspergillus flavus*. Infection and Immunity 62:79 - 85

Ramesh MV, Sirakova TD, Kolattukudy PE (1995) Cloning and characterization of the cDNAs and genes (*mep 20*) encoding homologous metalloproteinase from *Aspergillus flavus* and *A. fumigatus*. Gene 121:121 - 125

Rao MB, Taksale AM, Ghatge MS, Deshpande VV (1998) Molecular and biotechnological aspects of microbial proteases. Microbiology and Molecular Biology Reviews 62:597 - 635

Rawlings N, Barrett A (1993) Evolutionary families of peptidases. Biochemistry Journal 290:205 - 218

Rawlings N, Barrett A (1994) Families of serine peptidases. Methods in Enzymology 244:19 - 61

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

Ritch T, Gold M (1992) Characterization of a highly expressed liginin-peroxidase-encoding gene from the basidiomycete *Phanerochaete chrysosporium*. Gene 118:73 - 80

Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. Journal of Molecular Biology 232:584 - 599

Salamanca M, Barria C, Asenjo J, Andrews B (2002) Isolation, purification and preliminary characterization of cryophilic proteases of marine origin. Bioseparation 10:237 - 241

Samal B, Karan B, Boone T, Chen K, Stabinsky Y (1990) Isolation and characterization of the gene encoding a novel, thermostable serine proteinase from the mould *Tritirarchium album* limber. Molecular Microbiology 4:1789 - 1792

Samal BB, Karan B, Boone TC, Chen KK, Rohde MF, Stabinsky Y (1989) Cloning and expression of the gene encoding a novel proteinase from *Tritirarchium album* limber. Gene 85:329 - 333

Sambrook J, Fritish EF, Maniatis T (1989) Molecular cloning. A Laboratory Manual, Vol. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain terminating inhibitors. Proceedings of National Academy of Science USA 74:5463 - 5467

Santamaria F, Reyers F (1988) Proteases produced during autolysis of filamentous fungi. Tranactions of British Mycological Society 91:217 - 220

Schechter I, Berger A (1967) On the size of the active site in proteases. Biochemistry and Biophysics Research Communications 27:157 - 162

Schoen C, Reichard U, Monod M, Kratzin H, Ruchel R (2002) Molecular cloning of an extracellular aspartic proteinase from *Rhizopus microsporus* and evidence for its expression during infection. Medical Mycology 40:61 - 71

Screen SE, Leger RJS (2000) Cloning, Expression, and Substrate Specificity of a Fungal Chymotrypsin; Evidence for lateral gene transfer from an actinomycete bacterium. The Journal of Biological Chemistry 275:6689 - 6694

Segers R, Butt T, Carder J, Keen J, Kerry B, Peberdy J (1999) The subtilisins of fungal pathogens of insects, nematodes and plants, distribution and variation. Mycology Research 103:395 - 402

Shen H, Wang C, Lin W, Lai H, Tam M, Chou H, Wang S, Han S (2001) cDNA cloning and immunologic characterization of Pen o 18, the vacuolar serine protease major allergen of *Penicillium oxalicum*. Journal of Laboratory and Clinical Medicine 137:115 - 124

Sheng J, An K, Deng C, Li W, Bao X, Qiu D (2006) Cloning a Cuticle-Degrading Serine Protease Gene with Biologic Control Function from *Beauveria brongniartii* and Its Expression in *Escherichia coli* Current Microbiology 53:124 - 128

Shinde U, Inouye M (1993) Intramolecular chaperones and protein folding. Trends in Biochemical Science 18:442 - 446

Siezen R, de VW, Leunissen J, Dijkstra B (1991) Homology modeling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. Protein Engineering 4:719 - 737

Siezen R, Leunissen J (1997) Subtilases: The superfamily of subtilisin like serine proteases. Protein Science 6:501 -523

Sirakova TD, Markaryan A, Kolattukudy PE (1994) Molecular cloning and sequencing of the cDNA and gene for a novel elastinolytic metalloproteinase from *Aspergillus fumigatus* and its expression in *Escherichia coli*. Infection and Immunity 26:4208 - 4212

Smithson S, Paterson I, Bailey A, Screen S, Hunt B, Cobb B, Cooper R, Charnley A, Clarkson J (1995) Cloning and characterisation of a gene encoding a cuticle-degrading protease from the insect pathogenic fungus *Metarhizium anisopliae*. Gene 166:161 - 165

Southern E (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98:503

Sreeja Chellappan (2005) Alkaline protease production by marine fungus *Engyodontium album* BTMFS 10. Cochin University of Science and Technology

Sreeja Chellappan, Jasmin C, Soorej M. Basheer, Elyas K, Bhat SG, Chandrasekaran M (2005) Production, purification and partial characterization of a novel protease from marine Engyodontium album BTMFS10 under solid state fermentation. Process Biochemistry 41 (4): 956 - 961

St. Leger R, Frank D, Roberts D, Staples R (1992) Molecular cloning and regulatory analysis of the cuticle-degrading protease structural gene from the entomopathogenic fungus *Metarhizium anisopliae*. European Journal of Biochemistry 204:991 - 1001

St. Leger R, Joshi L, Bidochka MJ, Roberts DW (1996) Construction of an improved mycoinsecticide overexpressing a toxic protease. Proceedings of National Academy of Science USA 93:6349 - 6354

Stragier P (1996) Molecular genetics of sporulation in *Bacillus subtilis*. Annual Review in Genetics 30:297 - 341

Su N, Yu C, Shen H, Pan F, Chow L (1999) Pen c 1, a novel enzymic allergen protein from *Penicillium citrinum*. Purification, characterization, cloning and expression. European Journal of Biochemistry 26:115 - 123

Suarez B, Rey M, Castillo P, Monte E, Llobell A (2004) Isolation and characterization of PRA1, a trypsin-like protease from the biocontrol agent *Trichoderma harzianum* CECT 2413 displaying nematicidal activity Applied Microbiology and Biotechnology 65:46 - 55

Suresh P, Chandrasekaran M (1999) Impact of process parameters on chitinase production by an alkalophilic marine *Beauveria bassiana* in solid state fermentation. Process Biochemistry 34:257–267

Taguchi S, Odaka A, Watanabe Y, Momose H (1995) Molecular characterization of a gene encodingextracellular serine protease isolated from a subtilisin inhibitor-deficient mutant of *Streptomyces albogriseolus* S-3253. Applied and Environmental Microbiology 61:180 - 186

Takagi H, Morinaga Y, Inouye M (1988) Mutant subtilisin E with enhanced protease activity obtained by site-directed mutagenesis. Journal of Biological Chemistry 263:19592 - 19596

Takagi H, Takahashi T, Momose H, Inouye M, Maeda Y, Matsuzawa H, Ohta T (1990) Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease. Journal of Biological Chemistry 265:6874 - 6878

Takeuchi Y, Noguchi S, Satow Y, Kojima S, Kumagai, Miura K, Nakamura K, Mitsui Y (1991) Molecular recognition at the active site of subtilisin BPN': Crystallographic studies using genetically engineered proteinaonceous inhibitor SSI (Streptomyces subtilisin inhibitor). Protein Engineering:4501 - 4508

Tatsumi H, Murakami S, Tsuji R, Ishida Y, Murakami K, Masaki A, Kawabe H, Arimura H, Nakano E, Motai H (1991) Cloning and expression in yeast of a cDNA clone encoding *Aspergillus oryzae* neutral protease II, a unique metalloprotease. Molecular Gene and Genetics 228:97 - 103

Tatsumi H, Ogawa Y, Murakami S, Ishida Y, Murakami K, Murakami A, Masaki A, Kawabe H, Arimura H, Motai H (1989) A full length cDNA clone for the alkaline protease from *Aspergillus oryzae*: Structural analysis and expression in *Sacharomyces cerevisiae*. Molecular Gene and Genetics 219:33 - 38

Teo JW, Zhang L-H, Poh CL (2003) Cloning and characterization of a metalloprotease from *Vibrio harveyi* strain AP6. Gene 303:147 - 156

Thomas P, Russell A, Fersh A (1985) Tailoring the pH dependence of enzyme catalysis using protein engineering. Nature 318:375 - 376

Thomson J, Higgins D, Gibson T (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acid Research 22:4773 - 4680

Tonouchi N, Shoun H, Uozumi T, Beppu T (1986) Cloning and sequencing of a gene for Mucor rennin, an aspartate protease from *Mucor pusillus*. Nucleic Acid Research 14:7557 - 7568

Tsujibo H, Miyamoto K, Tanaka K, Kaidzu Y, Imada C, Okami Y, Inamori Y (1996) Cloning and sequence analysis of a protease-encoding gene from the marine bacterium *Alteromonas* sp. strain O-7. Bioscience Biotechnology and Biochemistry 60:1284 - 1288

Tsujibo H, Miyamoto K, Tanaka K, Kawai M, Tainaka K, Imada C, Okami Y, Inamori Y (1993) Cloning and sequence of an alkaline serine protease encoding gene from the marine bacterium *Alteromonas* Sp. strain O - 7. Gene 136:247 - 251

Tunga R, Shrivastava B, Banerjee R (2003) Purification and characterization of a protease from solid state cultures of *Aspergillus parasiticus*. Process Biochemistry 38:1553 - 1558

Valls L, Hunter C, Rothman J, Stevens T (1987) Proteinase sorting in yeast: the localization determinant of yeast vacuolar carboxy peptidase Y resides in the propeptide. Cell 48:887 - 897

van den Hombergh J, Jarai G B, FP, Visser J (1994) Cloning, characterization and expression of pepF, a gene encoding a serine carboxypeptidase from *Aspergillus niger*. Gene 151:73 - 79

van den Hombergh J, van de Vondervoort P, Fraissinet-Tachet L, Visser J (1997) Aspergillus as a host for heterologous protein production: the problem of proteases. Trends in Biotechnology 15:256 - 263

Vankuyk PA, cheetham BF, Katz ME (2000) Analysis of two *Aspergillus nidulans* genes encoding extracellular protease. Fungal Genetics and Biology 29:201 - 210

Venancio E, Daher B, Andrade R, Soares C, Pereira I, Felipe M (2002) The kex2 gene from the dimorphic and human pathogenic fungus *Paracoccidioides brasiliensis*. Yeast 19:1221 - 1223

Venugopal M, Saramma A (2006) Characterization of alkaline protease from *Vibrio fluvialis* strain VM10 isolated from a mangrove sediment sample and its application as a laundry detergent additive. Process Biochemistry Article in Press

Viterbo A, Harel M, Chet I (2004) Isolation of two aspartyl proteases from *Trichoderma asperellum* expressed during colonization of cucumber roots. FEMS Microbiology Letters 238:151 - 156

Voorhorst WG, Warner A, Vos WMd, Siezen RJ (1997) Homology modelling of two subtilisin-like serine proteases from the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus stetteri*. Protein Engineering 10:905 - 914

Wang M, Yang J, Zhang K (2006) Characterization of an extracellular protease and its cDNA from the nematode-trapping fungus *Monacrosporium microscaphoides*. Canadian Journal of Microbiology 52:130 - 139

Wedde M, Jacobs M, Stahl U (1999) Fungi: important organisms in history and today. In: Oliver RP, Schweizer M (eds) Molecular fungal biology, Vol 1. Cambridge University Press, Cambridge, p 377

Wells J, Powers D (1986) In vivo formation and stability of engineered disulfide bonds in subtilisin. Journal of Biological Chemistry 261:6564 - 6570

Wilson LM, Barbara J (2005) *Leptosphaeria maculans*, a fungal pathogen of *Brassica napus*, secretes a subtilisin-like serine protease. European Journal of Pathology 112:23 - 29

Wlodawer A, Li M, Gustchina A, Oyama H, Dunn B, Oda K (2003) Structural and enzymatic properties of the sedolisin family of serine-carboxyl peptidases. Acta Biochimica Polonica 50:81 - 102

Wolf WM, Bajorath J, Muller A, Raghunathan S, SIngh TP, Hinrichs W, Saeger W (1991) Inhibition of proteinase K by methoxysuccinyl-Ala-Ala-Pro-Ala-chloromethyl ketone. The Journal of Biological Chemistry 266:17695 - 17699

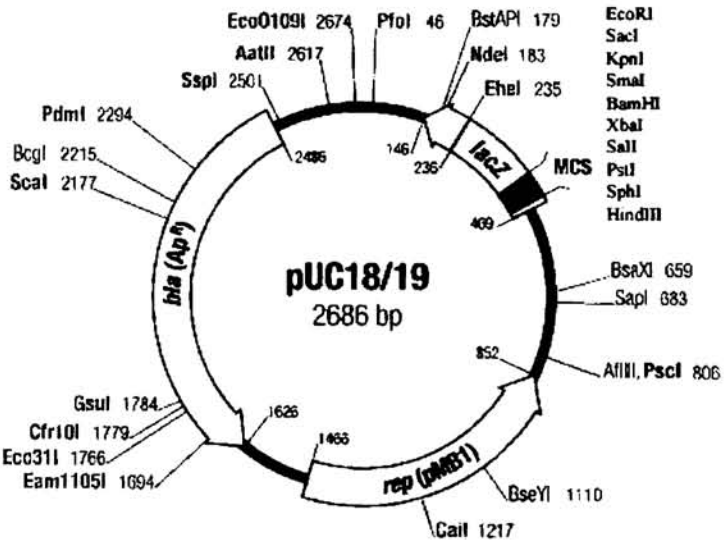
Woodfolk JA, Wheatley LM, Piyasena RV, Benjamin DC, Platts-Mills TAE (1998) Trichophyton Antigens Associated with IgE Antibodies and Delayed Type hypersensitivity sequence homology to two families of serine proteinases. The Journal of Biological Chemistry 273:29489 - 29496

Yabuta Y, Takagi H, Inouye M, Shinde U (2001) Folding pathway mediated by an intramolecular chaperone: Propeptide release modulates activation precision of pro-subtilisin. The Journal of Biological Chemistry 276:44427 - 44434

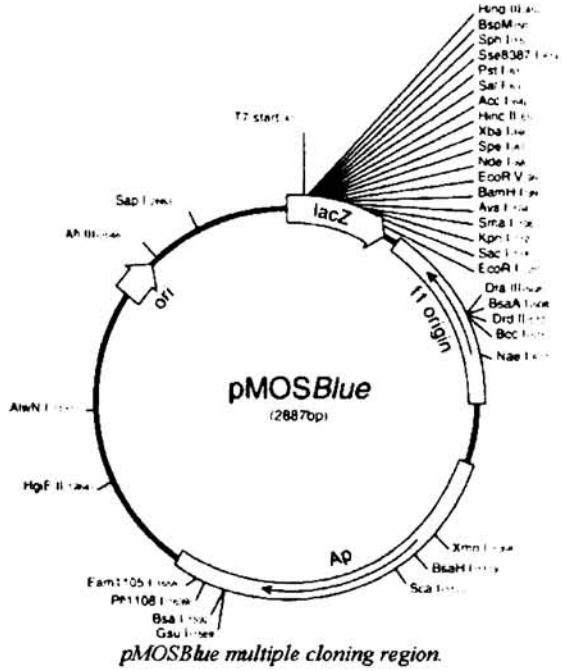
Yang J, Huang X, Tian B, Sun H, Duan J, Wu W, Zhang K (2005) Characterization of an extracellular serine protease gene from the nematophagous fungus *Lecanicillium psalliotae*. Biotechnology Letters 27:1329 - 1334

APPENDIX

pUC18 vector used for genomic DNA library construction



pMOS vector used for cloning PCR products



pMOSBlue multiple cloning region.

List of Publications

GenBank Accessions (NCBI, http//www.ncbi.nlm.nih.gov/)

Jasmin C, Sreeja C, Chandrasekaran M (2006) *Engyodontium album* 5.8S rRNA gene and Internal transcribed spacer1 and 2, complete sequence. NCBI accession No: DQ 872370

Jasmin C, Sreeja C, Chandrasekaran M (2006) *Engyodontium album* large subunit ribosomal RNA gene partial sequence. NCBI accession No: DQ 872371

Jasmin C, Sreeja C, Chandrasekaran M (2006) *Engyodontium album* large subunit ribosomal RNA gene, partial sequence. NCBI accession No: DQ 872372

Jasmin C, Sreeja C, Elyas K, Bhat S, Chandrasekaran M (2006) *Engyodontium album* alkaline protease gene sequence. NCBI accession No: DQ 268654

Jasmin C, Sreeja C, Elyas K, Bhat S, Chandrasekaran M (2006) *Engyodontium album* translation elongation factor-like mRNA, partial sequence. NCBI accession No: DQ268655

Manuscript under preparation

Jasmin C, Sreeja Chellappan, Elyas KK, Bhat SG, Chandrasekaran M (Under Preparation) Molecular cloning of alkaline serine protease gene of marine fungus *Engyodontium album* BTMFS10.